NOCIONES DE ESTADISTICA PARA PROFESIONALES DE LA SALUD

Dr. Emilio L. Morales Jiménez
Especialista de II Grado en Bioestadística
Profesor Consultante

Instituto Superior de Ciencias Médicas de la Habana
Facultad Finlay-Albarrán
Cuba

2006

INDICE

DERECHO DE AUTOR

El presente libro, **NOCIONES DE ESTADISTICA PARA PROFESIONALES DE LA SALUD**, ya sea en formato digital o impreso, está debidamente registrado en el Centro Nacional de Derecho de Autor, (CENDA), de la Republica de Cuba habiendo sido admitido en el área de depósito legal de esa Institución, a nombre del autor, Emilio Luis Morales Jiménez.

Hecho el depósito que establece la Ley.

No. Consecutivo: 0000207

No. del Registro: 503-2007

INDICE

A mis padres,

A mi esposa,

A mis hijos,

A mis nietos

INDICE

PROLOGO

Tradicionalmente el profesional y técnico de la salud, entiéndase fundamentalmente Médicos, Estomatólogos y Enfermeros han sido reacios a las matemáticas y a cualquier proceder que conlleve un cálculo numérico. Todavía el rechazo es mayor cuando es necesario interpretar y aplicar una fórmula matemática. Ello se evidencia desde que el estudiante de Medicina, Estomatología o Enfermería comienza la carrera donde muchos de ellos manifiestan abiertamente su aversión por las carreras de Ingeniería, Física y Matemáticas por su alto contenido matemático y algebraico.

Esta cuestión la hemos encontrado en diferentes países donde por distintas razones hemos asistido y hemos compartido estos criterios con sus especialistas. No es un problema nacional, sino una idiosincrasia de los que generalmente nos dedicamos a la medicina de una forma u otra.

Con el surgimiento de las nuevas tecnologías, primero calculadoras, sencillas y científicas, y luego las computadoras, sobre todo las microcomputadoras, los conocimientos sobre cálculos y operaciones matemáticas son cada vez menos en los estudiantes que nos llegan a la Universidad a estudiar las carreras de Ciencias Médicas. El estudiante puede saber hacer el cálculo con la calculadora y computadora pero es incapaz de razonarlo ni de hacerlo manualmente en la forma correcta en muchos casos. Ello tampoco es un fenómeno nacional. Ya en la década de los 70 y trabajando en la Dirección de Investigaciones del Instituto de Desarrollo de la Salud nos visitó un científico norteamericano, el Profesor Leslie Kish, que se ocupaba de enseñar Estadística en la carrera de Medicina en una Universidad de los EE.UU. compartiendo vivencias semejantes.

La Medicina, Ciencia y Arte a la vez, se basa en el método clínico que está muy distante de las matemáticas y las llamadas ciencias exactas aunque las ciencias médicas sean eminentemente probabilísticas. Por todo ello, es un reto tener que impartir una asignatura como la Bioestadística en las Ciencias Médicas.

No obstante, cada día mayor cantidad de profesionales de la salud, ven la necesidad de adentrarse en el campo de los números y utilizar procederes estadísticos para procesar información médica con vistas a presentar trabajos en eventos científicos, publicar artículos en revistas médicas, hacer una tesis de especialidad, maestría y doctorado o realizar una investigación científica. Eso los obliga a participar en cursos de post-grado de Metodología de la Investigación y de Bioestadística buscando las herramientas indispensables para su quehacer científico. Lamentablemente en muchas ocasiones el curso recibido, a pesar de tener un magnífico rigor académico, no es asimilable por ese profesional. La bibliografía que se utiliza no esta asequible al entendimiento de la mayoría de los participantes y al final, el profesional de la salud, no logra el objetivo que perseguía.

Otra forma de adquirir los conocimientos indispensables es la auto-preparación. Para ello se requiere bibliografía especializada. En este tipo de aprendizaje es imprescindible ir de lo simple a lo complejo. Los libros de Estadística ya sean de Estadística Matemática, Estadística Aplicada o de Estadística dirigida a las Ciencias Médicas o Biología por lo general están llenos de formulaciones matemáticas que dificultan mucho su asimilación por el profesional de la salud. ¿Qué hacer entonces? Lo más común es desistir y recurrir a personal especializado para que se ocupen del procesamiento y análisis de los datos.

De esta forma se delega en otros uno de los aspectos más creativos del quehacer científico. El procesamiento y análisis de los datos -ya bien sea en un trabajo científico, en una tesis de especialidad, maestría o doctorado o en una investigación científica,- por muy detallados que hayan sido planificados en la metódica o metodología del proyecto (protocolo) del trabajo en cuestión, siempre requiere de un análisis inquisitivo por parte del autor principal. Es muy frecuente, que durante este procesamiento y análisis surjan ideas y posibilidades de resultados que no se habían pensado con anterioridad. Realizarlos e introducirlos pueden dar un giro de 180 grados a los resultados haciéndolos más reales, profundos y explicativos. Esa es la importancia que tiene que el profesional de la salud sea capaz de participar activamente en esta fase de su quehacer científico y para ello requiere de conocimientos que se lo permitan.

El profesional de la salud requiere poder determinar como acometer el problema científico conjuntamente con el estadístico. No es que se convierta en estadístico, pero sí que tenga conocimientos de las herramientas estadísticas y de su posible utilización en el análisis concreto de su trabajo científico. La existencia de bibliografía asequible al razonamiento y manera de pensar del profesional de la salud, que le permita entender con relativa facilidad los procederes estadísticos, es por tanto una necesidad imperiosa en las Ciencias Médicas.

El libro Nociones de Estadística para Profesionales de la Salud se inscribe en este contexto. Es la recopilación enriquecida de los materiales docentes elaborados para los estudiantes de medicina siguiendo el programa de la asignatura a falta de un libro de texto entre 2001 y 2004. Tiene como antecedentes un libro del autor, elaborado y editado en Adén, Yemen, cuando estaba impartiendo la asignatura en la Facultad de Ciencias Médicas de la Universidad de Adén entre 1996 y 1997. Ese libro ya tiene dos ediciones. No es un libro de Estadística Matemática sino de la Estadística que más es utilizada por los profesionales de la salud. Está escrito en una terminología médica y minimizando las fórmulas matemáticas. Ello hace que pueda ser asimilado con más facilidad por el personal de la salud.

Aprovechamos la ocasión para agradecer infinitamente a todos aquellos que nos han impulsado y entusiasmado a terminar este empeño y en especial al Profesor Lic. Humberto Martínez Canalejo por sus consejos y experiencia.

Prof. Dr. Emilio L. Morales Jiménez

INDICE

CONTENIDO	Pag
Derecho de autor	**ii**
Dedicatoria	**iii**
Prólogo	**iv**
Indice	**vi**
Capítulo1. Introducción a la Estadística	**1**
◆ Concepto de Estadística.	
◆ Ramas de la Estadística: Estadística Descriptiva y Estadística Inferencial	
◆ Concepto de Universo y Muestra	
◆ El Método Estadístico. Etapas que lo constituyen: Planificación, Recolección, Procesamiento y Análisis e Interpretación. Características e importancia de cada etapa.	
◆ Aplicación del método científico y del método estadístico en problemas biomédicos.	
Capítulo 2. Recolección de la información	**11**
◆ Recolección de la información	
◆ Fuentes de recolección de información: Primaria y Secundaria	
◆ La observación y el interrogatorio. Características de cada una.	
◆ Formulario. Estructura. Elementos fundamentales a tener en cuenta en su confección.	
◆ Procedimientos para recolectar información según su frecuencia: Encuesta, Registro y Censo. Características.	
◆ Errores más comunes en la recogida de información.	
◆ Papel del profesional de la salud en a recolección de información.	
Capítulo 3. Procesamiento de la información	**23**
◆ Procesamiento de la información. Etapas que constituyen el procesamiento: organización, resumen y presentación.	
◆ Características de la organización y clasificación de los datos.	
◆ Concepto de variable.	
◆ Clasificación de la información atendiendo al tipo de variable: Cualitativa (nominal y ordinal) y Cuantitativa (discreta y continua)	
Capítulo 4. Distribución de frecuencias	**31**
◆ Distribuciones de frecuencia según el tipo de variable.	
◆ Concepto de clase o intervalo. Límite inferior y superior de la clase. Punto medio de la clase. Rango. Amplitud de clase.	
◆ Frecuencia absoluta, relativa y acumulada.	
Capítulo 5. Resumen de la información	**40**
◆ Resumen de la información	
◆ Medidas de tendencia central (media aritmética, mediana y moda). Características, propiedades, cálculo e interpretación.	
◆ Medidas de dispersión. Recorrido de la variable o amplitud total. Varianza, desviación estándar y coeficiente de variación. Características, propiedades, cálculo e interpretación. Variabilidad en Medicina y Biología.	
◆ Medidas de posición relativa (percentiles). Características, cálculo e interpretación. Su utilidad en los estudios de crecimiento y desarrollo.	
◆ Medidas para resumir datos cualitativos (razón, índice, proporción, porciento y tasa). Forma de cálculo e interpretación. Ejemplos en el campo médico.	

CONTENIDO **Pag**

Capítulo 6. Presentación de la información **58**

- Presentación de la Información.
- Cuadro o tabla estadística. Partes que la constituyen. Diferentes tipos de tablas estadísticas.
- Gráficos estadísticos. Tipos de gráficos acorde a la variable utilizada: gráfico de barras; gráfico de sector o pastel; histograma y polígono de frecuencias).
- Otros gráficos: gráfico aritmético simple y diagrama de dispersión.

Capítulo 7. Demografía y estadísticas de salud **76**

- Demografía. Concepto.
- Estadísticas de población. Su utilización en Salud Pública.
- Composición de la población. Pirámides de población. Diferentes tipos. Interpretación en países con diferentes grado de desarrollo.
- Concepto de estadísticas de salud. Estadísticas utilizadas en el campo de la salud
- Estadísticas de Natalidad. Principales conceptos. Medidas de fecundidad.
- Medidas de mortalidad. Principales conceptos. Mortalidad infantil y sus componentes. Mortalidad materna. Mortalidad general. Mortalidad específica.
- Medidas de morbilidad. Principales conceptos. Incidencia, Prevalencia, Letalidad. Morbilidad conocida y desconocida.
- Indicadores que miden los recursos y servicios disponibles y su grado de utilización. Médicos por mil habitantes. Camas por mil habitantes. Embarazadas captadas en primer trimestre. Mortalidad hospitalaria bruta y neta.
- Indicadores que miden los aspectos de higiene y saneamiento.

Capítulo 8. Introducción a la estadística inferencial **101**

- Conceptos de variable aleatoria y probabilidad.
- Modelo teórico de distribución o ley de una variable aleatoria.
- Modelo de la distribución normal. Parámetros de esta distribución. Propiedades.
- La distribución normal estándar. Ejemplos de aplicación.
- Tabla de Z.

Capítulo 9. Muestreo. Distribución muestral de la media aritmética cuando la distribución de la variable original es normal. Error estándar. **119**

- Muestreo. Ventajas y desventajas. Error de muestreo.
- Muestreo probabilístico y no probabilístico.
- Muestra representativa. Condiciones de una buena muestra.
- Esquemas de muestreo: muestreo aleatorio simple y muestreo estratificado.
- Distribución muestral de la media aritmética cuando la distribución de la variable original es normal. Error estándar.

Capítulo 10. Inferencia estadística. Estimación. **127**

- Estimación de parámetros poblacionales. Estimación puntual y por intervalos de confianza.
- Estimación por intervalo de confianza de la media poblacional (μ) con ó conocida y desconocida.
- Estimación por intervalo de confianza de una proporción poblacional (P).
- Precisión y confiabilidad de una estimación por intervalo.
- El tamaño de la muestra en función de la precisión (máximo error admisible) y la confiabilidad de la estimación.

CONTENIDO **Pag**

Capítulo 11. Inferencia estadística. Prueba de hipótesis. **140**

- Conceptos generales: Hipótesis nula e hipótesis alternativa. Prueba de hipótesis de una cola y de dos colas. Errores de Tipo I (á) y de Tipo II (â). Nivel de significación. Región crítica de una prueba de hipótesis.
- Prueba de hipótesis acerca de la media de una distribución normal cuando ó es conocida y desconocida.
- Prueba de hipótesis acerca de una Proporción Poblacional.
- Prueba de hipótesis sobre Diferencia de Medias.
- Prueba de hipótesis sobre diferencias de Proporciones Poblacionales.

Capítulo 12. Inferencia estadística. Método de Chi-cuadrado. **160**

- Relación entre variables.
- Tablas de contingencia.
- Estadígrafo Chi-cuadrado.
- Prueba no-paramétrica de independencia.
- Prueba no-paramétrica de homogeneidad.
- Tabla de Chi-cuadrado

Capítulo 13. Inferencia estadística. Correlación y regresión lineal. **172**

- Relación entre variables. Problemas de correlación y regresión lineal.
- Diagrama de dispersión. Utilidad.
- Coeficiente de correlación lineal poblacional o de Pearson **(**ñ**)** y Coeficiente de correlación lineal muestral **(r).**
- Prueba de hipótesis sobre el Coeficiente de correlación lineal poblacional.
- La regresión lineal.

Capítulo 14. Nociones de riesgo y su cuantificación. Método del valor predictivo. **184**

- Uso de la tabla de 2x2 para el cálculo del riesgo y del método del valor predictivo.
- Cálculo y uso del riesgo relativo (RR) y de la razón de productos cruzados o de ventajas (Odds Ratio - OR)
- Cálculo y uso del Método del Valor Predictivo. Sensibilidad, Especificidad, Valor predictivo positivo y Valor predictivo negativo.

Capítulo 15. La ciencia y la actividad científica. **195**

- El método científico.
- Etapas del método científico. El problema científico. Formulación de objetivos. Formulación de hipótesis. Planificación de la verificación de las hipótesis. Características e importancia.
- La observación científica. Importancia. Errores.
- Relación del método científico con los métodos de las ciencias particulares (método clínico y método epidemiológico)
- La investigación científica. Tipos de investigación científica.
- El proyecto de investigación científica. Importancia y partes que lo integran.
- El informe de investigación. Importancia y partes que lo integran.

Bibliografía Consultada **228**

INDICE

Capítulo 1
INTRODUCCION A LA ESTADISTICA

Autor: Dr. Emilio Morales Jiménez

Sumario

- *Concepto de Estadística.*
- *Ramas de la Estadística: Estadística Descriptiva y Estadística Inferencial*
- *Concepto de Universo y Muestra*
- *El Método Estadístico. Etapas que lo constituyen: Planificación, Recolección, Procesamiento y Análisis e Interpretación. Características e importancia de cada etapa.*
- *Aplicación del método científico y del método estadístico en problemas biomédicos.*

Concepto de Estadística.

La palabra Estadística proviene de la palabra ***"estado"*** dado el hecho que en la antigüedad, especialmente en China, Egipto y otras civilizaciones antiguas la recolección de datos se utilizaban con fines militares o de recaudación de impuestos.

Hoy en día suele usarse con 2 acepciones diferentes:

- estadísticas (minúscula y plural): Significa un conjunto de datos numéricos, tales como: estadísticas de salud, estadísticas deportivas, estadísticas agrícolas, etc...

- Estadística (mayúscula y singular): Se refiere a la ciencia de la Estadística que también se le ha llamado la ciencia de la experimentación. La Estadística es la ciencia que nos permite tomar decisiones frente a la incertidumbre.

Hay muchas definiciones de Estadística. Por ejemplo, Paul G. Hoel en su libro de "Estadística Elemental" la considera un sinónimo de método estadístico y la define como: ***"métodos para obtener conclusiones de una población por medio de muestras".*** En el libro de texto del ISCM-H, Bioestadística y Computación se define como ***"la ciencia que provee diferentes técnicas y procedimientos que permite organizar la recolección de datos, su procesamiento, análisis e interpretación".*** En el libro "Estadísticas de Salud y Metodología de Investigación" del Dr. Emilio Morales et al. se define como ***"la ciencia relacionada con métodos científicos para la recolección, organización, resumen, presentación y análisis de datos, así como derivar conclusiones válidas y tomar decisiones razonables sobre la base de dichos análisis".*** Esta última nos parece más completa pues incluye en la definición las dos partes didácticas básicas o ramas de la estadística: la estadística descriptiva y la estadística inferencial. No obstante esa última definición puede hacerse aún mas completa de la siguiente forma: ***"la ciencia relacionada con métodos científicos para la recolección, organización, resumen, presentación y análisis de datos, así como, teniendo en cuenta la incertidumbre, derivar***

conclusiones válidas y tomar decisiones razonables sobre la base de dichos análisis".

La Estadística como ciencia cobra cada día mayor importancia como parte fundamental del método científico, como herramienta básica de auxilio en la investigación.

Estadística descriptiva y estadística inferencial

Con fines docentes se distinguen en la Estadística dos grandes ramas: Estadística Descriptiva y Estadística Inferencial.

Estadística Descriptiva: Está ***relacionada con el procesamiento primario de los datos limitándose a la recolección, resumen y presentación de los datos***. Esta parte de la estadística provee diferentes procederes estadísticos que solamente permiten describir y analizar uno o más grupos de elementos sin que se puedan derivar conclusiones o inferencias de las características que exhiben los datos a un grupo mayor. Las conclusiones las obtiene el propio investigador según su juicio y experiencia sin tener en cuenta el grado de incertidumbre presente en los mismos. Es una descripción, una panorámica de la situación (como una fotografía instantánea de la muestra).

A pesar de sus limitaciones, la estadística descriptiva es muy útil en Medicina. Frecuentemente, en nuestro campo de aplicación, estamos trabajando con grupos de pacientes (por lo general pequeños grupos) siendo el interés principal el describir y analizar lo que se observa en dichos grupos. Prueba de ello lo tienen en la mayoría de los artículos que se publican en las revistas de Medicina o los que se presentan en eventos científicos que solamente hacen uso de la estadística descriptiva. Además esta rama es muy importante por ser la base para poder aplicar la estadística inferencial.

Estadística Inferencial: Es la ***parte de la Estadística que permite la toma de decisiones sobre los valores calculados con los datos de una muestra con relación a la población de donde se obtuvo la muestra.*** La estadística inferencial provee diferentes técnicas y procedimientos para el análisis de causas que son responsables o influyen en las características que exhiben los datos, descubrir relaciones entre los diversos factores involucrados (variables) y generalizar (inferir) los resultados a un grupo mayor (población o universo estadístico). Esta rama nos permite estudiar las muestras y hacer inferencia a la población, ayuda al investigador a obtener conclusiones teniendo en cuenta el grado de incertidumbre que puede estar presente en los datos. Esta parte de la Estadística se base fundamentalmente en la teoría de probabilidades.

Su utilidad en Medicina esta directamente relacionada con la investigación científica. De hecho, la mayoría de las investigaciones científicas realizadas en el campo de la

Medicina tienen que aplicar sus técnicas y procedimientos para llegar a conclusiones válidas y tomar decisiones razonables.

Concepto de universo y muestra

En el acápite anterior se utilizaron los términos población o universo y muestra. Debemos definirlos antes de seguir incursionando en la Estadística. En esta ciencia cuando se habla de universo o población se refiere a ***universo o población estadística***.

Universo o población: Conjunto o número total de elementos (personas, animales o cosas) que poseen al menos una característica común bien definida y susceptible de ser medida.

Muestra: Es un subconjunto cualquiera de la población o universo.

Los dos conceptos son frecuentemente utilizados en Estadística. Son conceptos relativos. Para la Estadística una población es el número total de elementos o todos los elementos de donde se requiere un análisis o estudio. También se le llama ***"universo o población de estudio"***. Por ejemplo:

- en un hospital o policlínico, las historias clínicas de todos los pacientes son el universo o población de historias clínicas del hospital o policlínico.
- en un área de salud de un consultorio de médico de familia, todos los pacientes dispensarizados por hipertensión arterial son la población de pacientes hipertensos dispensarizados del consultorio.
- en un área de salud de un consultorio de médico de familia, todas las embarazadas captadas son la población de embarazadas del consultorio.

Es de notar que cada población o universo estadístico tiene una definición que lo delimita, siendo una característica fundamental del mismo. No puede haber población o universo estadístico que no pueda ser definido.

La muestra que es una parte o subconjunto del universo, así se puede tener las siguientes muestras:

- un grupo de historias clínicas del archivo del hospital o policlínico fuera una muestra de las historias clínicas del archivo.
- un grupo de pacientes hipertensos dispensarizados de un área de un consultorio fuera una muestra de pacientes hipertensos dispensarizados de esa área.
- un grupo de embarazadas captadas en un área de un consultorio fuera una muestra de embarazadas captadas de esa área.

Se dijo que son conceptos relativos toda vez que una población o universo puede ser a su vez una muestra de un conjunto mayor. Así el universo o población de historias clínicas de todos los pacientes de un hospital o policlínico puede ser una muestra de

la población de historias clínicas de todos los hospitales o policlínicos del municipio. A su vez esta última población puede ser una muestra de todas las historias clínicas de los hospitales o policlínicos de una provincia.

Esto último recalca aún más la necesidad de tener siempre bien definido el universo o población de estudio.

El Método Estadístico. Etapas que lo constituyen: Planificación, Recolección, Procesamiento y Análisis e Interpretación. Características e importancia de cada etapa.

Método Estadístico.
Como toda ciencia, la Estadística tiene leyes, conceptos, técnicas y procedimientos particulares que le son inherentes. Para aplicarlas la Estadística provee un método científico llamado método estadístico. Puede ser definido como la ***forma organizada de aplicar las diferentes técnicas y procedimientos de la Estadística para el manejo de datos, o sea para la recolección, organización, procesamiento, resumen, presentación y análisis e interpretación de los datos, así como obtener conclusiones válidas y tomar decisiones razonables de un conjunto menor de datos, (muestra), hacia un conjunto mayor, (universo o población) de donde fueron extraídos.***

Este método lo proporciona la ciencia Estadística, siendo por tanto un método particular. Es la aplicación del método científico a la Estadística y es el método de la experimentación. Se considera como una de las herramientas de trabajo más útiles del Método Científico y tiene la particularidad de estar presente en todas las etapas del Método Científico.

El método estadístico puede ser aplicado por otras ciencias especialmente en a investigación científica. En Medicina y Salud Pública es obligado utilizarlo, no solamente para la investigación científica, sino también en el trabajo diario de los médicos e instituciones de salud. En efecto, para analizar los resultados de un tratamiento, para estudiar la sintomatología de una enfermedad, para evaluar la evolución de un grupo e pacientes, para analizar los indicadores de salud de una comunidad o para evaluar los servicios brindados por una unidad de salud entre una lista infinita de problemas médicos o de salud pública el método estadístico tiene un papel relevante.

Etapas del Método Estadístico
En el método estadístico se distinguen cuatro etapas:

1. Etapa de planificación
2. Etapa de recolección de datos
3. Etapa de elaboración o procesamiento de datos
4. Etapa de análisis e interpretación

- **Etapa de Planificación**

Es la primera etapa del método estadístico y permite organizar el trabajo requerido para cumplir todas las demás etapas. Algunos autores le llaman a esta etapa ***"planificación de la investigación"*** limitándola solamente a trabajos de investigación. El método estadístico es utilizado no solamente en investigación científica sino también en muchas otras actividades siendo mas adecuado nombrarlo de forma general: **PLANIFICACION**.

En esta etapa se crea un esquema organizativo de cómo vamos a utilizar dicho método, evitando improvisaciones en el mismo que en general introducen fuentes de errores capaces de invalidar el estudio. Dentro de esta etapa es necesario identificar cinco aspectos que se han denominado ***"sub-etapas, pasos, fases o partes"*** por diferentes autores. Ellas son:

- Planteamiento del problema
- Definición de objetivos
- Revisión y evaluación de información existente
- Formulación de hipótesis (Específicamente en la investigación científica)
- Planificación del trabajo o planificación de la verificación de la hipótesis

Estas sub-etapas no son inherentes al método estadístico sino a cualquier investigación, pero se han definido dentro de la planificación del método estadístico, por ser éste el método de la experimentación, el método de trabajo ideal para cualquier investigador.

- **Planteamiento del problema**

  Lo primero a realizar para aplicar el método estadístico es definir lo que se va a estudiar o analizar. Puede ser un problema científico, un estudio de morbilidad, el análisis de la evolución de un grupo de pacientes, en fin cualquier cosa puede ser analizada utilizando el método estadístico. Lo que ***es necesario es definir bien el problema, sin ambigüedades, conocer exactamente lo que se quiere.***

- **Definición de objetivos**

  Una vez planteado el problema se debe definir los objetivos. Los objetivos establecen lo que se quiere lograr y son las guías básicas para el trabajo a desarrollar.

- **Revisión y evaluación de la información existente**

  Cualquier análisis o estudio estadístico requiere de antecedentes bibliográficos de referencia. El estudio, análisis y evaluación de trabajos similares o sobre el problema son muy útiles para evitar duplicidades y errores, escoger el enfoque apropiado, comparar resultados y actualizar los conocimientos con la última información publicada sobre el mismo.

- **Formulación de hipótesis (Específicamente en la investigación científica)**
  Este aspecto es obligado en muchas investigaciones científicas, específicamente cuando la investigación se refiere a buscar una explicación o interpretación de un problema. Las hipótesis son las ideas que tienen los investigadores para explicar o interpretar el problema que van a ser verificadas mediante el trabajo de investigación.

- **Planificación del trabajo o planificación de la verificación de la hipótesis**
  Esta es la sub-etapa más importante de la planificación. Hay que definir y precisar claramente como el trabajo se va a realizar. Ello incluye:

  - la planificación de la recolección de datos
    - fuente de información,
    - procedimiento para la recolección de datos,
    - muestreo, etc..;

  - la planificación del procesamiento de datos
    - qué cómputos se realizarán,
    - cómo se presentará la información – plan de tabulación -, etc.);

  - la planificación del análisis de los datos
    - qué pruebas y técnicas estadísticas se utilizarán,
    - determinación de relaciones entre que variables,
    - determinación de riesgo,
    - aplicación de prueba de sensibilidad y especificidad, etc..);

  - la planificación de los recursos necesarios para el trabajo
    - recursos materiales y recursos humanos
    - costo –presupuesto-) y

  - un cronograma de todo el trabajo.

Etapa de Recolección de Datos

Es el paso donde se recoge la información y es una etapa importante que no debe ser subvalorada por el investigador. Solamente si la información es recogida con gran escrupulosidad y objetividad científica eliminando las fuentes de error podrán tener validez las conclusiones que de ellos se deriven La calidad de los datos es esencial de manera de obtener conclusiones confiables. Algunos de los aspectos fundamentales a tener en cuenta durante la recolección de datos son:

- Si se requiere una muestra, es imprescindible la selección adecuada de los elementos muestrales acorde con el método de muestreo empleado.
- La fuente de información debe ser confiable y cada dato debe ser recogido cuidadosamente.
- Se deben evitar errores durante la recolección de datos para evitar sesgo de la información que pueda llevar a conclusiones erróneas. Para lograr esto es

necesario chequear periódicamente la forma en que se obtienen y registran los datos.
- Solamente se deben recoger aquellos datos que van a ser procesados y analizados de manera de concentrar los esfuerzos en asegurar su calidad, ahorrar tiempo y recursos y facilitar el análisis al final del trabajo.
- Las técnicas utilizadas para la recolección de datos deben ser aplicadas correctamente.
- Se debe implementar una revisión y corrección periódica de los datos durante y al final de la recolección.

Resumiendo, si el dato recogido es exacto con buena calidad, los resultados serán confiables. De otra forma los resultados serán dudosos. Aquí es aplicable un término usado en computación, ***"BE-BO"*** que significa ***"basura entras – basura obtienes"***.

Etapa de Elaboración o Procesamiento de Datos

Una vez recogidos los datos, estos deben ser procesados de manera de poder ser analizados posteriormente. Usualmente al final de la recolección de datos, el médico, la enfermera, el estudiante de medicina, el trabajador de salud, o sea quien está haciendo el trabajo tiene una gran cantidad de cuestionarios u hojas de compilación llenos de datos o, si trabaja con computadoras, una base de datos llena. Es menester organizar, clasificar, resumir y presentar en tablas y gráficos estadísticos los datos recogidos o sea procesar los datos. El método estadístico provee la forma de realizar esto y con fines didácticos se ha dividido la etapa en tres sub-etapas:

- Organización y clasificación de los datos
- Resumen de los datos
- Presentación de los datos

- **Organización y clasificación de los datos**
  Primero que todo hay que organizar y clasificar los datos según las diferentes variables que se están estudiando.

- **Resumen de los datos**
  Es imposible hacer un análisis estadístico con los datos individuales de cada paciente u objeto estudiado en una investigación para llegar a conclusiones. Es obligado resumirlos para permitir el análisis e interpretación. Para ello, en dependencia del tipo de variable, existen diferentes procedimientos.

- **Presentación de los datos**
  Finalmente los datos deben ser presentados en una forma comprensible mediante tablas y gráficos estadísticos.

Es fácil de entender que el procesamiento de los datos es necesario para realizar cualquier trabajo estadístico. Este procesamiento puede hacerse manualmente o usando sistemas automatizados en dependencia de los recursos disponibles. Lo importante es utilizar el procedimiento adecuado. Frecuentemente esta etapa no se

cumple cabalmente pues se utilizan formas no adecuadas para organizar, clasificar, resumir y presentar los datos llevando a malas interpretaciones de los resultados o a falsas conclusiones.

Etapa de Análisis e Interpretación

Es la última etapa del método estadístico. Una vez organizada, resumida y presentada adecuadamente la información, es necesario analizar e interpretar los datos contrastando las hipótesis (si se formularon) con los datos ya procesados y llegar a determinadas conclusiones. La interpretación puede quedarse a nivel descriptivo o utilizar la estadística inferencial para realizar estudios comparativos. Existen diferentes técnicas y procedimientos en la estadística inferencial para el análisis e interpretación de los datos. La selección adecuada de ellas depende del tipo de estudio que se está realizando, de los objetivos, del método de muestreo, del tipo de variable, del número de elementos estudiados y otros factores. Un correcto análisis e interpretación de los resultados siempre lleva a conclusiones válidas, que es el principal objetivo de utilizar adecuadamente el método científico.

Aplicación del método científico y del método estadístico en problemas biomédicos.

En todos los países la Salud Pública está relacionada con la organización y administración de los diferentes recursos necesarios para los servicios y programas de salud. La aplicación del método científico y el método estadístico tienen un papel esencial en todas las actividades de Salud Pública.

Primero, es necesario conocer el estado de salud de la población lo que se logra, entre otros factores, por:

- Estadísticas demográficas para describir la composición, características y crecimiento de la población.
- Estadísticas vitales que describen a través de las estadísticas de natalidad y mortalidad los principales factores que afectan a la población en relación con los nacimientos y las causas de muerte.
- Estadísticas de morbilidad que brindan información sobre las enfermedades mas relevantes que afectan a la población.
- Estadísticas de la vivienda y el medio ambiente que permiten conocer las condiciones de las viviendas, centros de trabajo, higiene ambiental, etc. que pueden afectar la salud de la población.

Segundo, es necesario conocer la calidad y efectividad de los servicios de salud ofertados a la población. Esto se logra por:

- Estadísticas de recursos y servicios que permiten conocer la cantidad y calidad de los recursos, el uso de los recursos, el cumplimiento de programas de salud entre otras informaciones relevantes, necesarias para evaluar el trabajo de Salud Pública.

- Estadísticas hospitalarias que permiten conocer la calidad y cantidad de los servicios ofrecidos por estas instituciones.
- Análisis estadístico de los indicadores de salud de la población.

Tercera, la investigación en los Servicios de Salud Pública mediante estudios de terreno es necesaria para un adecuado control y desarrollo de las actividades de salud. Con relación a esto la Organización Mundial de la Salud (OMS) en la Serie de Informes Técnicos No. 510 de 1972 planteó que: ***..."La aplicación de estudios de terreno en los servicios de Salud Pública puede ser utilizada para solucionar algunos de los problemas involucrados en el desarrollo de los servicios de salud – un proceso que tiene tres fases: "***

"(1) Diagnóstico de la comunidad. En esta fase se valoran la frecuencia de las enfermedades y los recursos de salud en la comunidad, con énfasis en el grado de cumplimiento de las necesidades de prevención."

"(2) Investigación. Aquí se valoran rigurosamente nuevos métodos de control de enfermedades o cuidados de salud para determinar su eficacia en producir las acciones curativas o profilácticas requeridas."

"(3) Evaluación de servicios. La última fase consiste en estudios especiales de la efectividad con que los programas de salud basados en investigaciones preliminares están siendo aplicados en la práctica en la comunidad y de los resultados alcanzados."

Por otra parte en las últimas décadas el desarrollo de todas las ciencias y el surgimiento de nuevas tecnologías y nuevas ciencias han traído una mayor relación de interdependencia entre ellas. Hoy día, es común utilizar como medio diagnóstico complementario el ultrasonido computarizado, la tomografía axial computarizada (TAC), la resonancia magnética nuclear (RMN), electroencefalógrafos, electrocardiógrafos, ecocardiógrafos y electromiógrafos computarizados, equipos automáticos para conteo de células y análisis de sangre en laboratorio clínico, etc. Para el seguimiento del paciente crítico existen monitores computarizados, dispositivos automatizados para exámenes gasométricos y muchos otros equipos. Todas estas tecnologías de una forma u otra tienen incorporado análisis estadístico de datos que tiene que usarse e interpretarse por el médico o técnico que las opera.

Día a día la Medicina clásica está dando paso a una nueva Medicina donde los avances tecnológicos, la ingeniería genética e inmuno molecular, la biónica y la cibernética entre otras nuevas disciplinas tienen un papel predominante. La Estadística y el Método Científico tienen un papel relevante en todas ellas especialmente en la investigación científica. Ahora es muy común encontrar análisis estadísticos en los artículos de las revistas médicas de todas las especialidades y conclusiones avaladas por sólidos procederes estadísticos en la mayoría de los trabajos presentados en eventos científicos internacionales. Podemos decir que la

medicina cualitativa está siendo sustituida por una medicina cuali-cuantitativa, donde la Estadística tiene un papel predominante.

Muchos son los ejemplos en los cuales el médico y otros profesionales y técnicos de la salud necesitan utilizar del método estadístico y el método científico en su trabajo diario. Así, por mencionar algunos tenemos que:

- Para interpretar y analizar los resultados de un moderno electromiógrafo se requieren conocer procedimientos estadísticos de estandarización.
- Para interpretar y analizar los resultados de potenciales evocados en un electrencefalógrafo se requieren conocimientos estadísticos de promedios y estandarización.
- Para evaluar la efectividad de un nuevo proceder terapéutico tiene que usar procederes estadísticos.
- Para analizar la relación de diferentes factores con una enfermedad dada requiere usar técnicas estadísticas.
- Para llegar a conclusiones válidas con relación a una población de pacientes por medio del estudio de una muestra representativa de ella, necesita técnicas y procedimientos de la estadística inferencial.
- Para escribir un trabajo científico y mostrar los resultados de un estudio realizado necesita de procederes estadísticos.
- Para dar un pronóstico de la evolución de un paciente necesita del conocimiento estadístico sobre la evolución de la enfermedad que el paciente tiene.
- Para realizar una investigación científica en Medicina es obligatorio utilizar el método científico y el método estadístico.

INDICE

Capítulo 2
RECOLECCION DE LA INFORMACION.

Autor: Dr. Emilio Morales Jiménez

Sumario

- *Recolección de la información*
- *Fuentes de recolección de información: Primaria y Secundaria*
- *Técnicas de recolección de información. La observación y el interrogatorio. Características de cada una. Formulario. Estructura. Elementos fundamentales a tener en cuenta en su confección.*
- *Procedimientos para recolectar información según su frecuencia: Encuesta, Registro y Censo. Características.*
- *Errores más comunes en la recogida de información.*
- *Papel del profesional de la salud en a recolección de información.*

Recolección de la información

La recolección de información es la segunda etapa del método estadístico. Este es un paso importante, al que el investigador debe prestar especial atención, solo si la información ha sido recogida escrupulosamente, sin errores, podrá tener validez las conclusiones que de ella se derivan.

Esta etapa resulta a menudo olvidada por el investigador de lo cual resultan errores después insalvables. En el caso de cometerse errores en la recolección no se justifica siquiera el empleo de la Estadística ya que ninguna técnica podrá corregir los errores presentes en los datos.

Se debe recoger solamente la información necesaria. Recoger mayor información conspira contra la correcta interpretación de aquellos datos que son realmente necesarios, contra el costo, contra el tiempo invertido y, en general, contra el proceso investigativo. Los datos que deben ser recogidos, se establecen en el diseño experimental, basándose en los objetivos que persigue la investigación.

En la recolección de la información hay que tener en cuenta:

A. Fuentes de información
B. Método y procedimiento para la recolección de la información.
C. Diseño de formularios la recolección de la información.
D. Errores que pueden cometerse en la recolección de información.

Fuentes de recolección de información: Primaria y Secundaria

Teniendo en cuenta su origen, la información tiene dos posibles fuentes: primaria y secundaria.

Fuente primaria de información

Cuando un dato es recogido directamente del elemento que lo origina se define que proviene de una fuente primaria de información. La información no está registrada, los datos se recogen directamente del origen. Esta es la fuente de información más confiable y segura. En Medicina debe tratarse de usar siempre. En la práctica diaria del médico y del personal de enfermería, cada vez que se examina a un paciente se está recolectando información de la fuente primaria. En efecto, el paciente contesta todas las preguntas relativas al problema de salud que lo afecta y mediante el examen físico se recogen directamente del paciente los síntomas y signos. Cuando se está haciendo una investigación médica, los resultados serán mas confiables si la información es tomada directamente de los pacientes por los investigadores. Esto es así pues no media intermediario alguno, entre el que recolecta información y su fuente, que pueda introducir un sesgo informativo. ***La Medicina y la Salud Pública tienen al paciente como la principal fuente primaria de información.***

Fuente secundaria de información

Cuando el dato es recogido de una fuente intermedia, no directamente del elemento que la origina, se define que proviene de una fuente secundaria de información. En Medicina y Salud Pública las fuentes secundarias de información más comunes son las historias clínicas, los registros hospitalarios, los registros de nacimientos y defunciones y otros registros que pudieran existir en instituciones de salud u otras dependencias. Hay que diferenciar en las fuentes secundarias de información cual es confiable y cual no. Por lo general los registros oficiales del país son confiables, así los registros de nacimientos y defunciones deben ser confiables. En Cuba son confiables ambos. La historia clínica no es una fuente confiable de información. Existen circunstancias cuando no hay otra opción que utilizar fuentes secundarias de información; por ejemplo estudios retrospectivos de morbilidad y mortalidad que solamente pueden hacerse con registros hospitalarios y registro de defunciones. Sin embargo en investigaciones médicas sobre una determinada enfermedad, de la efectividad de un tratamiento, de la evolución de pacientes o similares se debe evitar utilizar fuentes secundarias realizando estudios prospectivos.

Los conceptos de fuente primaria y secundaria de información son dos conceptos relativos. Un dato puede ser clasificado de ambas formas al mismo tiempo. Por ejemplo, durante el examen de un paciente, el médico indica un análisis de sangre. Para el médico que los ordenó, los resultados provienen de una fuente secundaria de información, el registro del laboratorio clínico; pero para el técnico de laboratorio que procesó la muestra se utilizó la fuente primaria de información en este caso la sangre del paciente.

<u>Técnicas de Recolección de Información</u>

En este acápite abordaremos tres técnicas: la observación, el interrogatorio, y el formulario (cuestionario).

Las técnicas de recolección de datos permiten recolectar **sistemáticamente** la información del sujeto/objeto de estudio (paciente, experimento, u otro cualquiera) que se necesita para la investigación o el trabajo cotidiano. **Sistemáticamente se refiere a hacerlo siempre en la misma forma para evitar sesgos en la recogida de datos.**

El sesgo en la recolección de información se puede producir por:

- Por defecto en los instrumentos
  - Dificultades en la forma de hacer pregunta escrita u oral
  - Deficiente calibración y estandarización de los equipos de medición
- Sesgo del observador
- Sesgo del observado
- Sesgo del entorno

El sesgo se puede evitar con:

- planificación cuidadosa del proceso de recolección de datos
- prueba previa de las herramientas que se utilizarán.
- supervisión sistemática del proceso de recolección de datos

Siempre se deben tener en cuenta las consideraciones éticas durante la recolección de información. Para ello se debe prestar atención a lo siguiente:

- Se puede violar la privacidad mediante preguntas muy sensibles o accediendo a registros con información personal.
- No debe hacer observación sobre un sujeto sin su consentimiento.
- No respetar determinados valores culturales, tradiciones o tabúes del sujeto.

Las principales recomendaciones para cumplir con la ética son las siguientes:

- La obtención de consentimiento escrito del sujeto antes de comenzar el estudio.
- No explorar aspectos sensibles con el sujeto antes de haber establecido una adecuada relación con él.
- Asegurar la confidencialidad de los datos obtenidos.

<u>La observación y el interrogatorio. Características de cada una.</u>

La observación

La observación se define como el proceso perceptivo mediante el cual la atención es dirigida al elemento en estudio. La misma se analiza con profundidad en otro capítulo por lo que solamente nos limitaremos a su definición y señalar sus cinco factores componentes y algunas características:

Los cinco factores que intervienen en la observación científica son:

- El observador

- El sujeto u objeto observado
- El medio de observación
- El entorno que rodea la observación
- El sistema de conocimientos

Se debe tener en cuenta lo siguiente:
- Ha sido el método clásico de la investigación científica
- Siempre que sea posible se recomienda su uso por ser el procedimiento más objetivo.
- Exige entrenamiento previo del sujeto que la realiza.
- Cuando se requiere estudiar a grandes masas humanas se convierte en un proceso muy costoso.
- No es útil cuando el estudio necesita datos sobre manifestaciones subjetivas, hechos pasados e inquietudes futuras.

El interrogatorio

El interrogatorio puede ser:
- Directo (entrevista)
- Indirecto (por medio de cuestionarios)

- **La entrevista (interrogatorio directo)**

  La entrevista se define como un proceso activo de obtener información por medio de un interrogatorio oral y una observación del entrevistado por el entrevistador.

  En la entrevista hay que tener en consideración algunos factores:
  - Se debe preparar una guía de entrevista. (En la práctica médica diaria esta guía es la secuencia estructural de la historia clínica).
  - El entrevistador debe tener en cuenta la personalidad del entrevistado para evitar, durante el interrogatorio oral, respuestas incorrectas.
  - La inflexión de la voz y la expresión de la pregunta por el entrevistador no deben influenciar la respuesta.
  - Durante el interrogatorio oral se debe tener en cuenta el nivel cultural del entrevistado, debiéndose estar seguro que las preguntas se entendieron claramente.
  - También son importantes las condiciones del local, la fecha y la hora entre otros factores.

  En general la entrevista tiene las siguientes ventajas:
  - Combina la observación con la interrogación oral.
  - El entrevistador y entrevistado están en contacto directo.
  - Puede usarse para estudiar aspectos subjetivos, criterios y hechos retrospectivos.
  - Si existe duda en las respuestas o en la observación se pueden esclarecer durante la entrevista.

La entrevista tiene las siguientes desventajas:
- No es conveniente para estudiar gran cantidad de personas.
- Requiere personal altamente calificado
- Puede ser muy costosa.
- Los resultados del interrogatorio dependen de la memoria y las buenas intenciones del entrevistado.

En Medicina y Salud Pública este procedimiento tiene dos connotaciones especiales ya que es la técnica que el médico o cualquier personal de salud emplea diariamente. Ellas son:

- La entrevista médica incluye la observación (examen físico del paciente) y llenar un formulario (historia clínica) con los resultados del interrogatorio oral y el examen físico combinando las tres técnicas.
- Mediante la entrevista se realiza una cuestión muy importante en medicina **"la relación médico paciente"**.

La entrevista es la técnica más importante en Medicina y en Salud Pública por ser la forma en que la mayoría de las actividades profesionales se realizan. En efecto, es utilizada en la práctica médica diaria y en el trabajo de terreno de Salud Pública, en las investigaciones y en los estudios médicos. Como incluye todas las técnicas de recolección de datos es la más exacta y confiable para obtener información en nuestro campo de aplicación.

- **El Cuestionario (interrogatorio indirecto)**

El interrogatorio indirecto (cuestionario) se define como la aplicación de un cuestionario para ser llenado directamente por el paciente o un familiar.

Ventajas:
- Es adecuado para estudiar gran cantidad de personas.
- No requiere personal altamente calificado
- Puede ser utilizado para estudiar aspectos subjetivos, criterios o hechos retrospectivos.
- Es barato.

Desventajas:
- Las respuestas dependen de la memoria y la buena voluntad del que responde.
- Las respuestas dependen de la interpretación que hizo de la pregunta el que responde.
- Producirá diferentes resultados según el tipo de preguntas y la forma en que se formularon.
- No se puede garantizar que se reciban de retorno todos los cuestionarios que se entregaron para ser llenados en una investigación.

Se deben considerar las siguientes preguntas antes de diseñar el cuestionario:

- ¿Qué queremos saber exactamente?
- ¿Quién nos suministrará la información y que técnica emplearemos?
- ¿Entendemos el asunto lo suficiente como para diseñar el cuestionario o requerimos mas información para ello?
- ¿Cuál será el tamaño muestral? Si la muestra es grande generalmente se usan pocas variables y preguntas mayormente cerradas mientras que en muestras pequeñas se pueden estudiar más variables y emplear preguntas abiertas.

El diseño de un buen cuestionario lleva varias etapas o pasos y siempre lleva varios borradores:

- En el primer borrador nos debemos concentrar en el contenido.
- En el segundo, debemos mirar críticamente la formulación y secuencia de las preguntas.
- Luego debemos escudriñar el formato de las preguntas.
- Finalmente, debemos hacer una prueba para ver si el documento nos brinda realmente toda la información que se necesita. Por lo general, el cuestionario requerirá ajustes antes de que pueda ser usado para la recolección de datos.

La formulación de las preguntas debe atender a los siguientes aspectos:

- Formule una o más preguntas que proveerán la información que se necesita para cada variable. Analícelas y seleccione aquella que considere mejor.
- Tenga cuidado en formular preguntas específicas y precisas para evitar interpretaciones diferentes.
  Por ejemplo: ¿Dónde lleva Ud. a su hijo si tiene un trauma en el ojo?
  - Una madre puede pensar en una lesión grave con hemorragia y contestar el hospital especializado.
  - Otra madre puede pensar en una lesión de poca monta, sin hemorragia y decir el médico de la familia.
- La pregunta, por tanto, debe ser dividida en tantas partes como haga falta hasta hacerla específica y que no de lugar a interpretaciones distintas.

- Asegúrese que cada pregunta mida solamente una cosa de cada vez.
  Por ejemplo:
  ¿Qué tiempo de intervalo prefieren Ud. y su esposo entre dos hijos sucesivos?
  Sería preferible dividir esta pregunta en dos preguntas, una para el esposo y otra para la esposa pues pueden tener opiniones diferentes al respecto.

- Evite preguntas sugestivas
  Una pregunta sugestiva es aquella que sugiere la respuesta, por ejemplo,
  ¿Está Ud. de acuerdo en que si su hijo recibe un golpe en el ojo debe ser llevado al hospital?
  Esta pregunta no deja margen a la respuesta "no" u otra opción.
  Mejor sería, ¿Qué desearía se hiciera con su hijo de recibir un golpe en el ojo?

- Formule preguntas de control para chequeo cruzado de respuestas en preguntas "difíciles".
  Esto se debe hacer en preguntas sensibles o en aquellas donde es dificultoso obtener una respuesta precisa.

La secuencia de las preguntas debe asegurar lo siguiente:
- Un diseño del cuestionario en forma "amigable".
- La secuencia de las preguntas debe ser lógica.
- Uso de un lenguaje sencillo y cotidiano.
- Haga el cuestionario lo mas corto posible.

Formulario. Estructura. Elementos fundamentales a tener en cuenta en su confección.

Formulario
Un formulario se define como cualquier formato impreso dirigido hacia la recolección de datos que al mismo tiempo facilita la compilación de datos. En muchos casos el cuestionario tratado en párrafos precedentes cumple la misma función.

Son los documentos en los cuales se guarda la información recogida, son documentos impresos con el formato de preguntas y respuestas cuidadosamente seleccionadas y redactadas sobre la base de lo decidido en la etapa de planificación, de modo tal de no recoger datos innecesarios. Es el instrumento de recogida de información y en el se recogen dos tipos de datos:

a) de identificación.
b) específicos sobre el tema de estudio.

Estructura. Elementos fundamentales a tener en cuenta en su confección.
El formato del cuestionario debe tener en consideración lo siguiente:
- Encabezado para la identificación del cuestionario, número, fecha, datos generales, nombre del entrevistador para control de calidad.
- Las preguntas referentes a un mismo asunto deben estar juntas. Puede usar sub-encabezados por tópicos.
- Se debe dar el suficiente espacio para las respuestas de preguntas abiertas.

- Los escaques para marcar en las preguntas cerradas deben colocarse adecuadamente y si es posible hacia el mismo lugar de la página en todas las preguntas.

En el formulario se pueden colocar tres tipos de preguntas o acápites:
- preguntas cerradas
- preguntas abiertas
- preguntas semiabiertas o semicerradas

- **Preguntas Abiertas**
  Es aquella que permite una libre respuesta y que debe ser registrada con las propias palabras del que responde. Al que responde no se le dan opciones para escoger.

  ¿Qué Ud. haría exactamente si a su hijo le dan una pedrada en el ojo?
  ¿Cómo le comenzó el dolor?

- **Pregunta Cerrada**
  Es aquella que ofrece una lista de posibilidades o respuestas de donde se debe seleccionar.

  ¿Cuál es su estado civil? Soltero __; Casado __; Divorciado __; Viudo __;

  ¿Está trabajando? Si __; No__;

  ¿Cómo le comenzó el dolor? Haciendo esfuerzo físico __; Descansando __;
  Trabajo habitual __; Comiendo __; Durmiendo __;

  ¿Comió alguno de los siguientes alimentos ayer?
  - Frijoles si____ no ___
  - Pescado si ____ no ___
  - Huevos si ____ no ___
  - Arroz si ____ no ___

  En la pregunta cerrada se debe tratar de:
  - ofrecer una lista de opciones exhaustivas y mutuamente excluyentes. No obstante puede haber preguntas cerradas con mas de una respuesta como la de los alimentos formulada arriba.
  - tener el menor número de opciones posibles

  Se pueden usar para dar una opinión:
  ¿Cómo considera la atención recibida en el cuerpo de guardia?
  Muy rápida ___; Rápida ___; Normal ___; Lenta ___; Muy lenta ___;

- **Pregunta Semiabierta o Semicerrada**
  Es aquella que da opciones de posibles respuestas pero deja la posibilidad de una respuesta abierta.

  ¿Cómo le comenzó el dolor? Haciendo esfuerzo físico ___; Descansando ___;
  Trabajo habitual __; Comiendo __; Durmiendo __;
  Otra (especifique)_____;

  - Estas preguntas deben preguntarse como abiertas, no se debe brindar opción alguna.
  - Si la pregunta se hace en forma diferente por diferentes entrevistadores habrá sesgo.
  - El cuestionario debe esclarecer si se admite mas de una respuesta. La pregunta abierta o semiabierta generalmente permite mas de una respuesta.

Ventajas y Desventajas de Preguntas Abiertas y Cerradas y Condiciones para su Uso Óptimo

Pregunta abierta	Pregunta cerrada
Ventajas ❑ Pueden ser explorados algunos aspectos que no se pensaron inicialmente cuando se planificó el estudio, aportando elementos al problema. ❑ La información brindada espontáneamente por lo general es mas válida que las respuestas sugeridas en las opciones de donde se debe seleccionar. ❑ La información recogida textualmente del que responde puede ser de utilidad como ejemplos o para ilustrar de forma mas interesante el informe final.	**Ventajas** ❑ Las respuestas se registran rápido. ❑ El análisis es sencillo.
Desventajas ❑ Se requieren entrevistadores con experiencia para comenzar la conversación, dirigir la misma hacia los aspectos relevantes y registrar toda la información importante. ❑ El análisis es lento y requiere experiencia.	**Desventajas** ❑ Las preguntas cerradas son menos apropiadas en entrevistas cara a cara con personas de bajo nivel cultural. ❑ Los entrevistados pueden escoger opciones que nunca hubieran pensado por ellos mismos (sesgo por pregunta sugestiva). ❑ Se puede perder información importante si no se pregunta. ❑ El entrevistador y el entrevistado pueden perder interés después de un número determinado de preguntas.
Sugerencias ❑ Entrenar minuciosamente y supervisar a los entrevistadores o seleccionar personal experimentado. ❑ Preparar una lista de preguntas adicionales para tener a mano para usarla como “prueba” de las respuestas en forma sistemática. ❑ Pruebe previamente las preguntas y si es posible pre - categorize las respuestas más comunes, dejando espacio suficiente para otras respuestas.	**Sugerencias** ❑ Use preguntas cerradas sólo en los aspectos simples. ❑ Pruebe previamente las preguntas cerradas como abiertas para ver si las categorías cubren todas las posibilidades. ❑ Use preguntas cerradas en combinación con preguntas abiertas.

Ejemplos de formularios en el campo de la salud lo constituyen las historias clínicas, el certificado de defunción, certificados de nacimientos, fichas de enfermedades

transmisibles, los documentos que se elaboren para recoger información en investigaciones, modelos de encuestas, y muchos otros.

Debe decidirse de antemano (en la etapa de planificación, momento en que se confecciona este documento) si el procesamiento de la información será manual o automatizado, pues en general cuando se utilizan los procesamientos automatizadas, las preguntas no deben quedar abiertas y debe realizarse el formulario en función de las características del sistema que lo procesará.

Procedimientos para recolectar información según su frecuencia: Encuesta, Registro y Censo. Características.

Existen tres procedimientos para recolectar información según su frecuencia: el censo, la encuesta y el registro. Los tres son muy utilizados en Medicina y en Salud Pública.

La selección del método de recolección de información depende del propósito del estudio, de la naturaleza de la información y fundamentalmente de la frecuencia con que esta se recoja (ocasional, periódica o continuamente).

Censo. Sus características.

El censo se define como el procedimiento para recolectar información de todos los elementos de interés en un país o territorio determinado en un periodo específico de tiempo. En el censo se enumeran todos los elementos de interés y de cada uno se recolectan todos los datos a observar o a interrogar en un momento determinado. Por ejemplo, se puede definir un censo demográfico de la siguiente manera: proceso de recolectar, procesar y publicar datos demográficos, económicos y sociales de todos los habitantes de un territorio en un momento determinado.

Para la Medicina y la Salud Pública el censo mas importante, acorde a la información que suministra, es el Censo de Población y Vivienda. Este censo recoge información demográfica, económica y social de todos los habitantes de un país o de un territorio determinado en un periodo de tiempo. Brinda importante información sobre la población y las condiciones de vida necesarias a los médicos y funcionarios de Salud Pública como una ayuda para el diagnóstico del estado de salud de la población, como referencia en trabajos de terreno relacionados con la salud o una enfermedad dada, como un elemento necesario para planificar programas de salud y como información básica para las investigaciones o estudios médicos cuando se requiere una muestra representativa de la población, entre otros posibles usos.

El censo es como una fotografía o corte transversal de la población en un momento determinado, brindando información sobre ese momento. Usualmente se repite periódicamente en un periodo dado de tiempo. Tiene objetivos generales y propósitos universales pues sus resultados son usados por todos los posibles campos de aplicación. Dado su alto costo el censo de Población y Vivienda se realiza por lo general cada 10 años.

Encuesta. Sus características.
La encuesta se define como el procedimiento que se usa ocasionalmente y con un objetivo específico para recolección de información de una muestra o un sector dado de la población. Es el método mediante el cual se obtiene información de una parte específica de la población y con un propósito específico. Su propósito es parcial. Se realiza esporádicamente, cada vez que sea necesario conocer determinados datos, ejemplo, la encuesta de crecimiento y desarrollo que se realizó en nuestro país.

En Medicina y Salud Pública la encuesta es usada muy frecuentemente. La mayoría de las investigaciones y estudios en Medicina utilizan este procedimiento para la recolección de datos. En Salud Pública también es utilizado para la evaluación del cumplimiento de programas de salud, para la evaluación del estado de salud de la población o de un sector específico de ésta y para la evaluación de la situación de una enfermedad dada entre otros propósitos que pueden ser mencionados.

Registro. Sus características.
Un registro se define como el procedimiento para la recolección continua y sistemática de una información específica. Es un método de recoger información de forma sistemática y continua. Solamente se recoge la información relacionada con el evento que se registra. Un registro tiene objetivos específicos pero su propósito es universal porque la información es utilizada por diferentes campos de aplicación.

En Medicina y Salud Pública son muy utilizados. El registro de nacimientos, de defunciones, de enfermedades transmisibles, de enfermedades de declaración obligatoria, los registros hospitalarios, y otros son necesarios para obtener información básica que permite hacer: análisis de series cronológicas en relación con la salud de la población, análisis del comportamiento de determinada enfermedad, canales endémicos, estudios de causas de muerte y otros análisis. Algunas veces cuando se realiza un estudio de seguimiento de una enfermedad dada o estudios prospectivos se establecen registros especiales durante un periodo de tiempo, - 5 – 10 años- por un médico o una institución.

En el registro se recogen datos de todos los elementos, si es registro de nacimiento, se registraran a todos los niños nazcan donde nazcan.

Estos procedimientos se diferencian entre sí en una gama de aspectos como son frecuencia de realización, objetivos, la temporalidad y el propósito.

Se pueden resumir los aspectos vistos anteriormente en la siguiente tabla

Característica	Censo	Registro	Encuesta
Frecuencia de realización	Periódica	Continuo Sistemático	Ocasionalmente
Propósito	Universal	Universal	Parcial
Temporalidad	Transversal	Longitudinal	Transversal
Objetivos	Generales	Específicos	Específicos

Errores más comunes en la recogida de información.

Los errores que pueden cometerse en la recolección de información son los mismos que ya estudiamos en las partes precedentes de este texto. Es importante recordar que los errores pueden ser accidentales o sistemáticos y es necesario controlarlos. Los errores sistemáticos deben ser eliminados totalmente y son los más peligrosos pues se pueden producir sin que se percate el que está realizando el trabajo.

Papel del profesional de la salud en la recolección de información.

El médico junto con la enfermera y otros trabajadores de la salud juegan un papel muy importante en la recolección de información. Ellos son los que proporcionan los datos primarios que alimentan a los sistemas de información que se procesan nacionalmente. Un error o una inconsistencia a la hora de suministrar los datos puede tener graves consecuencias.

Esto es sumamente importante que lo entienda el futuro Especialista en Medicina General Integral, que tan importante eslabón juega en la cadena de nuestro sistema de salud.

Igualmente sucede cuando se está recogiendo datos en investigaciones o estudios de Medicina o Salud Pública.

INDICE

Capítulo 3
PROCESAMIENTO DE LA INFORMACION.

Autor: Dr. Emilio Morales Jiménez

Sumario

- *Procesamiento de la información. Etapas que constituyen el procesamiento: organización, resumen y presentación.*
- *Características de la organización y clasificación de los datos.*
- *Concepto de variable.*
- *Clasificación de la información atendiendo al tipo de variable: Cualitativa (nominal y ordinal) y Cuantitativa (discreta y continua)*

Procesamiento de la información. Etapas que constituyen el procesamiento: organización, resumen y presentación.

- El procesamiento de la información es la tercera etapa del método estadístico.
- Incluye varios procederes que permiten la organización, clasificación, resumen y presentación de los datos.
- Se puede hacer manualmente o con la ayuda de computadoras.

Pasos o sub-etapas:

- Organización y clasificación de los datos
- Resumen de los datos
- Presentación de los datos

Importancia

- Comúnmente realizado por estadísticos o bioestadísticos (universitarios o técnicos) y raramente por los médicos y personal de la salud.
- De hecho la mayoría del personal médico no está familiarizado ni está interesado en el procesamiento de datos.
- **Sin embargo, cuando se está realizando un estudio o haciendo una investigación médica es obligado que el personal médico involucrado en el trabajo participe en el procesamiento de manera de obtener los resultados necesarios para el análisis e interpretación (cuarta etapa del método estadístico).**
- Esto asegurará que la organización y clasificación, el resumen y la presentación de los datos se hagan teniendo en cuenta los puntos de vista médicos en relación con los objetivos del estudio o investigación.
- Para cumplir los objetivos médicos del trabajo es necesario establecer un buen trabajo en equipo entre ellos y el personal de la medicina.

Organización y clasificación de los datos

- Primer paso en el procesamiento de datos.
- Los datos deben ser ante todo organizados y clasificados para poder proceder a su procesamiento.
- Constatar que se haya recolectado efectivamente toda la información necesaria y no se haya recolectado datos innecesarios que nunca serán analizados.

Resumen de los datos

- Es el segundo paso.
- Permite mediante cálculos de la estadística descriptiva resumir la información
- Es un paso obligado pues es imposible analizar e interpretar los datos individualmente sino agrupados y resumidos.

Presentación de los datos

- Es el tercer y último paso.
- Obligado para poder mostrar lo que se ha realizado mediante tablas y gráficos estadísticos.

Características de la organización y clasificación de los datos

- ❖ Los datos recogidos deben ser primeramente revisados y si existen errores corregirlos de ser posible.
- ❖ Cuando los errores no pueden ser rectificados es preciso eliminarlos.
- ❖ Se debe tener en cuenta a la hora del resumen y presentación y señalar una categoría de **"sin datos"**.
- ❖ Cuidado con el tamaño de la muestra y por tanto contra lo planeado en el diseño muestral, de ahí la importancia de una atención directa del investigador a la etapa de recolección de datos.
- ❖ Una vez revisados los datos deben ser organizados en forma tal que permitan su resumen en una forma ágil y dinámica.

Si el trabajo se va a realizar manualmente hay varias opciones:

- Sábana de compilación de datos

Ejemplo de sábana de compilación de datos

No. de orden	Edad (años)	Sexo		Hábito de fumar			Número de cigarros
		Masc.	Fem.	Si	No	No respuesta	
1	30	X		X			10
2	41		X		X		-
3	23		X	X			15-20
Etc.							

- Tarjetas simples
  - ⇒ En esta opción se hace una tarjeta para cada elemento colocando los datos en la tarjeta en orden consecutivo.
  - ⇒ Esta opción facilita la clasificación de los elementos por distintas variables al mismo tiempo, seleccionando aquellas tarjetas que cumplan los requisitos acordes al dato de cada variable.
  - ⇒ Tiene el inconveniente que cuando el número de elementos es grande y se requieren gran cantidad de tarjetas se pueden dañar después de varias manipulaciones.

Función similar a las tarjetas se puede lograr con los propios formularios donde se recogió la información.

Si el trabajo se va a realizar por computadoras:

⇒ Se pueden utilizar sistemas de bases de datos o procesadores estadísticos donde se crea también una base de datos para organizar los datos.

⇒ Hay que tener especial cuidado al introducir los datos para no cometer errores.

⇒ Es aconsejable, toda vez que sea posible, utilizar la computación pues:
- ⇒ agiliza el trabajo
- ⇒ evita errores de clasificación y
- ⇒ se obtienen resultados más confiables.

Una vez organizados se procede a clasificar los datos.

- Es importante ya que facilita el procesamiento y análisis ulterior.
- Si tiene datos de diferentes poblaciones de estudio, obviamente numerará los cuestionarios por separado.
- En un **estudio comparativo**, es mejor clasificar los datos inmediatamente que son recogidos en los dos o tres grupos de comparación e inmediatamente numerarlos por separado.
- En un **estudio transversal** puede ser útil clasificar los datos en dos o más grupos dependiendo de los objetivos del estudio.

Una vez organizados y clasificados inicialmente los datos nos enfrentamos a la tarea de cómo hacer entendible el material de que disponemos.

- Para ello debemos **agruparlos según los criterios de clasificación (variables)** que hayamos definido en el diseño experimental (en la planificación de nuestra investigación).

Concepto de variable.

Se define variable como **cualquier cualidad o característica de una persona animal o cosa susceptible de ser medida.**

- Término de medida se utiliza en un amplio sentido.
  - Incluye la clásica medición matemática; por ejemplo, la medida de talla, peso, temperatura y otras.
  - Conteo de algunas características; por ejemplo, frecuencia cardiaca, frecuencia respiratoria y diferencial de sangre entre muchas otras.
  - Identificación de los elementos por una cualidad; por ejemplo, sexo, estado civil, síntomas y diagnóstico entre otras.

- En muchas ocasiones es fácil determinar como medir la variable, como en el sexo y cifra de hemoglobina, por citar algunas.
- En otras no resulta así teniendo dificultades en encontrar la forma de valorarlas, por ejemplo, accesibilidad de un servicio de salud, disponibilidad de recursos, nivel de conocimientos, estado nutricional, evolución de pacientes y respuesta a tratamiento por sólo mencionar las más frecuentes. **Entonces se requiere hacer que las variables sean operacionales mediante el uso de uno o varios indicadores precisos.**

Hacer que una variable sea operacional significa hacer que la variable sea medible.

Por ejemplo, el estado nutricional se puede medir por alguno de los siguientes indicadores: “peso para la edad”; “peso para la talla”; “talla para la edad” y “circunferencia del brazo”.

Una vez que todas las variables sean operacionales, es decir que estén bien definidas y se sepa exactamente como se van a medir, es cuando realmente sabemos que información tenemos que recolectar.

Clasificación de la información atendiendo al tipo de variable: Cualitativa (nominal y ordinal) y Cuantitativa (discreta y continua)

Acorde a la naturaleza de la característica que representa, las variables se clasifican en:

- **categóricas o cualitativas**
- **numéricas o cuantitativas**.

Una variable categórica o cualitativa es **una característica no-numérica por la cual un determinado elemento puede ser diferenciado de otro. O sea, las diferentes cualidades que caracterizan a un elemento.** Por ejemplo: sexo, raza, nacionalidad, síntomas, obesidad, hábitos, color de los ojos, etc. son todas variables categóricas o cualitativas.

Una variable numérica o cuantitativa es **una característica que se expresa por un número.** Por ejemplo: talla, peso, edad, temperatura, número de leucocitos, hemoglobina, glucosa en sangre, frecuencia cardiaca, etc. son todas variables numéricas o cuantitativas. En otras palabras, **las características que diferencian a un elemento de otro por medio de su cantidad es una variable numérica o cuantitativa.**

Escalas de medición

La escala de medición es la **forma en que cada elemento puede ser representado por una unidad de medida dada.**

Cualquier escala de medición tiene que cumplir dos requisitos:

- **Ser exhaustiva** o sea que la clasificación incluya todos los elementos.
- **Tener clases o categorías mutuamente excluyentes**, es decir que cada elemento tiene que ser considerado una sola vez.

Las diferentes escalas de medición utilizadas en Medicina y Salud Pública son:

- **Escalas numéricas o cuantitativas**
  1. Escala continua: entre dos valores cualesquiera puede haber otro valor.
     (Peso, talla, Hb, etc.)
  2. Escala discreta: entre dos valores no puede haber otro valor.
     (Frecuencia respiratoria, pulso, etc.)

- **Escalas categóricas o cualitativas**
  1. Escala nominal: Las categorías no tienen rango entre ellas.
     (Sexo, estado civil, localización dolor, etc.)
  2. Escala ordinal: Las categorías tienen un rango determinado que las identifica.
     (Hábito de fumar, rendimiento académico de un estudiante, grados de una neoplasia, estado de un paciente ingresado, etc.)

Cualquier variable numérica puede ser expresada en escala categórica ordinal; por ejemplo: Hemoglobina: hemoglobina baja (menos de 11 gm/dl);
hemoglobina normal (entre 11 y 13 gm/dl) y
hemoglobina alta (mas de 13 gm/dl)

Este último aspecto es muy importante pues es como más se usan las escalas en Medicina.

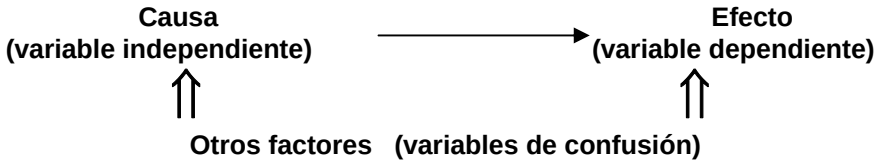
Otras clasificaciones de variables que por su utilización en los estudios e investigaciones médicas y de Salud Pública se deben conocer.

- **Variable dependiente y variable independiente**
  Cuando en un estudio se busca una explicación causal es importante distinguir entre variables dependientes y variables independientes.
  - La variable **dependiente** es aquella que se usa **para describir o medir el problema** objeto de estudio.
  - Las variables **independientes** son aquellas que se usan **para describir o medir los factores contribuyentes o asociados que presumiblemente causan o influencian en el problema**.

  El que una variable sea dependiente o independiente depende del planteamiento del problema y de los objetivos. Aunque normalmente hablamos de **CAUSAS** del problema, en lenguaje científico debemos hablar de **ASOCIACION** entre variables, a menos que se pueda probar una relación causal. Por ejemplo, si encontramos una asociación entre el hábito de fumar y el cáncer, sólo podremos concluir que el fumar causa el cáncer cuando demostremos que (1) el cáncer fue desarrollado después que el paciente comenzó a fumar y (2) no existe otro factor que puede haber causado ambos, el cáncer y el hábito de fumar.

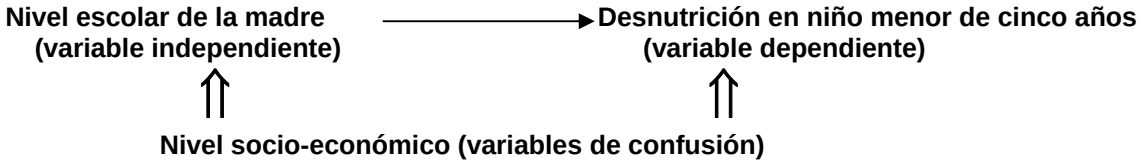
- **Variable de confusión**

  Una variable que es asociada con el problema y con una posible causa del problema es potencialmente una **VARIABLE DE CONFUSIÓN.**
  Una variable de confusión puede reforzar o debilitar una relación aparente entre el problema y una posible causa.



Por tanto, cuando se esté haciendo un estudio explicativo es muy importante considerar las posibles variables de confusión para determinar como se hará la planificación de la investigación y el análisis de los datos.

Por ejemplo: Se muestra una asociación entre bajo nivel escolar de la madre y desnutrición en niños menores de cinco años. Sin embargo, el nivel socio-económico puede estar relacionado al nivel escolar de la madre así como al estado de nutrición del niño.



El nivel socio-económico es potencialmente una variable de confusión. Para dar una idea real de la relación entre el nivel escolar de la madre y la desnutrición, debe ser considerado y medido el ingreso económico familiar. Esto puede ser incorporado en el diseño de la investigación, seleccionando solamente madres con determinado nivel de ingresos o tenerlo en consideración durante el análisis de los resultados haciendo el análisis del nivel educacional y la desnutrición por categorías de nivel socio-económico.

- **Variables de antecedentes**
  En la mayoría de las investigaciones, aparecen las **VARIABLES DE ANTECEDENTES**, tales como edad, sexo, nivel educacional, estado socio-económico, estado civil y religión. Estas variables de antecedentes están relacionadas frecuentemente con variables independientes de manera que influyen indirectamente en el problema. Si estas variables son importantes en el estudio deben ser medidas y tenidas en consideración. Las variables de antecedentes son notoriamente variables de confusión.

Para tener una idea de la importancia de definir adecuadamente una variable pondremos ejemplos de posibles alternativas cuando se está tratando de definir variables de uso frecuente. **(Tomado del libro de "Diseño y Conducción de Proyectos de Investigación en Sistemas de Salud" de la OMS)**

Ocupación	¿Ocupación para la que el sujeto fue entrenado, o trabajo que desarrolla actualmente? Si el sujeto está retirado o desempleado, ¿se utilizará la ocupación previa?
Educación	¿Número de años de educación, o último grado aprobado, o tipo de institución educacional a la que último asistió?

Ingreso económico	¿Ingreso personal, ingreso en la familia, o ingreso per cápita en la familia?
Hacinamiento	(Número medio de personas por cuarto en la casa). ¿Qué habitaciones se excluyen del índice (baños, cocina, sala, comedor, hall)?
Clase social	¿Basado en la ocupación, educación, hacinamiento, ingreso, barrio de residencia, comodidades de la casa, o la propia auto - percepción del sujeto? ¿Basado en uno de éstos o en una combinación de ellos?
Estado civil	¿En términos del status legal (soltero, casado, divorciado o viudo) o en términos de estabilidad (unión estable, unión casual)?
Paridad	¿Número total de embarazos previos o número total de partos?
Fecha de comienzo de enfermedad	¿Fecha en que el primer síntoma apareció, o fecha cuando fue primeramente diagnosticada, o fecha de la notificación?
Presencia de enfermedad crónica	¿Basada en la duración desde el comienzo? Si es así, ¿Que duración hace que se considere crónica – 3 meses, 6 meses, 1año? ¿Basada en la presencia de determinada enfermedad? ¿Se definen como crónicas algunas enfermedades independientemente de su duración? Si es así, ¿Cuales enfermedades? ¿Cómo considerar condiciones recurrentes (por ejemplo, faringitis frecuentes)?

INDICE

Capítulo 4
DISTRIBUCION DE FRECUENCIAS

Autor: Dr. Emilio Morales Jiménez

SUMARIO

- *Distribuciones de frecuencia según el tipo de variable.*
- *Concepto de clase o intervalo. Límite inferior y superior de la clase. Punto medio de la clase. Rango. Amplitud de clase.*
- *Frecuencia absoluta, relativa y acumulada.*

Distribuciones de frecuencia según el tipo de variable.

La distribución de frecuencias se define como la **agrupación o distribución de los datos en diferentes clases o categorías acorde a una escala dada.** Luego habrá distribuciones de frecuencias en escala categórica o cualitativa, en escala numérica o cuantitativa discreta y en escala numérica cuantitativa continua, las cuales veremos a continuación. En algunos libros de Estadística se considera la distribución de frecuencias como un método de clasificación de datos en otros se describe como un proceder para resumir datos. En nuestra opinión tiene ambas funciones, resumir y clasificar datos.

Concepto de clase o intervalo. Límite inferior y superior de la clase. Punto medio de la clase. Rango. Amplitud de clase.

Antes de seguir adelante es imprescindible definir algunos conceptos para poder entender como se realiza una distribución de frecuencia.

Clase o intervalo de clase

Clase o categoría se define como la **característica por la cual los datos de una variable son agrupados.**

En una variable categórica o cualitativa como el sexo las clases fueran: masculino y femenino; para el hábito de fumar las clases pudieran ser: no fumador, fumador severo, fumador moderado y fumador leve. En estos casos las características son las diferentes posibilidades o categorías que la variable tiene acorde con la escala de medición usada.

En la variable numérica o cuantitativa la característica es diferente. En una variable numérica discreta, por ejemplo el número de niños de una familia, las clases pueden ser: 0, 1, 2, 3, 4, 5, 6,.., o sea cada posible valor de la variable. En este caso es semejante a la variable categórica siendo cada número una **"categoría"**. Sin embargo si la variable tiene un rango amplio, digamos que el número de hijos es de 0 a 30, entonces es más conveniente hacer un arreglo utilizando otro criterio de agrupación. Ese criterio de agrupación es el que se usa para variables numéricas continuas. En estos casos las clases están formadas por lo que se llama **"intervalo de clase"**. Por ejemplo, suponga que estamos trabajando con la variable talla, medida en cm. Los intervalos de clase pueden ser: 140-144; 145-149; 150-154; 155-159;..... y sucesivamente. En el ejemplo

previo del número de niños pudieran ser: 0-3; 4-7; 8-11;.... y sucesivamente.

Aunque los términos de clase e intervalos de clase se pueden usar indistintamente, "**el intervalo de clase se puede definir como la forma en que las características de los datos numéricos son usadas con fines de agrupamiento"**.

El número de clases o intervalos de clases que se utilizan en una distribución de frecuencias depende de las características de los datos. Algunos autores dicen que el número de clases debe estar entre 5 y 20, otros que entre 4 y 12 y algunos prefieren que no excedan de 10. Nuestra opinión es que el número de clases no debe pasar de 12 dependiendo de los datos. No obstante el número de clases siempre es determinado por la persona que realiza la distribución de frecuencias.

En los intervalos de clase, (por ejemplo 140-144), los números se denominan **"límites de clase",** el número menor, **140, es el límite inferior** y el mayor, **144, es el límite superior.** En este caso como el intervalo de clase tiene limites inferior y superior se denomina **"intervalo de clase cerrado"**. El intervalo de clase que, al menos teóricamente, no tiene límite inferior o límite superior indicado, por ejemplo **"65 años y más"** se le llama **"intervalo de clase abierto"**.

Es importante conocer la **"frontera de clase"** o **límite real de la clase** cuando se trabaja con variable numérica continua. Por ejemplo supongamos que tenemos los siguientes intervalos de clase de la talla:

Talla (cms)	Lo que se interpreta normalmente hoy día es que los valores, por ejemplo, 144.3; 144.4 .... 144.999999 se incluyen en el intervalo de 140-144 y así sucesivamente. El límite real o frontera de clase está dado en este caso por 144.9999999999999.... en la superior y 139.9999999999.. en la inferior.
140-144	
145-149	
150-154	
155-159	
160-164	
165-169	

En el caso particular de variable numérica discreta, el límite de clase coincide siempre con la frontera de la clase.

Cuando se utilizan sistemas estadísticos automatizados se utiliza una notación distinta dejando claros las fronteras y límites de clases. Por ejemplo el sistema MICROSTA establece lo siguiente para distribuciones de frecuencia:

Talla (cms)	
140<145	Cada intervalo termina en < que un número, luego, por ejemplo, el primer intervalo termina en 144.999999999999.... y comenzaría en 140 y así sucesivamente.
145<150	
150<155	
155<160	
160<165	
165<170	

El sistema EXCELL al calcular distribuciones de frecuencia utiliza otro sistema:

Talla (cms)	
>140-144	Cada intervalo comienza en > que un número, luego, por ejemplo, el primer intervalo termina en 144 exacto y el que le sigue comenzaría en 144 con cualquier decimal. Es decir, 144.01, 144.1, 144.9999 se incluyen en el segundo intervalo el cual termina en 149 exacto y así sucesivamente.
>144-149	
>149-154	
>154-159	
>159-164	
>164-169	

En la mayoría de los trabajos científicos que se presentan en las revistas médicas las distribuciones de frecuencias se presentan en la primera forma descrita, sino se especifica otra cosa.

Rango
Rango es el **recorrido que tienen los datos**, es decir, **la diferencia entre el valor máximo y el valor mínimo de los datos**.

Longitud o amplitud de clase
La longitud o amplitud de clase es **la diferencia entre la frontera superior y la frontera inferior de la clase.** Los límites de la clase están incluidos en la amplitud de clase. Comúnmente las distribuciones de frecuencia tienen los intervalos de clase con la misma amplitud. Sin embargo en Medicina y Salud Pública algunas veces los intervalos de clase se hacen con diferentes amplitudes de clase, sobre todo cuando se trabaja la variable edad.

La amplitud de clase se calcula de la siguiente forma:

amplitud de clase = rango
número o cantidad de intervalos de clases deseados

o sea

amplitud de clase = valor máximo de los datos – valor mínimo de los datos
número o cantidad de intervalos de clases deseados

- Si el resultado de la división es un entero o un entero con decimal, la amplitud de clase debe ser el entero que le sigue. Por ejemplo, si el resultado de la división es 3 o 3.1 o 3.6, la amplitud de clase debe ser 4
- Si el resultado de la división es un decimal, la amplitud de clase debe ser el siguiente decimal superior. Por ejemplo, si el resultado de la división es 0.73 o 0.7 o 0.76, la amplitud de clase debe ser 0.8.

Punto medio de la clase

El punto medio de clase o marca de la clase es **la suma del límite superior e inferior dividido por dos.** Así la marca de clase para el intervalo 140-144 será (140+144)/2 = 142. La marca de clase es el **valor que representa al intervalo de clase en datos agrupados**.

Frecuencia absoluta, relativa y acumulada.

Frecuencia absoluta

La frecuencia absoluta, frecuencia o frecuencia de clase es **el número de observaciones que pertenecen a cada clase o en el intervalo de clase.** En otras palabras, es el número de elementos que tienen la característica expresada en la clase o el número de veces que aparecen los datos en cada clase o el número de ocurrencias en cada clase. La frecuencia absoluta se obtiene por simple conteo de los elementos.

Frecuencia relativa

La frecuencia relativa de una clase es **la frecuencia absoluta dividido entre el número total de observaciones o sea el número total de elementos.**

frecuencia relativa de la clase = frecuencia absoluta de la clase
número total de elementos

La frecuencia relativa es la medida del **"peso"** que cada frecuencia absoluta tiene con relación al número total de observaciones. Este concepto será retomado en estadística inferencial cuando se estudie probabilidades.

Si la frecuencia relativa se multiplica por 100, entonces se expresa como un porciento. La suma de las frecuencias relativas de todas las clases es 1 o 100% en dependencia de como se haya expresado.

Frecuencia acumulada

La frecuencia acumulada es **la suma de las frecuencias hasta una clase o intervalo de clase dado.** La frecuencia acumulada se calcula tanto con la frecuencia absoluta como

con la relativa. Como es fácil de entender, la frecuencia acumulada del último intervalo será la frecuencia total de todas las clases o sea el total de los elementos u observaciones.

Reglas generales para hacer una distribución de frecuencias

1. Identifique el tipo de variable acorde con la escala de medición empleada.
2. Seleccione las clases o calcule los intervalos de clase según los datos de la variable.
3. Distribuya cada elemento en la clase correspondiente.
4. Cuente el número de elementos en cada clase.
5. Calcule las frecuencias relativa y acumulada.

Veamos algunos ejemplos. Suponga que se ha realizado un estudio y en la muestra de 20 elementos se tomaron los datos que se presentan en la siguiente tabla.

Número del elemento	Sexo	Número de hijos	Talla en cms.
1	M	2	151.3
2	M	1	164.2
3	F	0	149
4	M	0	168.3
5	F	3	167.2
6	M	4	180
7	M	3	176.4
8	M	6	151
9	F	5	145.1
10	M	4	157.2
11	M	3	162.6
12	F	2	153.3
13	F	1	148.2
14	F	0	153.8
15	M	1	157
16	M	2	155
17	F	3	159.3
18	M	4	171.3
19	M	6	172.4
20	F	5	168

Se requiere hacer una distribución de frecuencias de cada variable.

Distribución de frecuencia del sexo.

- El sexo es una variable categórica o cualitativa nominal.

- Las clases solamente pueden ser dos categorías: masculino o femenino.

- Distribución de los elementos en las correspondientes clases.

Clases	Distribución de elementos
Masculino	IIIII IIIII II
Femenino	IIIII III

- Conteo del número de elementos en cada clase.

Clases	Frecuencia absoluta
Masculino	12
Femenino	8
Total	20

- Calculo de la frecuencia relativa y el porciento.

Clases	Frecuencia absoluta	Frecuencia relativa	Porciento
Masculino	12	0.6	60.0
Femenino	8	0.4	40.0
Total	20	1.0	100.0

Note que en la variable categórica nominal no se calculan frecuencias acumuladas. En las variables categóricas ordinales si se pueden calcular.

La distribución de frecuencia se presenta en forma de tabla estadística de la siguiente manera.

Pacientes según sexo		
Sexo	**Cantidad**	**Porciento**
Masculino	12	60.0
Femenino	8	40.0
Total	20	100.0

Por lo general sólo se emplea la frecuencia absoluta y la frecuencia relativa o el porciento.

Distribución de frecuencia del número de hijos

- El número de hijos es una variable numérica o cuantitativa discreta.

- Las clases pueden ser los valores que toma la variable (si no son muchos), o intervalos de clase (si los valores tienen un rango amplio). Como en el ejemplo los posibles valores son del 0 al 6, consideraremos cada valor como una clase.

- Distribución y conteo de los elementos.

Clases **Número de hijos**	**Distribución de elementos**	**Conteo o Frecuencia Absoluta**
0	I I I	3
1	I I I	3
2	I I I	3
3	I I I I	4
4	I I I	3
5	I I	2
6	I I	2
Total		20

Note que el número total de la frecuencia absoluta es el total de elementos de la muestra. Este número total es utilizado para chequear que todos los elementos hayan sido incluidos en la distribución. Si el total no coincide con el número de elementos de la muestra es que hubo algún error y hay que repetir la distribución.

- Cálculo de la frecuencia relativa, porciento y la frecuencia acumulada.

Número de hijos	**Frecuencia absoluta**	**Frecuencia relativa**	**Porciento**	**Frecuencia Acumulada**		
				Absoluta	**Relativa**	**Porciento**
0	3	0.15	15.0	3	0.15	15.0
1	3	0.15	15.0	6	0.30	30.0
2	3	0.15	15.0	9	0.45	45.0
3	4	0.20	20.0	13	0.65	65.0
4	3	0.15	15.0	16	0.80	80.0
5	2	0.10	10.0	18	0.90	90.0
6	2	0.10	10.0	20	1	100.0
Total	20	1	100.0	-------	-------	--------

Note que en las columnas de frecuencias acumuladas la fila del total está en blanco porque por definición está ya totalizado en la última clase. El uso de la frecuencia

acumulada es para saber a un golpe de vista algunas características de la variable. Por ejemplo, se puede saber de una forma rápida que el 65% de los elementos tienen 3 hijos o menos siendo 13 los elementos con esa característica.

Presentación en forma de tabla estadística.

Pacientes según número de hijos			
Número de hijos	**Cantidad**	**Porciento**	**Porciento acumulado**
0	3	15.0	15.0
1	3	15.0	30.0
2	3	15.0	45.0
3	4	20.0	65.0
4	3	15.0	80.0
5	2	10.0	90.0
6	2	10.0	100.0
Total	20	100.0	--------

Distribución de frecuencia de la talla.

- La talla es una variable numérica continua.

- Al ser una variable continua hay que calcular el intervalo de clase. Hagamos la distribución de frecuencias con 4 intervalos de clase. Para calcular la amplitud de clase necesitamos encontrar el valor máximo y el valor mínimo de los datos. Ellos son 180 y 145 respectivamente. Luego la amplitud de clase será:

$$\text{amplitud de clase} = \frac{\text{valor máximo} - \text{valor mínimo}}{\text{número de intervalos de clases}}$$

$$\text{amplitud de clase} = \frac{180 - 145}{4} = \frac{35}{4} = 8.7$$ Se aproxima a 9.

amplitud de clase = 9

Los intervalos de clases serán:

145 –153

154 – 162

163 – 171

172 – 180

Note que los límites de cada intervalo de clase se incluyen en la amplitud de clase. En efecto de 145 a 153 hay 9 valores.

- Distribución y conteo de los elementos.

Clases Talla en cms.	Distribución de elementos	Conteo o Frecuencia Absoluta
145 –153	I I I I I I I	7
154 – 162	I I I I I	5
163 – 171	I I I I I	5
172 – 180	I I I	3
Total		20

- Cálculo de la frecuencia relativa y acumulada. (Como se usan intervalos de clase se incluye en la tabla una columna para la marca de clase).

Talla (cms)	Marca de clase	Frecuencia absoluta	Frecuencia relativa	Porciento	Frecuencia acumulada		
					Absoluta	Relativa	Porciento
145 – 153	149	7	0.35	35.0	7	0.35	35.0
154 – 162	158	5	0.25	25.0	12	0.60	60.0
163 – 171	167	5	0.25	25.0	17	0.85	85.0
172 – 180	176	3	0.15	15.0	20	1	100.0
Total		20	1	100.0	---	---	---

Note que la marca de clase es la suma de los límites de clase dividido por 2; así para el primer intervalo de clase fuera (145 + 153)/2 = 298/2 = 149. Sucesivamente se calculan todas las marcas de clase o se le suma la amplitud de clase a la precedente obteniendo las demás. Esto es solamente válido cuando los intervalos de clases tienen todos la misma amplitud.

La presentación de los datos se hace mediante una tabla estadística semejante a las ya mostradas.

Las distribuciones de frecuencia son muy empleadas en Medicina y en Salud Pública. De hecho todos los trabajos científicos e investigaciones tienen distribuciones de frecuencia. Cuando se están presentando los resultados de un trabajo científico o investigación, por lo general, solamente se utilizan dos columnas además de las clases: la frecuencia absoluta y el porciento. La frecuencia acumulada se utiliza en trabajos muy específicos al igual que la frecuencia relativa.

INDICE

Capítulo 5
RESUMEN DE LA INFORMACIÓN

Autor: Dr. Emilio Morales Jiménez

SUMARIO

- *Resumen de la información*
- *Medidas de tendencia central (media aritmética, mediana y moda). Características, propiedades, cálculo e interpretación.*
- *Medidas de dispersión. Recorrido de la variable o amplitud total. Varianza, desviación estándar y coeficiente de variación. Características, propiedades, cálculo e interpretación. Variabilidad en Medicina y Biología.*
- *Medidas de posición relativa (percentiles). Características, cálculo e interpretación. Su utilidad en los estudios de crecimiento y desarrollo.*
- *Medidas para resumir datos cualitativos (razón, índice, proporción, porciento y tasa). Forma de cálculo e interpretación. Ejemplos en el campo médico.*

Resumen de la información

El resumen de la información es una parte importante de la etapa de procesamiento de datos. Aunque en el capítulo anterior, se consideró la distribución de frecuencias como un procedimiento para resumir datos, el mismo no da un valor que por si mismo represente a todo el conjunto de datos. Eso se estudiará en este capítulo. La estadística descriptiva incluye muchos procederes para resumir datos, los cuales son diferentes en dependencia si la caracterización es para variables numéricas o variables categóricas. Solamente abordaremos las mas frecuentemente utilizadas en Medicina y Salud Pública y que estén incluidas en el programa de estudio del estudiante de Medicina.

Se dividirá el estudio en:

- Medidas para resumir datos numéricos o cuantitativos
- Medidas para resumir datos categóricos u cualitativos

Los procederes relacionados con variables numéricas se corresponden con tres características principales de cualquier conjunto de datos numéricos:

- la tendencia central de los datos
- la dispersión de los datos y
- la posición relativa de los elementos en un conjunto ordenado de datos.

Los procederes relacionados con las variables categóricas se corresponden con las principales características de los datos cualitativos:

- la magnitud de la cualidad en un conjunto de datos
- la magnitud de la relación de una categoría con otra y
- la magnitud de una categoría en una muestra, en la población o con otras categorías.

Resumen de variables numéricas o cuantitativas
Se estudiaran las medidas de tendencia central, las medidas de dispersión y las medidas de posición relativa.

Medidas de tendencia central (media aritmética, mediana y moda). Características, propiedades, cálculo e interpretación.

Se define como medida de tendencia central **al *"estadígrafo"* que caracteriza a un conjunto de datos y que es calculado para la obtención de un valor que ocupe una posición central o intermedia entre los valores máximo y mínimo.** Se estudiarán tres de estas medidas: media aritmética, mediana y moda.

- **La media aritmética**

La media aritmética, también llamada solamente media o promedio, **se define como la suma de los valores de cada elemento de un conjunto de datos dividida entre el número total de elementos del conjunto.** La media se denota por $\bar{x}$ y se expresa matemáticamente como:

$$\bar{x} = \frac{X_1 + X_2 + X_3 + ...+ X_n}{N}$$

donde: X_1, X_2 .....X_n son los valores de cada dato en el conjunto, y

N es el número total de elementos en el conjunto.

La fórmula se escribe usualmente utilizando el símbolo de sumatoria **(Σ)** de la siguiente forma:

$$\bar{x} = \frac{\Sigma \mathbf{x}}{N}$$

(Al ser este texto para estudiantes de medicina y personal de la salud hemos obviado la notación en la sumatoria de i = 1 hasta n y creemos que con la explicación que se dio antes de la fórmula se sobreentiende que **x** representa el valor de cada elemento. Así se hará en todas las fórmulas.)

Por ejemplo si los valores de urea en 5 pacientes son:

17.5; 20; 23.5; 15; y 19 mg/dl.,

la media será:

$$\bar{x} = \frac{17.5 + 20 + 23.5 + 15 + 19}{5} = \frac{95}{5} = 19 \text{ mg/dl.}$$

Propiedades de la media

- La media siempre existe y siempre puede ser calculada para cualquier conjunto de datos numéricos.

- La media es única; cada conjunto de datos tiene una y sólo una media.

- La media, al considerar todos los valores del conjunto de datos, es altamente influenciada por valores extremos.

  En efecto, si al ejemplo anterior le agregamos el valor de un sexto paciente que tiene 52 mg/dl de urea, al calcular la media tendremos que:

$$\bar{x} = \frac{17.5 + 20 + 23.5 + 15 + 19 + 52}{6} = \frac{147}{6} = 24.5 \text{ mg/dl.}$$

$$\bar{x} = 24.5 \text{ mg/dl.}$$

  Este valor no representa realmente la tendencia central del conjunto. Note que de los 6 valores de urea promediados, 5 de ellos están por debajo de la media. Este es el inconveniente que tiene la media aritmética y cuando en el conjunto de datos hay valores extremos es preferible utilizar otra medida de tendencia central.

- La suma algebraica de las desviaciones del valor de cada elemento del conjunto con respecto a la media es cero.

  Por ejemplo, en el cálculo anterior donde la media fue 24.5, las desviaciones de los valores individuales son:

(24.5 – 17.5)	= + 7.0
(24.5 – 20.0)	= + 4.5
(24.5 - 23.5)	= + 1.0
(24.5 – 15.0)	= + 9.5
(24.5 – 19.0)	= + 5.5
(24.5 – 52.0)	= -27.5
Suma algebraica	0

- **La mediana**

La mediana se define como **el valor que divide al *conjunto ordenado de datos* en dos partes iguales.**

El valor de la mediana es el punto medio del conjunto ordenado de datos; por debajo del valor de la mediana se encuentran los valores del 50% de los elementos al igual que por encima los valores del otro 50%.

Para calcular una mediana, antes es obligado ordenar el conjunto de datos según los valores de cada elemento, ya bien sea en orden ascendente o en orden descendente.

Por ejemplo, a 7 pacientes se le determina el valor de proteínas totales siendo:

5; 6; 6; 8; 7; 5 y 7 mg/dl.

Para calcular la mediana es necesario primero que todo, ordenar los datos. Los datos ordenados en orden ascendente fueran:

5 5 6 6 7 7 8

Es fácil encontrar el punto medio cuando el conjunto de datos es de pocos números. En este caso la mediana es 6 mg/dl, (la 4ta. posición). En efecto por encima de esa posición hay tres valores (7; 7; 8) y por debajo de ese valor también hay tres valores (6; 5; 5).

Una regla general para el cálculo de la posición de la mediana si el número de elementos del conjunto es impar, **una vez ordenados los datos**, se encuentra por:

$$(N+1)/2$$

Esto es, el número total de elementos del conjunto más 1 dividido entre 2.

En el ejemplo, la posición de la mediana fuera: (7+1)/2 = 8/2 = 4.; y

Posición	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Valor de los elementos	5	5	6	6	7	7	8

El valor de la mediana es el valor del elemento que ocupa la cuarta posición, que es 6 mg/dl.

Si el número de elementos es par, entonces la mediana es la media aritmética o semisuma de los dos valores centrales del conjunto ordenado de datos. La posición de los dos valores centrales se encuentra por:

$$(N/2) \quad y \quad (N/2 + 1).$$

Por ejemplo, se tiene el valor de proteínas totales en 16 pacientes:

5.8; 6; 4.1; 7.3; 8; 6.6; 7.9; 8.4; 6.5; 6; 7.1; 5.3; 5; 6.2; 7.8 y 8 mg/dl.

Primero los datos tienen que ser ordenados:

Posición de ordenamiento:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Valor de cada elemento:	4.9	5	5.3	5.8	6	6	6.2	6.5

Posición de ordenamiento:	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
Valor de cada elemento:	6.6	7.1	7.3	7.8	7.9	8	8	8.4

Los valores centrales tendrán las posiciones:

$(N/2)= 16/2= 8$ y $(N/2+1)= 16/2 +1= 8+1=9$

Note que los valores 8 y 9 son de las posiciones centrales no del valor de la urea. El valor de la posición 8 es 6.5 mg/dl y el de la posición 9 es 6.6 mg/dl. La mediana será la media aritmética de los dos valores centrales, a saber:

$\frac{(6.5+6.6)}{2} = \frac{13.1}{2} = 6.55$ mg/dl. La mediana = 6.55 mg/dl.

Note que en el conjunto ordenado de datos, el valor 6.55 divide al conjunto en 8 valores por debajo y en 8 por encima.

Propiedades de la mediana

- La mediana siempre existe y siempre puede calcularse para cualquier conjunto de datos numéricos.

- La mediana es única; cada conjunto de datos numéricos tiene una y sólo una mediana.

- La mediana no es influenciada por valores extremos, siendo la medida de tendencia central adecuada cuando en el conjunto de datos existen valores extremos.
  Por ejemplo, veamos el cálculo de la mediana para el caso que se mostró en las medias (vea párrafos precedentes en propiedades de la media). La mediana de urea sería:

Posición:	**1**	**2**	**3**	**4**	**5**	**6**
Valores de urea:	15	17.5	19	20	23.5	52

Al haber 6 elementos, la posición de los valores centrales fuera:

$$6/2 = 3 \qquad y \qquad 6/2 + 1 = 3 + 1 = 4$$

Los valores de las posiciones centrales son 19 y 20 mg/dl, luego la mediana es:

$$\frac{(19 + 20)}{2} = \frac{39}{2} = 19.5 \text{ mg/dl.}$$

Vea que en este caso la mediana, 19.5 mg/dl, es más característica como medida central del conjunto de datos que la media (24.5 mg/dl).

- La mediana se calcula con pocas operaciones matemáticas.

- Es obligado hacer un ordenamiento del conjunto de datos antes de calcular la mediana.

- **La moda**

La moda se define como **el valor que ocurre con mayor frecuencia en un conjunto de datos.** En otras palabras, es el valor que más se repite.

Para calcular la moda se deben ordenar los datos primero, bien en orden ascendente o descendente.

Por ejemplo:

Conjunto de datos A: 2; 2; 5; 9; 9; 9; 9; 10; 11; 13. La moda es 9.
Conjunto de datos B: 0; 1; 2; 3; 4; 5; 14. No hay moda.
Conjunto de datos C: 1; 2; 4; 4; 4; 5; 7; 7; 7; 8; 8; 8; 9; 9; 9. La moda es 4, 7, 8 y 9.

Propiedades de la moda

- La moda puede no existir en un conjunto de datos.

- La moda puede ser más de una.

- La moda puede usarse en datos categóricos o cualitativos.

De las tres medidas de tendencia central explicadas, la más utilizada en Medicina y Salud Pública es la media aritmética. Sin embargo, en muchas ocasiones existen valores extremos en un conjunto de datos y en lugar de la media se debe utilizar la mediana. El médico, el estudiante de Medicina y el trabajador de la salud tiene que darse cuenta que el objetivo de calcular una medida de tendencia central es obtener un valor central de los datos. Si la media no cumple ese objetivo debe utilizarse otra medida que es la mediana. La moda se utiliza raramente en Medicina.

Cuando en un conjunto de datos la media, la mediana y la moda tienen el mismo valor es indicación de que la curva que representa a la distribución de frecuencias de los datos es simétrica, como por ejemplo la curva normal que se estudiará mas adelante.

Medidas de dispersión. Recorrido de la variable o amplitud total. Varianza, desviación estándar y coeficiente de variación. Características, propiedades, cálculo e interpretación. Variabilidad en Medicina y Biología.

Las medidas de dispersión **se definen como las medidas que representan el grado al cual tienden a dispersarse los datos numéricos con relación a un valor de tendencia central.**

La dispersión o variación de los datos es otra característica a tener en consideración cuando se describen datos numéricos. Las medidas de tendencia central por si solas no brindan suficiente información sobre un conjunto de datos. Tomemos, por ejemplo dos grupos de 5 pacientes y a cada paciente se le determina el nivel de glucosa en sangre con los siguientes resultados:

Paciente	Grupo 1	Grupo 2
No. 1	2.0	3.8
No. 2	3.0	3.9
No. 3	4.0	4.0
No. 4	5.0	4.1
No. 5	6.0	4.1
$\overline{X}$	**4.0**	**4.0**

La media aritmética es igual en ambos grupos, 4.0 g/l. Sin embargo los grupos son diferentes en la variación o dispersión de los datos. El Grupo 1 tiene una variación mayor, en efecto entre 2.0 y 6.0 mientras que el segundo grupo tiene una pequeña variación, entre 3.8 y 4.2.

El conocimiento de la dispersión de los datos es necesario en Medicina y en Salud Pública porque es importante para el análisis e interpretación de muchos resultados en estudios e investigaciones médicas. Por ejemplo, para evaluar la efectividad de un nuevo proceder terapéutico, es necesario considerar la dispersión de las variables que se utilizan en la evaluación.

Hay diferentes medidas de dispersión. Nos ocuparemos de las más comunes: el rango también llamado recorrido o amplitud total, la varianza y la desviación estándar.

- **El rango (recorrido o amplitud total)**

El rango de un conjunto de datos es **la diferencia entre el valor mayor y el valor menor de un conjunto de datos.** En el ejemplo previo de glucosa en sangre fuera:

Para el primer grupo: 6.0 – 2.0 = 4.0 g/l
Para el segundo grupo: 4.2 – 3.8 = 0.4 g/l

El primer grupo tiene un rango 10 veces superior al segundo. Los datos están mucho más dispersos.

El rango es la medida de dispersión más sencilla, pero tiene falta de especificidad en algunos casos. Tomemos, por ejemplo tres grupos de 7 niños por debajo de 15 años de edad siendo la edad de cada uno la siguiente:

Grupo A: 1; 3; 5; 7; 9; 11; y 13.
Grupo B: 1; 2; 3; 7; 11; 12; y 13.
Grupo C: 1; 5; 6; 7; 8; 9; y 13.

Para ver la dispersión o variación de los datos hagamos una representación tabular:

Grupo	Edad en años												
	1	2	3	4	5	6	7	8	9	10	11	12	13
A	x		x		x		x		x		x		x
B	x	x	x				x				x	x	x
C	x				x	x	x	x	x				x

El rango es el mismo en los tres grupos, (13 – 1) = 12 años, pero la variación de los datos es muy distinta en los tres grupos. El primer grupo tiene las edades uniformemente distribuidas, el segundo agrupadas a los extremos y el tercero agrupadas al centro. Por esta razón el rango hay que usarlo con cautela y en circunstancias adecuadas.

- **La varianza**

La varianza de un conjunto de datos es definida como la sumatoria del cuadrado de las desviaciones de cada elemento con la media aritmética, dividido por el número total de elementos menos 1. La varianza se denota por S^2 y se calcula por la fórmula:

$$S^2 = \frac{\Sigma (X - \bar{X})^2}{N-1}$$

donde $(X - \bar{X})^2$ es el cuadrado de la diferencia de cada elemento con la media, y
$N-1$ es el total de elementos del conjunto menos 1

Para calcular la varianza se puede utilizar un medio bastante práctico que se

ejemplifica a continuación:

- Cálculo de la varianza de glucosa en sangre. (Ejemplo ya mencionado)

G R U P O 1		
X	**X - $\bar{x}$**	**$(X - \bar{x})^2$**
2.0	- 2.0	4.0
3.0	- 1.0	1.0
4.0	0	0
5.0	+ 1.0	1.0
6.0	+ 2.0	4.0
	$\sum$	10.0
	S^2 = 10.0/4	**= 2.5 g^2/l^2**

($\bar{x}$ = 4.0)

G R U P O 2		
X	**X - $\bar{x}$**	**$(X - \bar{x})^2$**
3.8	- 0.2	0.04
3.9	- 0.1	0.01
4.0	0	0
4.1	+ 0.1	0.01
4.2	+ 0.2	0.04
	$\sum$	0.10
	S^2 = 0.10/4	**= 0.025 g^2/l^2**

La varianza tiene el inconveniente de que al ser calculada con valores elevados al cuadrado, la unidad de medida también está elevada al cuadrado, dando valores grandes.

- **La desviación estándar**

La desviación estándar es **definida como la raíz cuadrada de la varianza.** Se denota por S y se calcula por la fórmula:

$$S = \sqrt{S^2} = \sqrt{[\sum (X - \bar{x})^2 / N-1]}$$

Para el cálculo de la desviación estándar primero hay que calcular la varianza. En el ejemplo de la glucosa en sangre fuera:

G R U P O 1	**G R U P O 2**
Varianza (**S^2**)=2.5 g^2/l^2.	Varianza (**S^2**)=0.025 g^2/l^2.
S = $\sqrt{2.5}$ = 1.58 g/l.	**S** = $\sqrt{0.025}$ = 0.158 g/l.

La desviación estándar mide la dispersión en la misma unidad de medida de la variable, dando valores mas adecuados a los datos. La desviación estándar es la medida de dispersión más utilizada como se verá en estadística inferencial.

La varianza y la desviación estándar tienen las propiedades comunes siguientes:

- Siempre son un número positivo.

- A mayor valor del estadígrafo, mayor es la dispersión de los datos en relación con la media.

- Si el valor es 0, significa que no hay dispersión y solamente puede ser posible si todos los datos tienen el mismo valor.

En Medicina y en Salud Pública las medidas de dispersión más usadas son la varianza y la desviación estándar. Se necesitan específicamente cuando se realizan procedimientos de la estadística inferencial. **Ambas medidas pueden también ser calculadas con la mediana como medida de tendencia central.**

Sin embargo, en nuestro campo de acción, muy frecuentemente es necesario comparar la dispersión o variación en dos conjuntos de datos con la misma o diferentes variables. En estos casos al usar las medidas de dispersión ya descritas suceden dos problemas:

1. Si las variables son diferentes se cuantifican por diferentes medidas haciendo imposible la comparación.
2. Si la unidad de medida es la misma, (una variable en dos conjuntos de datos), la medida de dispersión es influenciada por el valor de los datos. Por ejemplo el peso expresado en kilogramos de personas adultas es mucho mayor que el peso de niños por debajo de tres años, dando, por tanto, valores para la medida de dispersión acorde a ellos.

Con la finalidad de evitar esto, es mejor no utilizar esas medidas también llamadas medidas de ***"dispersión absoluta"*** y utilizar en su lugar una medida de ***"dispersión relativa"***. Una de la más común es el coeficiente de variación.

- **El coeficiente de variación**

El coeficiente de variación se **define como la desviación estándar dividida entre la media y expresado como un porciento.**

El coeficiente de variación expresa la desviación estándar como un porciento de la media. Esta medida de dispersión relativa no depende de la unidad de medida ni de los valores absolutos de los datos dando una buena idea de la variabilidad para comparar dos o más conjuntos de datos. Tiene la desventaja de que no es exacta cuando la media está muy cerca de cero. Se denota por **C.V.** y se calcula por:

$$\mathbf{C.V.} = \frac{S}{\bar{x}} \times 100$$

Veamos un ejemplo: Se estudiaron dos grupos de pacientes masculinos. Un grupo estaba integrado por pacientes de 25 años de edad y el otro por pacientes de 11 años. Se tomó el peso en kilogramos encontrándose lo siguiente:

	G R U P O 1	**G R U P O 2**
Edad:	25 años	11 años
Peso medio:	72.5 kgs.	40 kgs.
Desviación estándar:	5 kgs.	5 kgs.

Si analizamos la dispersión del peso en los grupos por la desviación estándar podemos concluir que es igual en ambos grupos, en efecto, 5 kgs.

Se calcula el coeficiente de variación en cada grupo:

G R U P O 1	**G R U P O 2**
C.V.= $\frac{5}{72.5}$ x100	**C.V.**= $\frac{5}{40}$ x100
C.V.= 6.9%	**C.V.**= 12.5%

Por el coeficiente de variación se concluye que el peso en el grupo 2 tiene más variabilidad que en el grupo 1.

Medidas de posición relativa (percentiles). Características, cálculo e interpretación. Su utilidad en los estudios de crecimiento y desarrollo.

Las medidas de posición relativa se **definen como las medidas que describen la posición de un dato con relación al número total de datos.** Solamente estudiaremos los **percentiles**. Una condición obligada de estas medidas es que el conjunto de datos tiene que ser ordenado por orden de magnitud, generalmente en orden ascendente.

- **Percentiles**
  Los percentiles dividen el conjunto ordenado de datos en 100 partes iguales. Se denotan por la letra **P** seguido del número del percentil. Luego, $\mathbf{P_5}$ sería el percentil 5. Como los percentiles dividen el conjunto ordenado de datos en 100 partes iguales, el significado del percentil es que debajo de su valor se encontraría un porciento determinado de elementos y por encima otro. Por ejemplo:
  - ► $\mathbf{P_5}$ divide el conjunto ordenado de datos en dos partes: por debajo del valor del $\mathbf{P_5}$ estaría el 5% de los datos y por encima el 95%;
  - ► $\mathbf{P_{35}}$ divide al conjunto ordenado de datos en dos partes: por debajo de su valor estaría el 35% de los elementos y por encima el 65%.

Note que la suma de los dos porcentajes de los datos, en los que el percentil los divide, es 100%.

Cálculo del percentil

- Primero es obligado ordenar el conjunto de datos, usualmente en orden ascendente.

- Segundo aplicar la siguiente fórmula para encontrar la posición del percentil.

$$\text{Posición del percentil} = \frac{\text{Total de elementos en el conjunto de datos} \times \text{Número del percentil deseado}}{100}.$$

que sale de despejar la siguiente igualdad (regla de tres):

$$\frac{\text{Posición del percentil (X)}}{\text{Total de elementos del conjunto de datos (N)}} = \frac{\text{Número del percentil deseado}}{100}$$

La aplicación de la fórmula fuera la siguiente. Se desea calcular el percentil 15, $\mathbf{P_{15}}$

$$\text{Posición del percentil} = \frac{N \times 15}{100}$$

donde: N = Número total de elementos en el conjunto de datos
15 = El percentil deseado.
100 = Número de partes en que percentil divide al conjunto de datos.

- Análisis del resultado de la fórmula

  - Si el resultado es un entero, el valor del percentil es la media aritmética (semisuma) de los valores de los datos ocupando esa posición y la que sigue en orden ascendente.

  - Si el resultado es un entero con decimales, entonces la posición del percentil es el entero superior y el valor del percentil es el valor del dato que ocupa esa posición.

Por ejemplo, en un estudio se pesan 85 niños de 7 años de edad.

El peso de cada uno, ya ordenado en orden ascendente se muestra a continuación:

Posición:	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
Peso:	9.5	9.5	9.7	9.8	9.8	9.8	9.9	9.9	10	10	10	10
Posición:	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
Peso:	10	10	10.1	10.6	11	11.1	11.2	11.7	11.8	12	12	12.1
Posición:	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**
Peso:	12.6	12.7	12.9	13	13	13.1	13.8	13.9	14	14	14.1	14.3
Posición:	**37**	**38**	**39**	**40**	**41**	**42**	**43**	**44**	**45**	**46**	**47**	**48**
Peso:	14.5	14.5	14.6	14.7	14.7	14.8	14.9	15	15	15	15.1	15.1
Posición:	**49**	**50**	**51**	**52**	**53**	**54**	**55**	**56**	**57**	**58**	**59**	**60**
Peso:	15.1	15.1	15.1	15.2	15.2	15.3	15.6	15.7	16	16	16.1	16.2
Posición:	**61**	**62**	**63**	**64**	**65**	**66**	**67**	**68**	**69**	**70**	**71**	**72**
Peso:	16.4	16.5	16.6	17	17	17	17.3	17.6	17.8	18	18	18.1
Posición:	**73**	**74**	**75**	**76**	**77**	**78**	**79**	**80**	**81**	**82**	**83**	**84**
Peso:	18.6	19	19.2	19.6	19.7	19.8	20	20	20	20.1	20.1	20.3
Posición:	**85**											
Peso:	20.4											

Veamos algunos ejemplos: (N = 85, i.e., el estudio es de 85 niños.)

- Percentil 40 (**P_{40}**).

$$\text{Posición} = \frac{85 \times 40}{100} = \frac{3400}{100} = 34$$

Como el resultado es un entero, el valor de **P_{40}** será la media aritmética de las posiciones 34 y 35.

Posición 34--- Valor 14 kgs.
Posición 35--- Valor 14.1 kgs.

$$\text{El } P_{40} = \frac{14 + 14.1}{2} = 14.05 \text{ kgs.}$$

- Percentil 5 ($\mathbf{P_5}$).

$$\text{Posición} = \frac{85 \times 5}{100} = \frac{425}{100} = 4.25$$

Como el resultado tiene decimales, la posición es el entero que le sigue, o sea, la posición 5. El valor de la posición 5 es el percentil deseado, así $\mathbf{P_5}$ = 9.8 kgs.

- Percentil 50 ($\mathbf{P_{50}}$)

$$\text{Posición} = \frac{85 \times 50}{100} = \frac{4250}{100} = 42.5$$

Como el resultado tiene decimales, la posición es el entero que le sigue, o sea, la posición 43. El valor de la posición 43 es el percentil 50, $\mathbf{P_{50}}$ = 14.9 kgs.

El **percentil 50 es la mediana**, ya que divide al conjunto de datos en dos partes iguales, el 50% de los elementos por encima y el 50% por debajo.

Uso de los percentiles en Medicina y en Salud Pública

Los percentiles se utilizan en Medicina y en Salud Pública para encontrar valores de normalidad de una variable numérica. El uso más común es en las tablas de peso y talla, y en especial en Pediatría. En estas tablas, los percentiles de peso y talla se expresan por sexo y edad, permitiendo al Pediatra u otro personal de la salud evaluar el desarrollo de un niño en específico. Aunque no existe un criterio definitivo en cuanto a cuales son los percentiles que establecen el rango de normalidad, la mayoría de los pediatras coinciden de utilizar un rango entre $\mathbf{P_5}$-$\mathbf{P_{95}}$ o $\mathbf{P_3}$-$\mathbf{P_{97}}$.

En Cuba la encuesta de Crecimiento y Desarrollo de la Población llevada a cabo con una muestra representativa de la población y dirigida por el Ministerio de Salud Pública ha permitido establecer tablas de peso y talla por edades y sexo para su uso en Pediatría, utilizando los percentiles.

El percentil puede ser utilizado en otros estudios o investigaciones médicas, todo depende de lo que se quiera obtener como resultado.

Medidas para resumir datos cualitativos (razón, índice, proporción, porciento y tasa). Forma de cálculo e interpretación. Ejemplos en el campo médico.

Las variables categóricas o cualitativas son muy importantes en Medicina y Salud Pública. Los datos de variables categóricas son datos no-numéricos que solamente permiten el conteo de su presencia. (Vea distribución de frecuencias). Aunque las frecuencias absolutas y relativas de datos categóricos son importantes y resumen la información, por lo general son insuficientes para la mayoría de los análisis que se requieren.

En nuestro campo de aplicación son necesarias medidas que, además de resumir datos categóricos, sean capaces de expresar la relación entre categorías o entre variables y que puedan facilitar la comparación de resultados. Esto permitirá describir, interpretar y analizar un problema o situación específica. Estas medidas se conocen como **"indicadores de salud"**.

Un indicador de salud se define como la expresión, valor o número calculado que permite medir la magnitud de una situación de salud que está siendo estudiada. La Estadística tiene diferentes medidas para datos categóricos. Aquellos que se usan como indicadores de salud son la razón, el índice, la proporción, el porciento y la tasa.

- **La razón**
  La razón se **define como una fracción de la forma a/b, donde a y b son números de elementos que tienen diferentes cualidades en una variable o son diferentes variables.** La razón expresa la relación cuantitativa entre dos cualidades o categorías o entre dos variables. O sea, cuantos elementos tienen una cualidad o pertenecen a una categoría dada, con relación a un elemento que tiene otra cualidad o pertenece a otra categoría.

  Por ejemplo, si en un hospital existen 300 enfermeras y 50 médicos, la razón de enfermeras/médicos expresa cuantas enfermeras hay por cada médico.

  razón enfermeras/médicos = a/b donde a = número de enfermeras
  b = número de médicos

  razón = 300/50 = 6 . **La razón es 6 enfermeras por médico.**

Razones que se utilizan comúnmente en hospitales son:
- razón de camas/salas
- razón de equipos de rayosX/técnicos de rayosX
- razón de trabajadores por cama, y otras muchas.

- **El índice**
  El índice se **define como la razón multiplicada por 100.** El índice expresa una relación similar a la razón pero en relación con 100 elementos.

  El índice se calcula por la fórmula: **a/b x 100**.

  Los índices se utilizan cuando el numerador es menor que el denominador. Por ejemplo, un índice muy común utilizado en hospitales es el índice ocupacional. Este índice mide el número de camas ocupadas por pacientes con relación al número total de camas del hospital en un periodo dado de tiempo. La unidad de medida es días-cama, o sea, el número de camas en un día. Por ejemplo, si el hospital tiene 20 camas, en un día tiene 20 días-cama, pero en una semana tiene (20 x 7) = 140 días-cama. Los días-cama de los pacientes ingresados (también llamados días-paciente), son el número de días que un paciente ingresado ha estado ocupando una cama.

  Suponga que un hospital tiene 50 camas y se desea saber el índice ocupacional para un periodo de un mes (30 días), durante el cual los pacientes ingresados utilizaron un total de 1000 días-cama.

$$\text{Indice ocupacional} = \frac{\text{días-cama de pacientes}}{\text{Número total de días-cama del hospital}} \times 100$$

días camas de pacientes = 1000
días-cama del hospital en el periodo dado = 50 x 30 = 1500

$$\text{Indice ocupacional} = \frac{1000}{1500} \times 100 = 66.6 \text{ días-paciente por 100 diás-cama.}$$

y se interpreta que el hospital ha estado utilizando sus camas a un 66.6%

Por lo general, en los hospitales sólo se dice: Indice ocupacional es 66.6. El índice ocupacional adecuado de un hospital debe estar entre 85-90.

Otro índice utilizado en hospitales quirúrgicos es el índice de infección hospitalaria que mide la cantidad de infecciones intra-hospitalarias con relación al número de operaciones limpias en un periodo de tiempo dado. (Operación limpia es aquella realizada a un paciente sin infección). Se calcula por:

$$\text{Indice infección hospitalaria} = \frac{\text{Numero de pacientes infectados en el hospital}}{\text{Numero de operaciones limpias}} \times 100$$

Para que el índice de infección hospitalaria sea considerado de bien debe estar por debajo de 6.5 infecciones por cada 100 operaciones limpias realizadas.

- **La proporción**
  La proporción se **define como la fracción de la forma a/N, donde a es el número de elementos que tiene una característica y N el número total de elementos.** La proporción mide **"el peso"** de una cualidad o categoría o de una variable con relación al total de elementos del conjunto de datos.

  Por ejemplo, la proporción de enfermeros en un hospital, si el hospital tiene 1200 trabajadores y de ellos 300 son enfermeros sería:

  Proporción = a/N, donde **a** = número de enfermeros = 300
  N = número total de trabajadores = 1200

  $$\text{Proporción de enfermeros} = \frac{300}{1200} = \frac{1}{4} = 0.25$$

  El resultado puede ser expresado como ¼, esto es un enfermero por cada 4 trabajadores, ó 0.25 enfermeros por trabajador.

  Tome nota que la proporción se calcula igual que la frecuencia relativa que se vio en la distribución de frecuencias.

  Las proporciones se utilizan con mucha frecuencia en los estudios e investigaciones médicas.

- **El porciento o porcentaje**
  El porciento se **define como la proporción multiplicada por 100.** Es una medida muy común en todos los campos de aplicación. Se denota por **p** y su fórmula es:

  $$\mathbf{p = a/N \times 100.}$$

  En el ejemplo anterior, el porciento de enfermeros del hospital sería:

  $$p = \frac{300}{1200} \times 100 = 0.25 \times 100 = 25\%$$

  El 25% de los trabajadores del hospital son enfermeros.

- **La tasa**
  La tasa se **define como la relación entre el número de elementos en que ocurre un específico fenómeno y el total de elementos expuestos a la ocurrencia del fenómeno.** La tasa expresa el riesgo o posibilidad de ocurrencia de un fenómeno o situación en una población dada en un periodo de tiempo dado. Se calcula por la fórmula:

$$\textbf{tasa} = a/N \times 10^n$$

donde: **a** = frecuencia con que un fenómeno o situación ha ocurrido en un periodo determinado de tiempo.

N = Número de personas expuestas a la ocurrencia del fenómeno o situación en el mismo periodo de tiempo.

$\mathbf{10^n}$ = Factor de multiplicación de manera de evitar tener solamente decimales en los resultados, ya que el denominador es siempre mayor que el numerador. Por tanto, acorde con el denominador este factor sería $10^2 = 100$; $10^3 = 1000$; $10^4 = 10000$; ..... 10^n

Las tasas se usan mucho en Medicina y en Salud Pública para expresar el riesgo de morir, (tasas de mortalidad), el riesgo de enfermar, (tasas de morbilidad) y otros aspectos importantes relacionados con la salud. Se verán en detalle en el capítulo dedicado a las estadísticas de salud.

INDICE

Capítulo 6
PRESENTACION DE LA INFORMACION

Autor: Dr. Emilio Morales Jiménez

SUMARIO

- *Presentación de la Información.*
- *Cuadro o tabla estadística. Partes que la constituyen. Diferentes tipos de tablas estadísticas.*
- *Gráficos estadísticos. Tipos de gráficos acorde a la variable utilizada: gráfico de barras; gráfico de sector o pastel; histograma y polígono de frecuencias).*
- *Otros gráficos: gráfico aritmético simple y diagrama de dispersión.*

Presentación de la información.

Como se ha visto anteriormente la etapa de Procesamiento de la Información tiene tres sub-etapas: Organización y clasificación de los datos; Resumen de los datos y Presentación de los datos. La Presentación de los datos es la tercera y última sub-etapa.

Una vez que los datos han sido clasificados y resumidos es necesario presentarlos en una forma adecuada que permita su análisis e interpretación. Hay dos formas de presentar los datos: la presentación tabular por medio de tablas estadísticas y la presentación gráfica mediante los gráficos estadísticos. **La forma principal y esencial de presentación de los datos es la tabla estadística la cual se complementa por el gráfico estadístico para resaltar los principales resultados mostrados en ella.** Es un error presentar resultados de datos que han sido procesados solamente con gráficos estadísticos debido a que el lector no puede tener una idea completa de los mismos.

En Medicina y en Salud Pública las tablas estadísticas complementadas por gráficos estadísticos se usan con mucha frecuencia. En casi todos los trabajos científicos, artículos publicados en revistas médicas, en libros de Medicina y en los análisis periódicos del trabajo realizado la representación tabular y gráfica de los resultados son una necesidad obligada. Ellos hacen más comprensible y esclarecen las explicaciones escritas brindadas en los textos.

Cuadro o tabla estadística. Partes que la constituyen.

La tabla estadística es la tabulación de los datos procesados de manera de presentarlos en una forma resumida y comprensible. Toda tabla estadística tiene tres partes que la constituyen: **el título; el cuerpo o tabla propiamente dicha y el pie, calce o notas explicativas.**

En la tabla que se presenta a continuación se señalan cada uno de ellos.

TITULO **CUERPO O TABLA PROPIAMENTE DICHA**

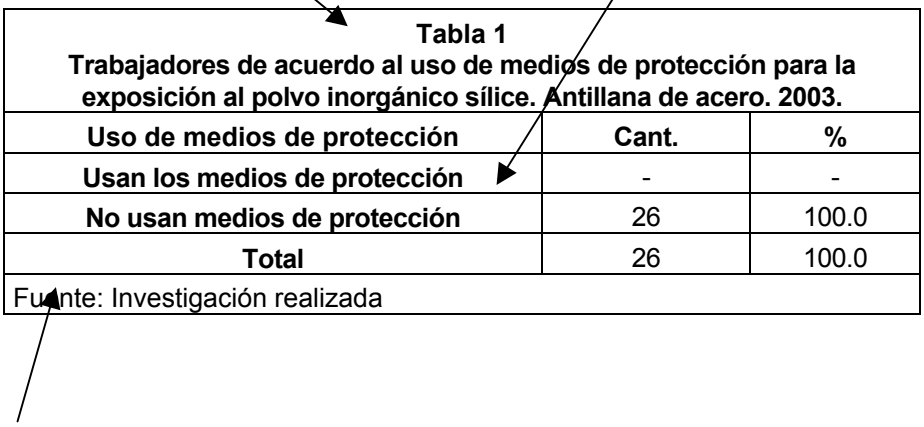


Tabla 1 **Trabajadores de acuerdo al uso de medios de protección para la exposición al polvo inorgánico sílice. Antillana de acero. 2003.**		
Uso de medios de protección	**Cant.**	**%**
Usan los medios de protección	-	-
No usan medios de protección	26	100.0
Total	26	100.0
Fuente: Investigación realizada		

PIE O CALCE DE LA TABLA (NOTAS EXPLICATIVAS)

- **TITULO**

El título es el encabezado de la tabla estadística y debe describir en forma concisa, clara y completa el contenido de la tabla. Esos tres requisitos del título**: conciso, claro y completo** son muy importantes. El título para ser completo tiene que responder a las preguntas – ¿Qué? -¿Cómo? -¿Dónde? -¿Cuándo?

- Que se está presentando en la tabla y como se esta presentando.
- Donde se realizó el trabajo.
- Cuando se realizó el trabajo.

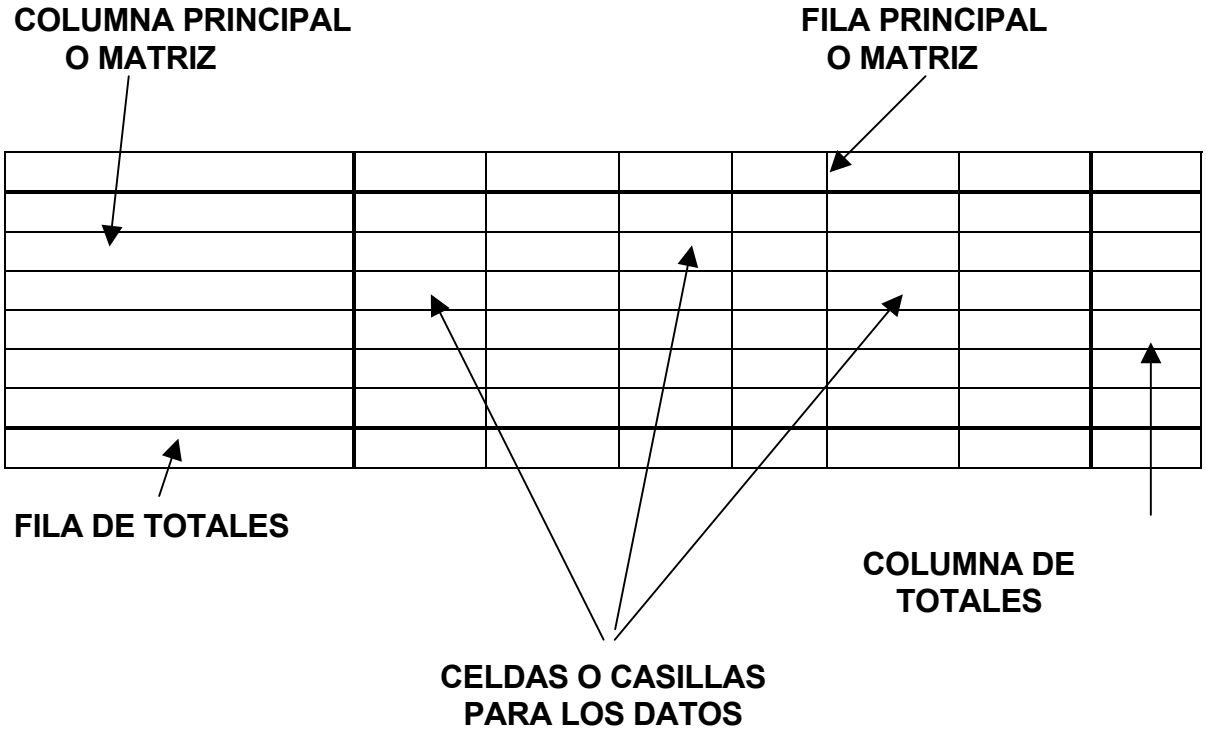
En el título del ejemplo anterior:
- ¿Qué? Trabajadores
- ¿Cómo? De acuerdo al uso de medios de protección contra el polvo inorgánico sílice.
- ¿Dónde? Antillana de Acero
- ¿Cuándo? 2003

Como por lo general se presentan los datos en más de una tabla, las mismas se numeran comenzando por el número 1. La numeración forma parte del título y sirve además para identificar la tabla y referirse a ella en una forma abreviada.

❖ **CUERPO O TABLA PROPIAMENTE DICHA**

El cuerpo de la tabla es su parte fundamental. Los datos procesados se tabulan en esta parte, o sea, se colocan en filas y columnas facilitando el análisis. Luego, el cuerpo esta formado por un conjunto de celdas o casillas delimitadas por las filas y columnas.

ESQUEMA DEL CUERPO O TABLA PROPIAMENTE DICHA



Fila principal o matriz: Es, por lo general, la primera fila y se utiliza para escribir los encabezados de las columnas, esto es, la descripción del contenido de cada columna. Frecuentemente, el encabezado de las columnas requiere más de una línea, en estos casos la fila matriz tiene varias líneas o subdivisiones para dar detalles específicos de los datos. La fila matriz debe tener una doble línea para separarla del resto del cuerpo de la tabla.

Columna principal o matriz: Es la primera columna y es utilizada para escribir los encabezados de las filas, o sea la descripción del contenido de cada fila. También puede tener subdivisiones como la fila principal en dependencia de los datos. La columna matriz debe tener una doble línea para separarla del resto del cuerpo de la tabla.

Columna de totales: Es la última columna y es utilizada para el total de las filas. Esta columna se puede omitir si los totales por filas no se necesitan. Si existe debe estar separada por una doble línea del resto del cuerpo de la tabla.

Fila de totales: Es la última fila y es utilizada para el total de las columnas. Esta fila puede ser omitida si los totales por columna no se necesitan. Si existe debe estar separada por una doble línea del resto del cuerpo de la tabla.

Celdas o casillas para los datos: Las celdas o casillas están delimitadas por el resto de las filas y columnas de la tabla. En ellas se colocan los valores resultantes del procesamiento de los datos.

❖ PIE, CALCE O NOTAS EXPLICATIVAS

El PIE es la última parte de la tabla. Se escribe debajo del cuerpo de la tabla e incluye la fuente de los datos así como cualquier información complementaria necesaria para esclarecer el contenido de la tabla.

DIFERENTES TIPOS DE TABLAS ESTADISTICAS

Según el número de variables utilizadas en la representación tabular las tablas se clasifican en:

❖ Tablas de una sola entrada

Tabla 5
Trabajadores expuestos al polvo inorgánico sílice con Neumopatías según enfermedad respiratoria específica.
Antillana de acero. 2003.

Enfermedad respiratoria específica	Cant.	%
Bronquitis crónica	7	87.5
Fibrosis pulmonar	1	12.5
Silicosis	-	-
Cáncer del pulmón	-	-
Asma ocupacional	-	-
Total	8	100.0

Fuente: Investigación realizada

La tabla de una sola entrada solamente muestra una variable. Este tipo de tabla, aunque se utiliza con mucha frecuencia, solamente permite analizar una variable, en este caso las enfermedades pulmonares encontradas en los trabajadores expuestos al polvo inorgánico sílice con Neumopatías.

❖ Tabla de dos entradas

Tabla 4
Trabajadores expuestos a polvo inorgánico sílice según presencia de Neumopatías y hábito tabáquico. Antillana de acero. 2003.

Hábito tabáquico	Con Neumopatías		Sin Neumopatías		Total	
	Cant	%	Cant	%	Cant	%
Fumadores	7	83.3	9	50.0	16	61.5
No fumadores	-	-	4	22.2	4	15.4
Ex-fumadores	1	16.7	5	27.8	6	23.1
Total	8	30.8 *	18	69.2 *	26	100.0

Fuente: Investigación realizada
Nota: * El porcentaje calculado del total de la fila. El resto de los porcentajes calculados del total de la columna.

La tabla de dos entradas muestra dos variables. Este tipo de tabla es el más frecuentemente utilizado. El mismo ahorra espacio en los trabajos científicos y en los informes finales de investigaciones y permite analizar cada variable por separado o en relación de una con la otra.

En el ejemplo se muestra el hábito tabáquico por las filas y la presencia de Neumopatías por las columnas. Los totales de las filas y las columnas permiten analizar independientemente el hábito de fumar y la presencia de Neumopatías en los trabajadores expuestos al polvo inorgánico sílice. Utilizando los otros datos se analiza la presencia o no de Neumopatía en relación con el hábito de fumar, o sea relacionando las dos variables pudiendo hacer comparaciones útiles para el análisis de los datos.

Hay dos tablas específicas de dos entradas que se usan muy frecuentemente en Estadística Inferencial. Ellas son la **"tabla de contingencia"** y la **"tabla de 2 x 2".** La primera se utiliza en la prueba del Chi-cuadrado y la segunda se utiliza en la cuantificación de factores de riesgo y en el método del valor predictivo. Se estudiarán en estadística inferencial.

❖ Tablas de tres entradas, cuatro entradas, etc..

Estas tablas raramente se utilizan pero se pueden realizar. En ellas se combinan tres, cuatro o más variables respectivamente. Son tablas muy complejas de analizar y es recomendable no utilizarlas.

REGLAS GENERALES PARA LA CONSTRUCCION DE TABLAS ESTADISTICAS

1. El título debe cumplir los requerimientos de ser claro, conciso y completo.
2. Los encabezados de las columnas y las filas deben ser claros y definir claramente el contenido de las celdas o casillas de datos.
3. La unidad de medida utilizada debe ser incluida en los encabezados del cuerpo de la tabla.

4. Si la tabla incluyera estadígrafos tales como la media aritmética, porcentaje, tasas, etc.., los valores básicos con que fueron calculados deben estar incluidos en la tabla. Por ejemplo, una tabla que solamente muestre los porcentajes y no muestre las frecuencias absolutas, (de donde se obtuvo el porcentaje), está incompleta.
5. La tabla debe incluir todos los datos en relación con la variable que está siendo tabulada. La publicación de solamente parte de los datos de una distribución es incorrecta ya que no permite verificar al lector la validez de los argumentos del autor.
6. La tabla no debe estar sobrecargada con datos de manera que se pueda entender con facilidad.
7. La fuente de los datos debe aparecer siempre en el CALCE.
8. Si debido a los datos existiera alguna posibilidad de confusión para hacer el análisis, ello debe aclararse adecuadamente en el calce.
9. Si para presentar los resultados del procesamiento de datos se requieren varias tablas estadísticas los totales deben corresponderse entre ellas. Es imprescindible que entre las tablas no existan contradicciones en los datos.

COMO SE ANALIZA UNA TABLA ESTADISTICA

1. Leer y entender el título.
2. Leer y entender el calce y notas explicativas si las hubiera.
3. Leer los encabezados de las columnas y las filas de manera de conocer la organización de la información dentro de la tabla.
4. Precisar la unidad de medida y darse cuenta de la forma en que están calculados los porcentajes u otro estadígrafo que se utilice.
5. Observar la información de las casillas, comprendiendo su significado, de manera de familiarizarse con el comportamiento general de los datos.
6. Analizar primero los totales.
7. Relacionar los totales con los valores máximos y mínimos en las columnas y las filas.
8. Analizar las diferencias que presentan los datos en relación con las variables.
9. Si hay varias tablas estadísticas, busque las relaciones entre ellas.

Gráficos estadísticos

El gráfico estadístico es la representación de los datos procesados mediante figuras geométricas (puntos, líneas, rectángulos, circunferencias, etc..) con dimensiones proporcionales a la magnitud de los datos. El gráfico estadístico es un complemento de la tabla estadística. Su principal objetivo es mostrar los aspectos relevantes de los datos tabulados de manera que el lector se de cuenta de cuales son con una sola ojeada.

Los gráficos estadísticos tienen dos limitantes con relación a las tablas estadísticas:

- Los gráficos no pueden presentar tanta información como la presentada en las tablas estadísticas.
- Muchas veces los datos se presentan en los gráficos con datos aproximados, mientras que en las tablas se presentan con el valor exacto.

El gráfico estadístico tiene tres partes: **el título, cuerpo o gráfico propiamente dicho y el pie o calce (notas explicativas).** El título y el pie o calce tienen las mismas características que lo ya descrito para las tablas estadísticas.

- **CUERPO O GRAFICO PROPIAMENTE DICHO**

El cuerpo del gráfico es su parte principal. Tiene dos componentes: la representación pictórica de los datos y la leyenda.

Representación pictórica

La representación pictórica de los datos depende del tipo de variable. Para las **variables categóricas o cualitativas y numéricas o cuantitativas discretas** se utilizan:

- **gráficos de barra y**
- **gráfico circular o pastel.**

Para las **variables numéricas o cuantitativas continuas** se utilizan:

- **histograma y**
- **polígono de frecuencias.**

Existen otros gráficos que se utilizan con mucha frecuencia en Medicina y Salud Pública. Cuando se necesita representar **datos relacionados con el tiempo** se utiliza el **gráfico aritmético simple**. Cuando se necesita **representar dos valores cuantitativos continuos de un conjunto de datos para ver su distribución en un eje de ordenadas X y Y** se utiliza el **diagrama de dispersión.**

La leyenda

La leyenda es la forma en que los diferentes datos están identificados en el gráfico. No todos los gráficos requieren leyenda, pero cuando se representa en un gráfico más de una variable o se representan distintas categorías de una misma variable es obligado el uso de la leyenda para que se pueda entender.

TIPOS DE GRÁFICOS ACORDE A LA VARIABLE UTILIZADA

- **Variables categóricas o cualitativas y numéricas o cuantitativas discretas**

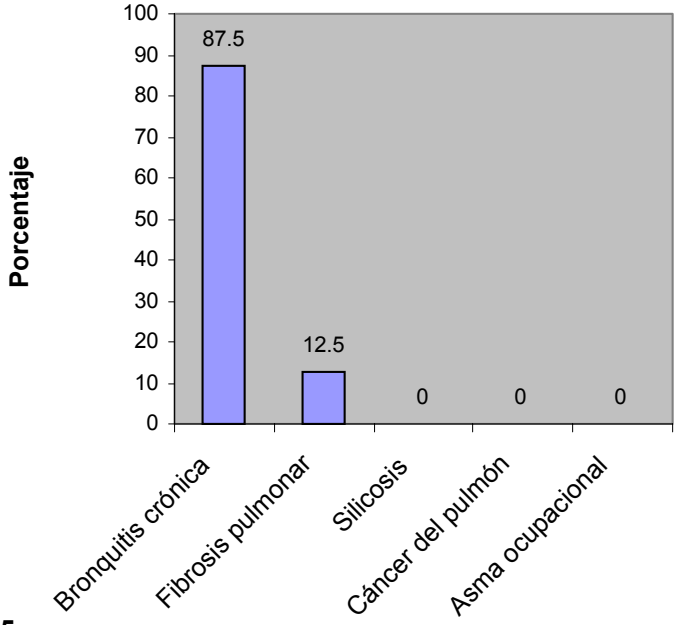
  - **Gráfico de barras**
    - ⇒ Las reglas generales para este tipo de gráfico son:
      - Las barras deben tener el mismo ancho.
      - El espacio entre barras debe ser igual y no debe ser menor que la mitad del ancho de la barra ni mayor que el ancho de la barra.
      - La escala de medida debe comenzar en cero con intervalos de igual tamaño y magnitud.
    - ⇒ Existen diferentes tipos de gráficos de barras. Los que se utilizan con más frecuencia son: los gráficos de barra simple, barra doble, barra compleja y barra proporcional.

GRAFICO DE BARRA SIMPLE

Esta gráfico se utiliza para representar los datos de una variable. Por ejemplo, veamos el gráfico que correspondería a la tabla 5 de los párrafos anteriores.

Gráfico 1

Trabajadores expuestos a polvo inorganico silice con Neumopatias segun tipo de Neumopatia. Antillana de acero. 2003

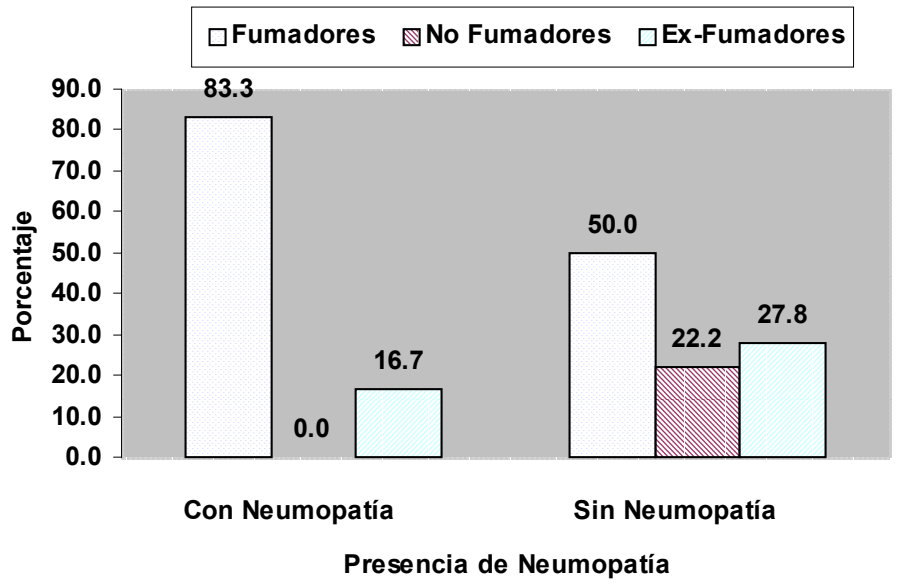


Fuente: Tabla 5

GRAFICO DE BARRAS DOBLE

Este gráfico se utiliza para representar los datos de dos variables. Por ejemplo, hagamos el gráfico de la tabla 4 descrita en párrafos anteriores de las variables presencia de Neumopatías según el hábito tabáquico.

Gráfico 2
Trabajadores expuestos al polvo inorgánico sílice según presencia de Neumopatías y Hábito Tabáquico. Antillana de Acero. 2003



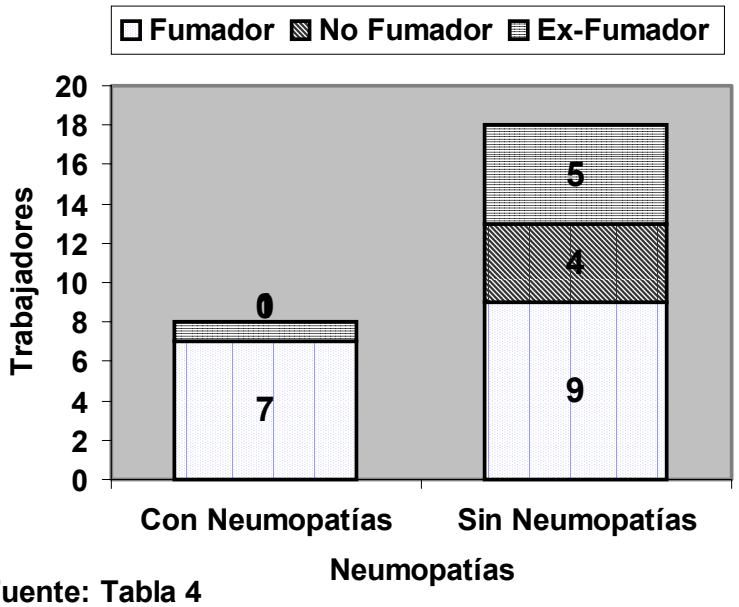
Fuente: Tabla 4

En los dos gráficos precedentes cada barra representa una característica o una categoría de una variable y se llama barra simple. La escala de medida se encuentra en el eje de las Y mientras las categorías de las variables se hallan en el eje de las X. Las barras pueden ser lineales o tridimensionales y la frecuencia que representan puede o no mostrarse en la parte superior de la barra. La escala de medida utilizada está en relación con los valores que se van a representar.

GRAFICO DE BARRAS COMPLEJAS (COMPUESTA)

Este gráfico se utiliza para representar al menos dos variables. En el cada barra está integrada con todas las categorías de una variable. Por ejemplo, hagamos el gráfico con los mismos datos que los utilizados en la barra doble.

Gráfico 3
Trabajadores expuestos a polvo inorgánico sílice según presencia de Neumopatía y hábito tabáquico. Antillana de Acero. 2003

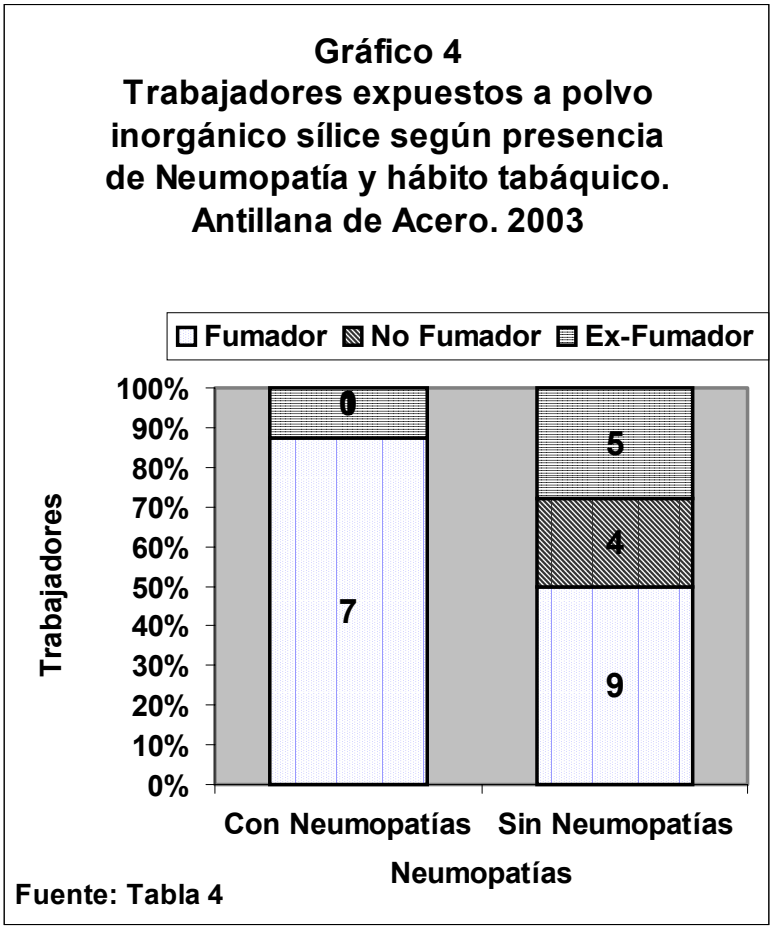


Fuente: Tabla 4

Como puede verse cada barra está formada por las tres categorías de la variable hábito tabáquico. La categoría con mayor valor se coloca en la parte inferior de la barra. La altura de la barra es la suma de las frecuencias de las categorías. Cada barra tiene dos informaciones: la frecuencia de cada categoría y la frecuencia total que equivale a la altura de la barra. Esa frecuencia total corresponde ya a la otra variable, presencia de Neumopatías. La barra de la izquierda representa los que presentan Neumopatías y la de la derecha los que no la presentan.

GRAFICO DE BARRA PROPORCIONAL

Este gráfico es utilizado para representar en cada barra un conjunto completo de datos mostrando la proporción (porcentaje) que cada categoría presenta. Cada barra, por tanto, representa 100 % y cada parte el porcentaje de la categoría. Comúnmente se muestra dentro de cada parte o a su lado la frecuencia o el porcentaje correspondiente. La escala de medidas siempre representará el porcentaje hasta el 100 % por lo que todas las barras tienen la misma altura. Hagamos el gráfico con los mismos datos que los anteriores.



Cualquier gráfico de barras puede tener orientación vertical, como los presentados, u horizontal.

GRAFICO CIRCULAR O PASTEL

Este gráfico se usa muy frecuentemente en Medicina y en Salud Pública. Es muy similar al gráfico proporcional en el sentido de que se representa todo el conjunto de datos. Este gráfico se confecciona tanto con los valores absolutos como los porcentajes de esos valores. Tiene como inconvenientes que solamente se puede representar una variable y que requiere de algunos cálculos.

Hagamos el gráfico con los totales de las filas de la tabla 4. Ellos son:

Total		
Hábito Tabáquico	Cant	**%**
Fumador	16	61.5
No Fumador	4	15.4
Ex- Fumador	6	23.1
Total	26	100.0

El gráfico es una circunferencia y cada categoría de la variable será representada por una tajada (desde el centro de la circunferencia hacia el borde) con un ángulo calculado de la siguiente manera:

⇒ La circunferencia tiene un ángulo de 360° representando el 100 %.
⇒ Para cada categoría se debe calcular el porcentaje que el valor absoluto representa para el total.
⇒ El cálculo del ángulo que represente al porcentaje de la categoría se realiza por el siguiente procedimiento:

$$\textbf{Angulo de la tajada de la categoría} = \frac{360 \times \text{Porcentaje de la categoría}}{100}$$

En nuestro ejemplo:

❖ Fumadores = 16 y el % = 61.5

$$\text{El ángulo es} = \frac{360 \times 61.5}{100} = 221.4°$$

❖ No Fumadores = 4 y el % = 15.4

$$\text{El ángulo es} = \frac{360 \times 15.4}{100} = 55.4°$$

❖ Ex-Fumadores = 6 y el % = 23.1

$$\text{El ángulo es} = \frac{360 \times 23.1}{100} = 83.2°$$

Los ángulos para cada sector o tajada del gráfico fueran:

Categoría	%	Angulo de la tajada o sector del gráfico
Fumador	61.5	221.4°
No Fumador	15.4	55.4°
Ex-Fumador	23.1	83.2°
Total	100.0	360°

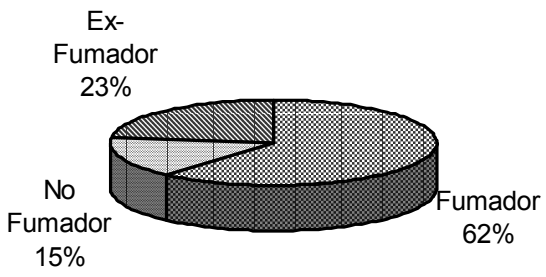
Una vez que se han calculado los ángulos se procede a realizar el gráfico, dividiendo la circunferencia en tajadas o sectores acorde a los ángulos.

El gráfico se puede realizar también con las frecuencias absolutas. Entonces la fórmula utilizada sería la siguiente:

$$\text{Angulo de la tajada de la categoría} = \frac{360 \times \text{Frecuencia absoluta de la categoría}}{\text{Número total de elementos}}$$

El gráfico fuera el siguiente:

Grafico 5
Trabajadores expuestos al polvo inorgánico sílice según hábito tabáquico. Antillana de acero. 2003



Fuente: Tabla 4

- **Gráficos para variables cuantitativas continuas**

HISTOGRAMA

El Histograma se realiza utilizando los límites de clase y por lo general se hace con orientación vertical. El mismo es un gráfico de barras simples pero sin separación entre las columnas por representar una variable continua.

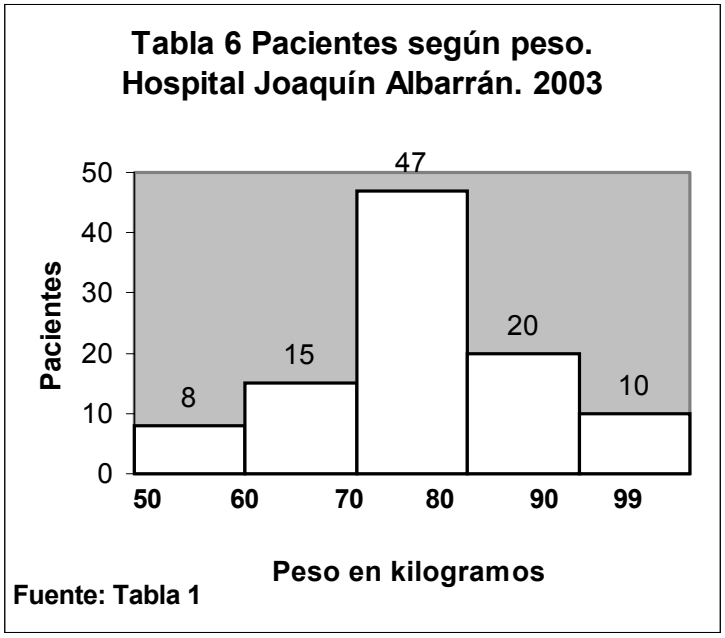
La siguiente tabla estadística muestra una variable cuantitativa continua, el peso de un grupo de pacientes.

Tabla 1
Pacientes estudiados según peso.
Hospital Joaquín Albarrán. 2003

Peso (en kilogramos)	Cant.	%
50 – 59	8	8.0
60 – 60	15	15.0
70 – 79	47	47.0
80 – 89	20	20.0
90 – 99	10	10.0
Total	100	100.0

Fuente: Investigación realizada

El histograma fuera el siguiente:

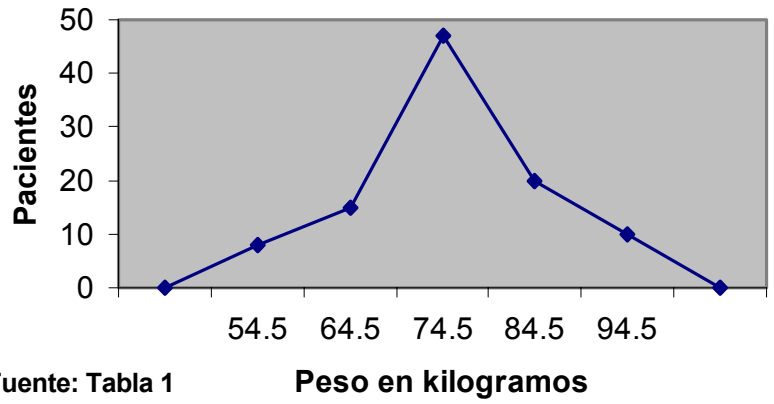


El ancho de cada barra es proporcional a la amplitud del intervalo de clase. La altura de cada barra es la frecuencia absoluta o cantidad de elementos de cada clase. Fíjese que en el eje de las X se colocan los límites de clase.

POLIGONO DE FRECUENCIAS

Este gráfico se hace en forma muy similar al histograma, pero utilizando la marca de clase. Una manera práctica de realizarlo es dibujar el histograma, marcar el punto central del límite superior de la barra y conectar esas marcas por una línea continua. Después se borran las barras. Los dos extremos de la línea trazada se bajan al eje de las X. en el punto central de intervalos clases colocados a ambos extremos. Esto permite que el área total bajo la curva sea la misma que la del histograma.

Tabla 7
Pacientes según peso. Hospital Joaquín Albarrán. 2003



Fuente: Tabla 1

Otros tipos de graficos

En Medicina y Salud Pública se utilizan también, con mucha frecuencia, otros tipos de gráficos. Cuando se va a representar gráficamente una serie cronológica, o sea datos con relación al tiempo se utiliza el gráfico aritmético simple. Cuando se quiere establecer una relación lineal entre dos variables cuantitativas continuas se realiza el gráfico de dispersión.

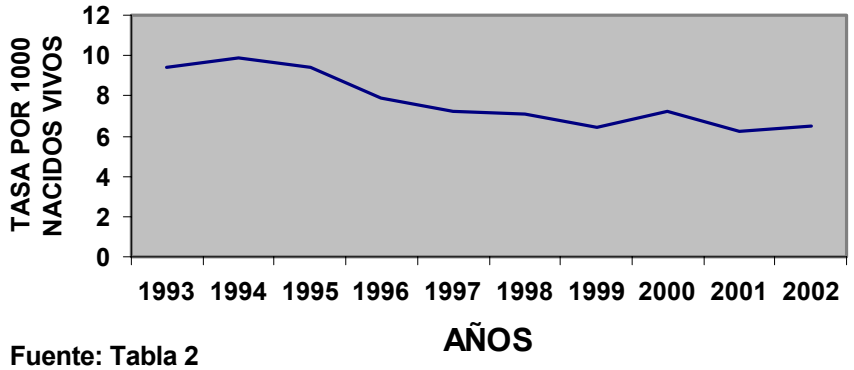
GRAFICO ARITMETICO SIMPLE

Este gráfico se utiliza mucho cuando se está analizando el comportamiento de la morbilidad, mortalidad u otra medida en un lapso de tiempo. El eje horizontal, o de las X se utiliza para el tiempo y el eje vertical o de las Y para los valores numéricos de los datos. Para ejemplificar tomaremos los datos de 10 años de la mortalidad infantil en Cuba.

Tabla 2 Mortalidad Infantil en Cuba 1993 – 2002	
Año	**Tasa de Mortalidad Infantil por 1000 nacidos vivos**
1993	9.4
1994	9.9
1995	9.4
1996	7.9
1997	7.2
1998	7.1
1999	6.4
2000	7.2
2001	6.2
2002	6.5
Fuente: Dpto. Estadísticas MINSAP	

El Gráfico aritmético simple fuera:

Gráfico 8
Mortalidad Infantil. Cuba. 1993 - 2002



Fuente: Tabla 2

DIAGRAMA DE DISPERSION

Este gráfico es muy sencillo de hacer. Se basa en dos ejes de coordenadas (X,Y). El eje de las X es el horizontal y se le asignan los valores de la variable que consideraremos independiente. El eje de las Y es vertical y se le asignan los valores de la variable que consideraremos como dependiente.

Para realizar el gráfico, una vez establecidas las escalas en ambos ejes, se coloca un punto donde se entrecrucen los valores de las variables de ambos ejes. Cada punto va estar formado por un valor del eje de las X y otro del eje de las Y.

Por ejemplo, tenemos los valores de peso y talla de 20 jóvenes.

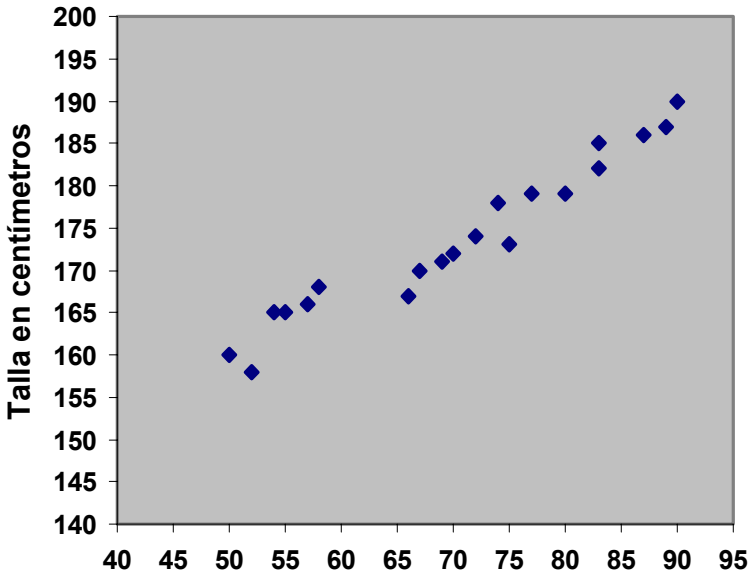
Tabla 3
Jóvenes según peso y talla. Municipio Playa. 2003

Peso (kilogramos)	Talla (centímetros)	Peso (kilogramos)	Talla (centímetros)
80	179	70	172
83	182	67	170
66	167	58	168
50	160	54	165
75	173	87	186
90	190	52	158
89	187	74	178
83	185	69	171
55	165	57	166
72	174	77	179

Fuente: Investigación realizada

El Diagrama de dispersión fuera el siguiente:

Gráfico 9
Diagrama de dispersión de peso y talla de jovenes. Municipio Playa. 2003



Fuente: Tabla 3

Cada punto del gráfico corresponde donde se cruza el peso con la talla en cada joven. De esta forma se aprecia la dispersión de los pares de datos. En este ejemplo esta evidente que a medida que el peso aumenta la talla también se incrementa.

INDICE

Capítulo 7
DEMOGRAFIA Y ESTADISTICAS DE SALUD

Autor: Dr. Emilio Morales Jiménez

SUMARIO

- *Demografía. Concepto.*
- *Estadísticas de población. Su utilización en Salud Pública.*
- *Composición de la población. Pirámides de población. Diferentes tipos. Interpretación en países con diferentes grado de desarrollo.*
- *Concepto de estadísticas de salud. Estadísticas utilizadas en el campo de la salud*
- *Estadísticas de Natalidad. Principales conceptos. Medidas de fecundidad.*
- *Medidas de mortalidad. Principales conceptos. Mortalidad infantil y sus componentes. Mortalidad materna. Mortalidad general. Mortalidad específica.*
- *Medidas de morbilidad. Principales conceptos. Incidencia, Prevalencia, Letalidad. Morbilidad conocida y desconocida.*
- *Indicadores que miden los recursos y servicios disponibles y su grado de utilización. Médicos por mil habitantes. Camas por mil habitantes. Embarazadas captadas en primer trimestre. Mortalidad hospitalaria bruta y neta.*
- *Indicadores que miden los aspectos de higiene y saneamiento.*

Demografía. Concepto

Ciencia que estudia las poblaciones humanas, esto es, su tamaño, distribución geográfica, composición por diferentes características (edad, sexo, nivel escolar, lugar de vivienda y otros), migraciones ó movimientos poblacionales, variaciones y las causas de esas variaciones.

Tiene dos campos de estudio:

- El análisis demográfico que investiga los componentes de las variaciones de la población y sus cambios.
- Estudios sobre población que incluye los estudios propios de la demografía relacionado con la fecundidad, mortalidad y migraciones; y el estudio de las relaciones entre los cambios de la población y la influencia que sobre ellos pueden haber ejercido variables propias de otras ciencias como la economía, medicina y sociología entre otras.

Estadísticas de población. Su utilización en Salud Pública.

Las estadísticas de población forman parte de las estadísticas demográficas y se centran en la composición de la población acorde a diferentes atributos y características tales como sexo, edad, estado civil, lugar de nacimiento y situación ocupacional entre muchos otros. Existen estudios de composición de una determinada población por sexo, por grupos de edades, por estado civil, por nivel de escolaridad y por cualquier característica de interés.

Las características demográficas más importantes son la edad y sexo. El análisis conjunto de la edad y sexo de una población determinada proporciona **la estructura de la población,** elemento básico para el estudio de cualquier población. La representación gráfica de la distribución por edad y sexo de la población se denomina **pirámide de población**.

Utilizacion en la salud publica

En la Salud Pública se requiere conocer la composición de la población por muchas características. El ser humano es un ser bio-psico-social y las enfermedades y problemas de salud que lo afectan están condicionados en muchas ocasiones por la composición de la población. Baste señalar dos ejemplos;

- Para elaborar los distintos programas de salud hay que tener siempre en cuenta la composición de la población hacia los cuales se dirigen.
- Para planificar la atención médica es menester conocer la composición de la población por determinadas características.

Composición de la población. Pirámides de población. Diferentes tipos. Interpretación en países con diferentes grado de desarrollo.

La composición de la población, por lo general se basa en su distribución y análisis por edad y sexo. Ello se representa mediante un gráfico donde se muestra la composición de la población por edad y sexo.

Ese gráfico son dos histogramas horizontales por grupos de edades, uno por cada sexo, opuestos por la base. Convencionalmente el sexo masculino se coloca del lado izquierdo y el sexo femenino del lado derecho. La magnitud de la población por grupos de edades se puede reflejar de forma absoluta ó relativa en porcentaje. Por lo general, los grupos de edad son de una amplitud de 5 años siendo los siguientes: 0-4; 5-9; 10-14; y así sucesivamente.

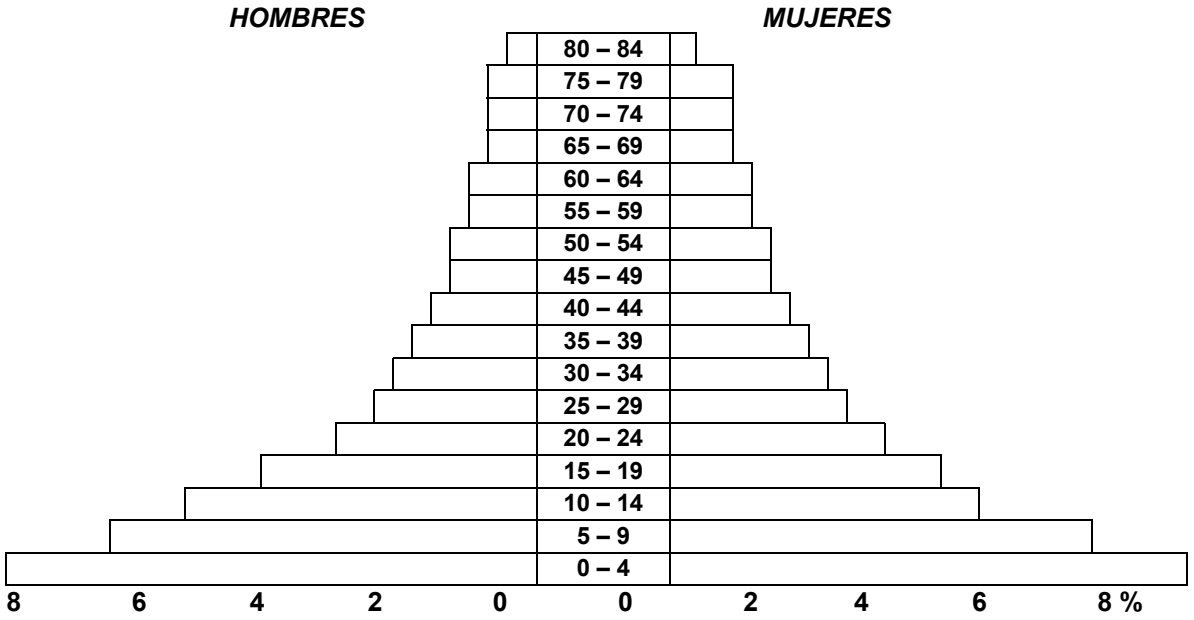
El rectángulo que representa un grupo de edad no es igual para las diferentes poblaciones. **Su longitud depende del comportamiento de las tres variables demográficas que determinan los cambios poblacionales: mortalidad, fecundidad y migración.** Sobre la base del análisis de la estructura de la población se pueden distinguir tres tipos de poblaciones:

- ⇒ Población joven = Pirámide joven
- ⇒ Población vieja = Pirámide vieja
- ⇒ Población intermedia = Pirámide intermedia

Pirámides de población

PIRAMIDE JOVEN

COSTA RICA 1963

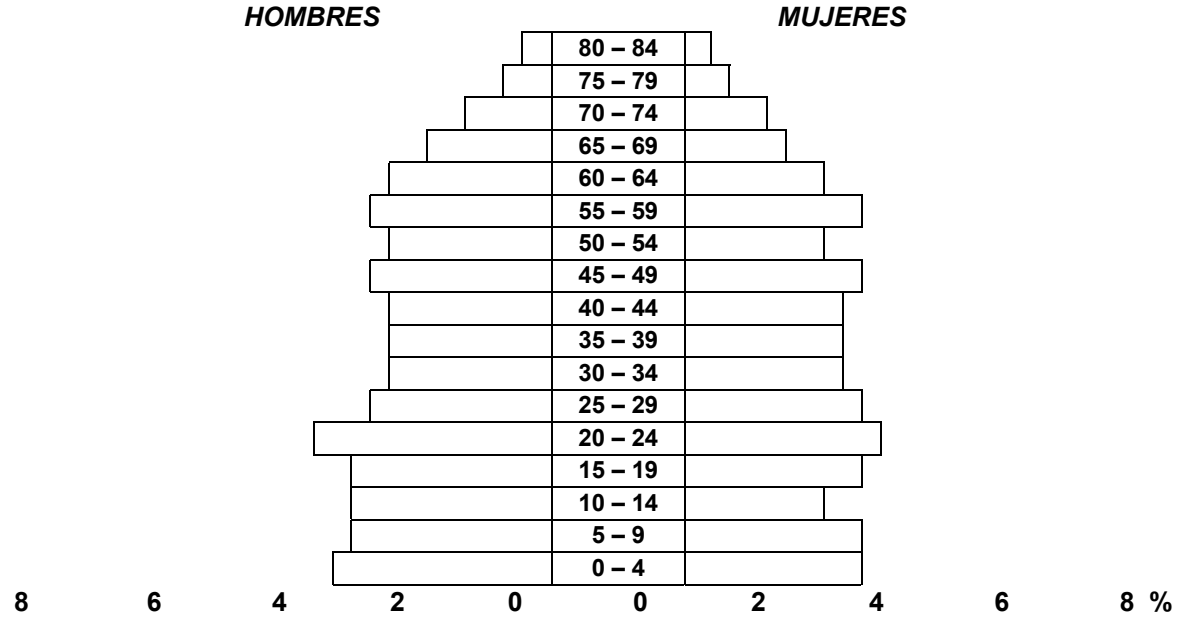


Caracteristicas

- ✓ Base ancha
- ✓ Rápida disminución hacia la cúspide
- ✓ Poca altura
- ✓ Población joven
- ✓ Representan alta natalidad y alta mortalidad
- ✓ Nivel de salud inadecuado
- ✓ Corresponde a países subdesarrollados

PIRAMIDE VIEJA

SUECIA 1969

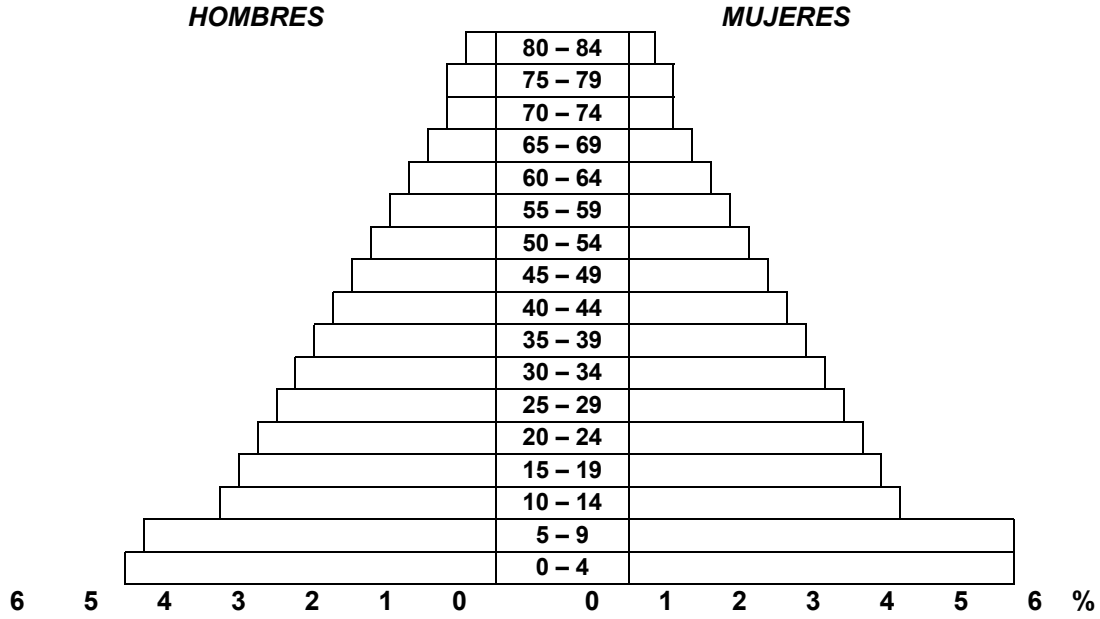


Caracteristicas

- ✓ Base estrecha
- ✓ Presenta sinuosidades o abultamientos (relacionadas con guerras, migraciones, epidemias)
- ✓ Gran altura
- ✓ Población vieja
- ✓ Representan baja natalidad y baja mortalidad
- ✓ Nivel de salud adecuado
- ✓ Corresponde a países desarrollados

PIRAMIDE INTERMEDIA

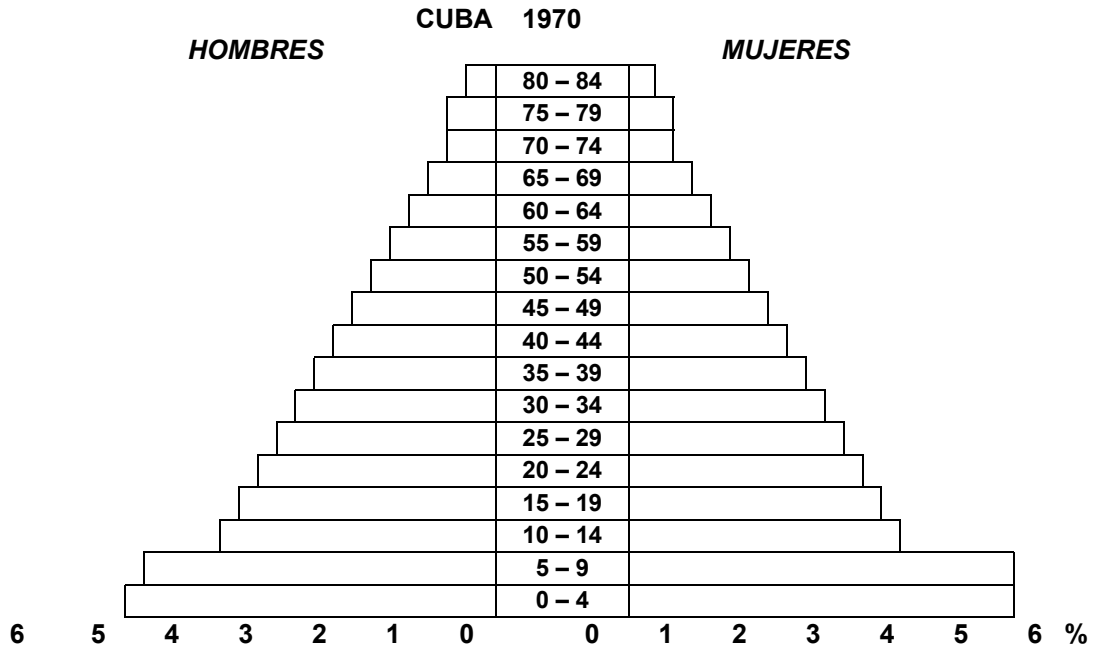
CUBA 1970

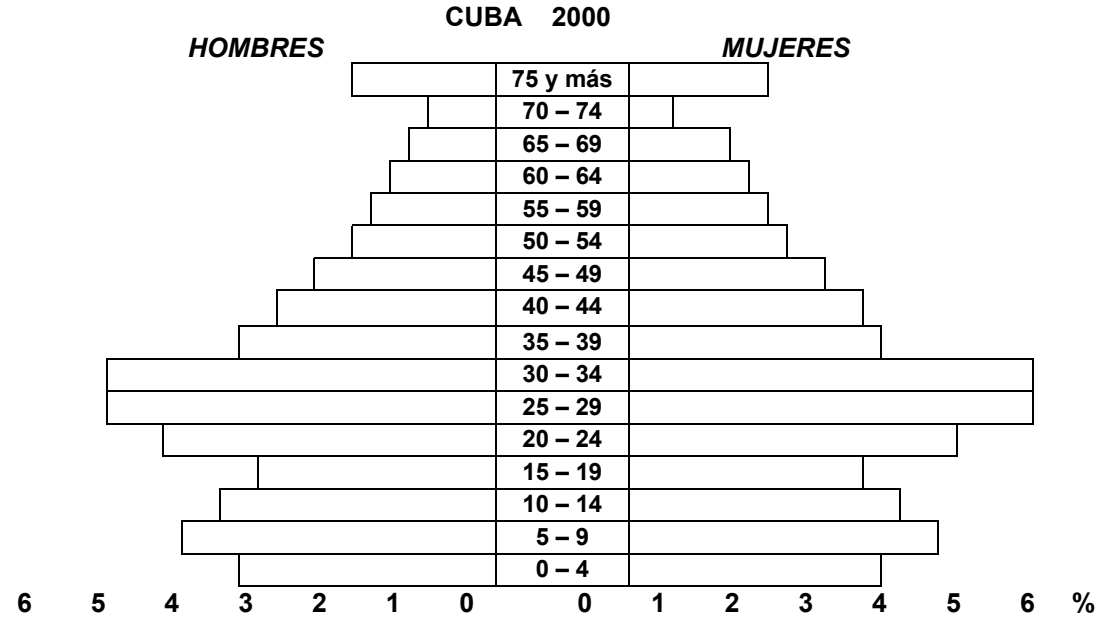


Caracteristicas

- ✓ Ni tan ancha como la joven, ni tan estrecha como la vieja
- ✓ La disminución hacia la cúspide se va haciendo mas uniforme
- ✓ La altura comienza a crecer
- ✓ Población en transición a envejecer, se eleva la esperanza de vida
- ✓ Comienza a disminuir la natalidad y la mortalidad
- ✓ Incremento paulatino y sistemático del nivel de salud
- ✓ Corresponde a países en vías de desarrollo

Evolucion de la piramide de la poblacion cubana





Caracteristicas:

- Lento crecimiento de la población en últimos 30 años; descenso rápido e intenso de la tasa de fecundidad (129.5 principios década del 70 a 49.4 en 1998).
- Descenso de la mortalidad fundamentalmente la infantil (menores de un año); De 38.7 por 1000 nacidos vivos en 1970 a 7.2 en el año 2000. (2002 fue de 6.5).
- Envejecimiento poblacional: En 1970 la población de 60 años y + era el 9%, en el 2000 es el 12.9%. Aumento sistemático de la esperanza de vida al nacer – en 1999 fue de 74.83 años. Amplio desarrollo y cobertura de los servicios y programas de salud.

Concepto de estadísticas de salud. Estadísticas utilizadas en el campo de la salud

Se entiende por estadísticas de salud a todas las estadísticas relacionadas con la salud. En ella se incluyen:

⇒ Estadísticas vitales:
  ⇒ Estadísticas de natalidad que miden los nacimientos de la población.
  ⇒ Estadísticas de mortalidad que miden el riesgo de morir que tiene la población.

⇒ Estadísticas de morbilidad que miden el riesgo de enfermar que tiene la población.

⇒ Estadísticas de recursos y servicios que miden la disponibilidad y utilización de los recursos y servicios de salud de que se dispone para la atención de la población.

⇒ Estadísticas de higiene y saneamiento que miden la situación higiénica y de saneamiento en el entorno físico y social de la población. (Muchos de los indicadores que se utilizan en este aspecto son indicadores socioeconómicos).

Estadísticas de Natalidad. Principales conceptos. Medidas de fecundidad.

Algunos conceptos necesarios

Natalidad

Natalidad se define como la frecuencia de nacimientos vivos en una población.

Fecundidad

Fecundidad se define como la procreación efectiva en una población.

Fertilidad

Fertilidad se define como la capacidad de procreación de una población, o sea la capacidad que tiene una población de reproducirse. El elemento más significativo a tener en cuenta en la fertilidad son las mujeres en edad fértil o sea las mujeres entre 15 y 49 años.

Nacido vivo

Es la expulsión o extracción completa del cuerpo de la madre, independientemente de la duración del embarazo de un producto de la concepción que después de dicha separación, respire o de cualquier otra señal de vida tal como: palpitaciones del corazón, pulsaciones del cordón umbilical, ó movimientos de los músculos de contracción voluntaria, tanto si se ha cortado ó no el cordón umbilical y esté o no desprendida la placenta.

Medidas de fecundidad

Tasa bruta de natalidad (TBN). También se le conoce como Tasa General de Natalidad, Tasa Absoluta de Natalidad, ó simplemente Tasa de Natalidad.

- Mide la frecuencia de nacimientos vivos con relación a la población total en un determinado periodo de tiempo.

- Se calcula por la fórmula siguiente:

$$\text{TBN} = \frac{\text{Número de nacidos vivos}}{\text{Población media total}} \times 1000 \text{ (mil)}$$

- El resultado se expresa por 1000 habitantes

Población media total: Es la que existe a mitad del periodo, día central del año (1ro de Julio).

Características

- Forma más sencilla de medir el crecimiento de la población.

- Tiene como limitante que toma en consideración a toda la población incluyendo a aquellos que no pueden procrear como niños y ancianos.

- Es un indicador poco específico siendo altamente influenciado por la composición de la población por edad y sexo.

- No es recomendable para comparar entre países a no ser que se usen tasas estandarizadas.

Ejemplo de cálculo de Tasa Bruta de Natalidad.

En una ciudad, la población media total de un año determinado fue de 250,000 habitantes. El número de nacidos vivos fue de 4000 en ese año.

La Tasa Bruta de Natalidad fue:

$$\text{TBN} = \frac{4000}{250{,}000} \times 1000 = 16$$

TBN = 16 nacidos vivos por cada 1000 habitantes.

Tasa de Fecundidad General (TFG). También se conoce como Tasa General de Fecundidad.

- Mide la frecuencia de nacimientos vivos con relación a la población de mujeres de 15 a 49 años de edad en un determinado periodo de tiempo.

- Se calcula por la fórmula siguiente:

$$\text{TFG} = \frac{\text{Número de nacidos vivos}}{\text{Población femenina de 15 a 49 años (Mujeres en edad fértil)}} \times 1000 \text{ (mil)}$$

- El resultado se expresa por 1000 mujeres de 15 a 49 años ó mujeres en edad fértil.

Características

- Elimina la limitante de la Tasa Bruta de Natalidad.

- Se utiliza para hacer comparaciones entre países.

Ejemplo de cálculo de la Tasa de Fecundidad General.

En una ciudad, la población media total de un año determinado fue de 250,000 habitantes, de los cuales las mujeres de 15 a 49 años fueron 50,000. El número de nacidos vivos fue de 4000 en ese año.

La Tasa de Fecundidad General fue:

$$\text{TFG} = \frac{4000}{50{,}000} \times 1000 = 80$$

TFG = 80 nacidos vivos por cada 1000 mujeres de 15 – 49 años de edad.
80 nacidos vivos por cada 1000 mujeres en edad fertil.

Tasa de Fecundidad Específica por edad (TFE).
Características

- Mide la fecundidad de las mujeres pero por grupos de edades.
- Por lo general se utilizan grupos de edad con intervalo de cinco años (15-19; 20-24; 25-29; 30-34..... 45-49)
- Permite el análisis detallado de la fecundidad y conocer el aporte de cada edad o grupos de edad a la fecundidad.
- Tiene como ventaja que elimina las diferentes proporciones de mujeres en las distintas edades.
- Tiene como dificultad la obtención de datos del nacimiento según la edad de la madre.
- Facilita establecer programas de salud específicos atendiendo a la tasa en los diferentes grupos de edades, sobre todo en los grupos de riesgo; mujeres jóvenes (15-19 años) y mujeres añosas (35-39; 40-44 y 45-49 años).

- Se calcula por la fórmula siguiente:

$$\text{TFE} = \frac{\text{Número de nacidos vivos a una edad determinada (grupo de edad) de la madre}}{\text{Población femenina de esa edad}} \text{ X 1000 (mil)}$$

- El resultado se expresa por 1000 mujeres de una edad determinada.

Los niveles máximos de fecundidad se observan entre 20-24 años de edad y los mínimos entre 45-49 años.

Ejemplo de cálculo de Tasa de Fecundidad Específica por edades.
En una ciudad determinada la población de mujeres de 15-19 años es de 5000; de 20-24 años es de 6000 y de 45 a 49 años es de 2000. El número de nacidos vivos entre madres de cada grupo de edad son 800, 1500 y 300 respectivamente.

Las tasas de fecundidad especifica por edades fueron:

Grupos de edades	**15-19**	**20-24**	**45-49**
Número de mujeres	5000	6000	2000
Número de nacidos vivos	800	1500	300

$$\text{TFE} = \underset{\textbf{15-19}}{\frac{800}{5000}} \text{ X 1000} \quad \underset{\textbf{20-24}}{\frac{1500}{6000}} \text{ X 1000} \quad \underset{\textbf{45-49}}{\frac{300}{2000}} \text{ X 1000}$$

TFE = 160 por 1000 mujeres de 15 a 19 años | **250 por 1000 mujeres de 20-24 años** | **150 por 1000 mujeres de 45 a 49 años**

La tasa de Fecundidad Específica, como cualquier tasa específica se puede calcular con cualquier grupo poblacional. El único requerimiento es el cómputo del evento en los grupos poblacionales específicos. Por ejemplo si se quiere comparar la TFE en mujeres de 15 años de edad que viven en áreas rurales y las mujeres de 15 años de edad que viven en áreas urbanas de un territorio y en un periodo de tiempo determinado se procede de la manera siguiente:

En un determinado territorio el número de mujeres de 15 años de edad que viven en áreas rurales es 1500 y parieron 50 nacidos vivos; mientras que aquellas que viven en áreas urbanas son 2500 y parieron 60 nacidos vivos. Las Tasas de Fecundidad Especificas son:

$$\text{TFE en mujeres de 15 años de edad en áreas rurales} = \frac{\text{Número de nacidos vivos de mujeres de 15 años de edad de áreas rurales}}{\text{Mujeres de 15 años de edad de áreas rurales}} \times 1000$$

$$= \frac{50}{1500} \times 1000 = 33.3$$

TFE = 33.3 nacidos vivos por cada 1000 mujeres de 15 años de edad que viven en áreas rurales

$$\text{TFE en mujeres de 15 años de edad en áreas urbanas} = \frac{\text{Número de nacidos vivos de mujeres de 15 años de edad de áreas urbanas}}{\text{Mujeres de 15 años de edad de áreas urbanas}} \times 1000$$

$$= \frac{60}{2500} \times 1000 = 24$$

TFE = 24 nacidos vivos por cada 1000 mujeres de 15 años de edad que viven en áreas urbanas

Los grupos poblacionales para las tasas específicas se forman acorde a las necesidades de los indicadores de salud que se requieran. Se pueden hacer con cualquier característica de la población: edad, sexo, nivel educacional, estado ocupacional, nivel socioeconómico y cualquier otro, así como con combinaciones de ellos.

Medidas de mortalidad. Principales conceptos. Mortalidad infantil y sus componentes. Mortalidad materna. Mortalidad general. Mortalidad específica.

Importancia

Según la OMS **"las estadísticas de mortalidad son una de las fuentes principales de información en salud y en muchos países constituyen la información de salud más confiable"**. Por las estadísticas de mortalidad se puede conocer en otros, los siguientes aspectos:

- el nivel socioeconómico y condiciones ambientales de una sociedad;
- el estado de salud de la población;
- la cobertura de los servicios de salud;
- el cumplimiento de programas de salud tales como de inmunización y control de enfermedades transmisibles entre otros y
- en general el grado de desarrollo de una sociedad.

Por ejemplo:

En los países subdesarrollados la tasa de mortalidad infantil es alta reflejando pobres servicios de salud, baja cobertura de inmunización, deficiente atención pre y post natal y problemas nutricionales de la población. Las principales causas de muerte son las enfermedades transmisibles, diarrea y enfermedades gastrointestinales y enfermedades infecciosas agudas.

En países desarrollados la tasa de mortalidad infantil es baja reflejando adecuados servicios de salud, buena cobertura de inmunización y eficiente atención pre y post natal. Las principales causas de muerte son los problemas cardiovasculares, tumores malignos y afecciones cerebro-vasculares.

Algunos conceptos necesarios

Mortalidad

Es la acción de la muerte sobre una población.

Defunción o muerte

Es la desaparición permanente y total de todo signo de vida, en un momento posterior al nacimiento.

Defunción fetal

Es la muerte del producto de la concepción antes de la expulsión o extracción completa del cuerpo de la madre, independientemente de la duración del embarazo. Indica la defunción, la circunstancia que después de la separación, el feto no respire ni de cualquier otra señal de vida tal como: palpitaciones del corazón, pulsaciones del cordón umbilical o movimientos efectivos de los músculos de contracción voluntaria tanto si se ha cortado ó no el cordón umbilical y esté ó no desprendida la placenta.

Medidas de mortalidad
Se expresan por las Tasas de Mortalidad que miden el riesgo de morir que tiene una población.

Tasa bruta de mortalidad (TBM). También se le conoce como Tasa General de Mortalidad, Tasa Absoluta de Mortalidad, ó simplemente Tasa de Mortalidad.

- Mide la frecuencia de fallecidos con relación a la población total en un determinado periodo de tiempo; o sea, **mide el riesgo de morir que tiene una población en general**.

- Se calcula por la fórmula siguiente:

$$\text{TBM} = \frac{\text{Número total de fallecidos}}{\text{Población media total}} \quad \text{X 1000 (mil)}$$

- El resultado se expresa por 1000 habitantes.

Características

- Es la forma más sencilla de conocer el impacto de la muerte en una población.

- Es un indicador poco específico siendo altamente influenciado por la composición de la población por edad y sexo.

- No es recomendable para comparar entre países a no ser que se usen tasas estandarizadas.

Ejemplo de cálculo de Tasa Bruta de Mortalidad.

En una ciudad, la población media total de un año determinado fue de 200,000 habitantes. El número de fallecidos fue de 1500 en ese año.

La Tasa Bruta de Mortalidad fue:

$$\text{TBM} = \frac{1500}{200{,}000} \text{ X 1000 (mil)} = 7.5$$

TBM = 7.5 fallecidos por cada 1000 habitantes.

Tasas específicas de mortalidad

Miden el riesgo de morir que tiene una población o un grupo poblacional ya sea en forma general ó por una causa determinada. Estas tasas se calculan por sexo, edad, áreas geográficas, nivel educacional, estado socioeconómico y causas de muerte entre otras características, bien en forma individual o combinadas. Son más adecuadas para el conocimiento de la mortalidad ya que permiten un análisis detallado por diferentes grupos poblacionales y principales causas de muerte en esos grupos brindando información relevante para establecer programas adecuados de salud.

- Forma general de cálculo de Tasas Específicas de Mortalidad (TEM):

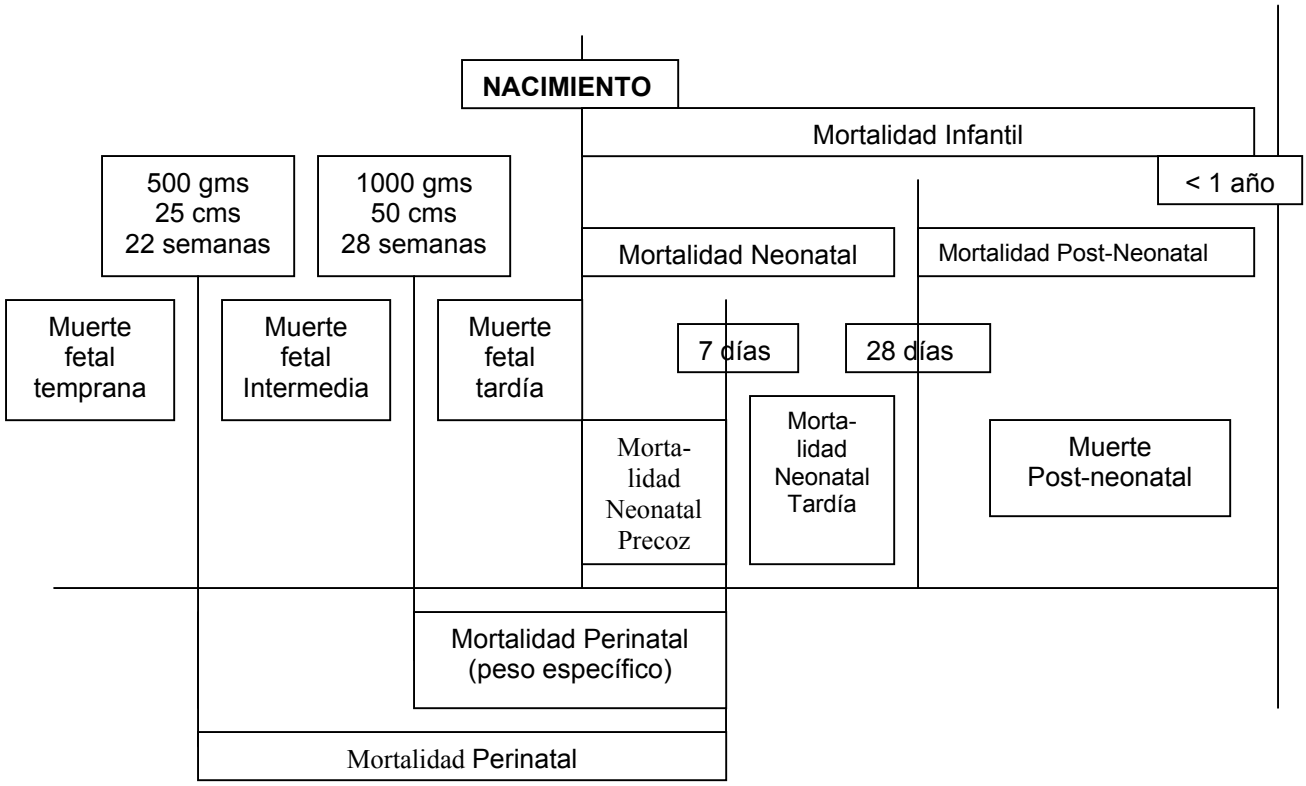
$$TBM = \frac{\text{Número de fallecidos en el grupo poblacional i}}{\text{Población del grupo i}} \quad X\ 10^{n}$$

Hay tasas de mortalidad específica muy importantes, que incluso son solicitadas por la OMS a todos los países. Ella son:

- Tasa de Mortalidad Infantil y sus componentes
- Tasa de Mortalidad Perinatal
- Tasa de Mortalidad Materna

Tasa de Mortalidad Infantil y sus componentes

La tasa de mortalidad infantil mide el riesgo de morir en los niños menores de un año. Los componentes de esta tasa según la OMS se exponen en el esquema siguiente:



El periodo que abarca la mortalidad infantil es desde el nacimiento hasta menos de 1 año. Se divide en:

- Periodo Neonatal (de nacimiento a menos de 28 días)
- Periodo Post-neonatal (de 28 días a menor de un año)

El periodo Neonatal se divide en:

- Neonatal Precoz (de nacimiento a 7 días completos)
- Neonatal Tardío (de más de 7 días a menos de 28 días completos)

La mortalidad post-neonatal es fácil de disminuir. Ella es provocada por diarrea, enfermedades gastrointestinales, infecciones respiratorias agudas y enfermedades transmisibles. Todas son fáciles de prevenir y controlar con adecuados programas y servicios de salud.

La mortalidad neonatal no es fácil de disminuir. Es provocada por enfermedades congénitas, membrana hialina, enfermedad hemolítica del recién nacido y otras enfermedades difíciles de prevenir y controlar.

- Modo de cálculo:
  - El denominador será siempre la cantidad de nacidos vivos en el periodo determinado.
  - Siempre se multiplica por 1000 (mil).
  - Por ende, la respuesta será por 1000 nacidos vivos.

$$\text{Tasa de Mortalidad Infantil (TMI)} = \frac{\text{Muertes en niños menores de un año}}{\text{Nacidos vivos}} \times 1000$$

$$\text{Tasa de Mortalidad Neonatal} = \frac{\text{Muertes en niños menores de 28 días completos de vida}}{\text{Nacidos vivos}} \times 1000$$

$$\text{Tasa de Mortalidad Postneonatal} = \frac{\text{Muertes en niños de 28 días completos de vida hasta menores de un año}}{\text{Nacidos vivos}} \times 1000$$

$$\text{Tasa de Mortalidad Neonatal Precoz} = \frac{\text{Muertes en niños hasta 7 días completos de vida}}{\text{Nacidos vivos}} \times 1000$$

$$\text{Tasa de Mortalidad Neonatal Tardía} = \frac{\text{Muertes en niños de más de 7 días hasta 28 días completos de vida}}{\text{Nacidos vivos}} \times 1000$$

Ejemplo de cálculo de la Tasa de Mortalidad Infantil y sus componentes.
En una ciudad el número de nacidos vivos en un año determinado fue de 3000 y el número de muertes de niños menores de un año fue el siguiente:
muertes menores de un año----------200
de ellos:
muertes entre 28 días y menor de un año de vida---------- 150
muertes menores de 28 días de vida---------- 50
de ellos:
muertes hasta 7 días completos de vida---------- 35
muertes de más de 7 días hasta menor de 28 días de vida----- 15

$$\text{Tasa de Mortalidad Infantil (TMI)} = \frac{200}{3000} \times 1000 = 66.6$$

= 66.6 fallecidos menores de un año por cada 1000 nacidos vivos

$$\text{Tasa de Mortalidad Post-neonatal} = \frac{150}{3000} \times 1000 = 50$$

= 50 muertes postneonatales por cada 1000 nacidos vivos

$$\text{Tasa de Mortalidad Neonatal} = \frac{50}{3000} \times 1000 = 16.6$$

= 16.6 muertes neonatales por cada 1000 nacidos vivos

$$\text{Tasa de Mortalidad Neonatal Precoz} = \frac{35}{3000} \times 1000 = 11.6$$

= 11.6 muertes neonatales precoces por cada 1000 nacidos vivos

$$\text{Tasa de Mortalidad Neonatal Tardía} = \frac{15}{3000} \times 1000 = 5$$

= 5 muertes neonatales tardías por cada 1000 nacidos vivos

La mortalidad infantil se clasifica de la siguiente manera:

- **Muy alta** – 100 o más fallecidos menores de un año por cada 1000 nacidos vivos
- **Alta** – Entre 50 y 99 fallecidos menores de un año por cada 1000 nacidos vivos
- **Mediana** – Entre 30 – 49 fallecidos menores de un año por cada 1000 nacidos vivos
- **Baja** – Entre 15 – 29 fallecidos menores de un año por cada 1000 nacidos vivos
- **Muy baja** – Menor de 15 fallecidos menores de un año por cada 1000 nacidos vivos

Tasas de Mortalidad Perinatal

Según la OMS, el periodo perinatal **"comienza en 22 semanas completas, (154 días) de gestación, (el tiempo en que normalmente el peso del nacimiento es 500 gm.), y termina a los siete días completos después del nacimiento".**

Estas tasas miden el riesgo de morir alrededor del momento del parto. El estudio de la mortalidad durante este periodo es fundamental para evaluar la atención de salud recibida por la madre durante todo el embarazo, en específico en el último trimestre, y también la atención de salud recibida por el recién nacido durante la primera semana de vida.

- Modo de cálculo
  - El denominador será siempre el total de nacimientos, esto es la suma de los nacidos vivos con las muertes fetales en un periodo de tiempo determinado.
  - Siempre se multiplican por 1000.
  - El resultado se dará siempre por 1000 nacimientos.

$$\text{Tasa de Muertes Fetales} = \frac{\text{Muertes fetales}}{\text{Total de nacimientos}} \times 1000$$

$$\text{Tasa de Mortalidad Perinatal} = \frac{\text{Muertes fetales + muertes neonatales precoces}}{\text{Total de nacimientos}} \times 1000$$

$$\text{Tasa de Mortalidad Perinatal de (peso específico)} = \frac{\text{Muertes fetales con peso de 1000 g. ó más + muertes neonatales precoces de recién nacidos que pesaron 1000g. ómás al nacer}}{\text{Total de nacimientos que pesaron 1000g. o más}} \times 1000$$

Ejemplo de cálculo de las Tasas de Mortalidad Perinatal

En una ciudad el número de nacidos vivos es de 2500 en un año, el número de muertes fetales es de 1000 y el número de recién nacidos hasta 7 días completos de vida fallecidos es de 60.

$$\text{Tasa de Muerte Fetal} = \frac{1000}{2500 + 1000} \times 1000 = \frac{1000}{3500} \times 1000 = 285.7$$

= 285.7 muertes fetales por cada 1000 nacimientos.

$$\text{Tasa de Mortalidad Perinatal} = \frac{1000 + 60}{2500 + 1000} \times 1000 = \frac{1060}{3500} \times 1000 = 302.8$$

= 302.8 muertes perinatales por cada 1000 nacimientos.

Tasa de Mortalidad Materna
Mide el riesgo de morir que tiene una mujer embarazada.

Concepto de muerte materna:
Muerte materna es la muerte de una mujer durante el embarazo ó durante los 42 días después de terminado el embarazo, sin tener en cuenta la duración y el sitio del embarazo, de cualquier causa relacionada o agravada por el embarazo ó su manejo, pero no por causas accidentales.

- Modo de cálculo

$$\text{Tasa de Mortalidad Materna (TMM)} = \frac{\text{Muertes maternas}}{\text{Nacidos vivos}} \times 10^{n}$$

Los resultados se expresan por 10^{n} nacidos vivos.

- El valor de n depende del denominador de la fórmula. El objetivo es obtener al menos un entero en la respuesta. Una manera práctica de determinar el número al cual se eleva 10 es contando cuantos números tiene el denominador y n sería esa cantidad menos 1.

Ejemplo de cálculo de la Tasa de Mortalidad Materna
En una ciudad el número de nacidos vivos en un año fueron 50,000 y las muertes maternas fueron 40. La tasa de Mortalidad Materna es:

$$TMM = \frac{40}{50{,}000} \times 10^{4} = \frac{40}{50{,}000} \times 10{,}000 = 8$$

= 8 muertes maternas por cada 10,000 nacidos vivos.

Otras tasas específicas de Mortalidad
- Todas se calculan por 10^{n}.
- Se calculan por la siguiente fórmula general

$$TEM = \frac{\text{Número de fallecidos en el grupo poblacional i}}{\text{Población del grupo i}} \times 10^{n}$$

Los resultados se expresan por 10^{n} habitantes del grupo i.

Estas tasas son muy necesarias en Medicina y Salud Pública para conocer detalles específicos de la mortalidad en la población.

Ejemplo de cálculo de Tasas Específicas de Mortalidad

En una ciudad la población media total es de 300,000 habitantes, la población media de sexo femenino es de 151,000 y la de niños menores de 14 años es 40,000. El número total de fallecidos de tuberculosis (TB) es 150, de ellos 60 fueron del sexo femenino y 100 en niños menores de 14 años. Se pueden calcular las tasas específicas de mortalidad por TB siguientes:

$$\text{Tasa de Mortalidad por TB} = \frac{\text{Total de fallecidos por TB}}{\text{Población media total}} \times 10^n$$

$$= \frac{150}{300{,}000} \times 10^5 = \frac{150}{300{,}000} \times 100{,}000 = 50$$

= 50 muertes por TB por cada 100,000 habitantes.

$$\text{Tasa de Mortalidad por TB en sexo femenino} = \frac{\text{Muertes por Tb en sexo femenino}}{\text{Población media de sexo femenino}} \times 10^n$$

$$= \frac{60}{151{,}000} \times 10^5 = \frac{60}{151{,}000} \times 100{,}000 = 39.7$$

= 39.7 muertes por TB en el sexo femenino por cada 100,000 habitantes del sexo femenino.

$$\text{Tasa de Mortalidad por TB en niños < 14 años} = \frac{\text{Muertes por Tb en < 14 años}}{\text{Población media < 14 años}} \times 10^n$$

$$= \frac{100}{40{,}000} \times 10^4 = \frac{100}{40{,}000} \times 10{,}000 = 25$$

= 25 muertes por TB en niños < 14 años por cada 10,000 habitantes < de 14 años.

Analizando las tasas vemos que el grupo más afectado es el de niños < de 14 años. En efecto tiene la mayor tasa 25 por 10,000 mientras que en la población total es de 50 por 100,000 y en el sexo femenino es de 39,7 por 100,000.

Medidas de morbilidad. Principales conceptos. Incidencia, Prevalencia, Letalidad. Morbilidad conocida y desconocida.

Algunos conceptos necesarios

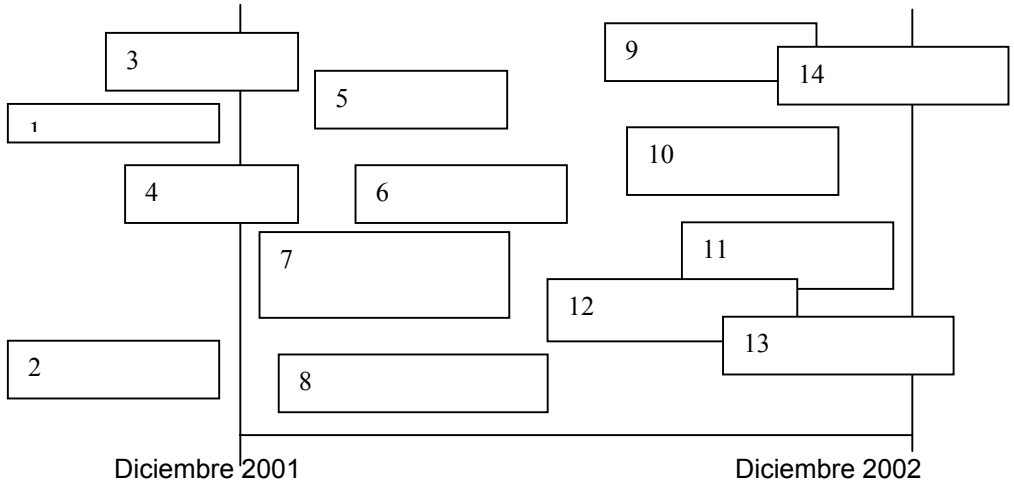
Morbilidad
Es la acción de las enfermedades sobre una población.

Incidencia
Es el número de casos nuevos de una enfermedad dada que ocurren en un territorio en un periodo de tiempo específico.

Prevalencia
Es el número total de casos de una enfermedad dada que existen en un territorio en un periodo de tiempo específico.

Letalidad ó Fatalidad
Es el número de personas que fallecen por una enfermedad determinada con relación al número total de personas que padecen la enfermedad en un territorio en un periodo de tiempo específico.

- La diferencia entre incidencia y prevalencia se presenta en el siguiente esquema:



INCIDENCIA AÑO 2002 = 5; 6; 7; 8; 9; 10; 11; 12; 13; 14 = 10 CASOS

PREVALENCIA AÑO 2002 = 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14 = 12 CASOS

Algunas características de la morbilidad
Se incluyen en el concepto de morbilidad los defectos físicos, traumatismos y secuelas e inhabilidades de enfermedades, malformaciones congénitas y traumas. Las estadísticas de morbilidad se requieren para un conocimiento completo del cuadro de salud de una población, pero son difíciles de obtener dado los siguientes factores:

- Muchos enfermos no concurren a los servicios de salud a menos que se encuentren en muy mal estado.
- Muchas enfermedades afectan a parte de la población con síntomas leves o manifestaciones subclínicas no siendo detectadas ni por el propio enfermo.
- Falta de registros de morbilidad ambulatoria en muchos países y en aquellos que existe poca confiabilidad de la información.

Se considera que la morbilidad es como un témpano de hielo (iceberg) donde la **morbilidad conocida** es solamente la pequeña parte del témpano de hielo que sobresale del agua. Son las enfermedades que podemos conocer. La **morbilidad desconocida u oculta** es la gran masa de hielo que se encuentra sumergida. Son las enfermedades imposibles de detectar.

Medidas de morbilidad

- Las medidas de morbilidad son las siguientes:
  - Tasa de incidencia
  - Tasa de prevalencia
  - Indice de letalidad

Tasa de incidencia
Mide el riesgo de contraer una enfermedad.

- Modo de cálculo

$$\text{Tasa de incidencia} = \frac{\text{Número de casos nuevos de una enfermedad dada}}{\text{Población media total}} \times 10^n$$

Los resultados se expresan por 10^n habitantes.

Tasa de prevalencia
Mide el riesgo de padecer una enfermedad.

- Modo de cálculo

$$\text{Tasa de prevalencia} = \frac{\text{Número total de casos de una enfermedad dada}}{\text{Población media total}} \times 10^n$$

Los resultados se expresan por 10^n habitantes

Indice de letalidad
Mide el riesgo de morir que tienen los enfermos de una enfermedad dada.

- Modo de cálculo

$$\text{Indice de letalidad} = \frac{\text{Número de fallecidos por una enfermedad dada}}{\text{Número total de casos de la enfermedad dada}} \times 100$$

Los resultados se expresan en porciento.

Ejemplo de cálculo de Tasas de Morbilidad
En una ciudad determinada se tiene la información siguiente para un periodo de tiempo:
Total de casos de TB ----------300
Nuevos casos de TB ---------- 25
Muertes por TB ---------- 15
Total de casos de TB en grupo de edad 25-34 años ------- 50
Población media total ---------- 200,000
Población media en grupo de edad 25-34 años ---------- 10,000

$$\text{Tasa de Incidencia de TB} = \frac{25}{200{,}000} \times 10^5 = \frac{25}{200{,}000} \times 100{,}000 = 12.5$$

= 12.5 nuevos casos de TB por cada 100,000 habitantes.

$$\text{Tasa de Prevalencia de TB} = \frac{300}{200{,}000} \times 10^5 = \frac{300}{200{,}000} \times 100{,}000 = 150$$

= 150 casos de TB por cada 100,000 habitantes.

$$\text{Indice de letalidad de TB} = \frac{15}{300} \times 100 = 5$$

= 5 %

$$\text{Tasa de prevalencia de TB en personas de 25-34 años} = \frac{50}{10{,}000} \times 10^4 = \frac{50}{10{,}000} \times 10{,}000 = 50$$

= 50 casos de TB en el grupo de edad de 25-34 años por cada 10,000 habitantes de 25-34 años.

El ejemplo por sí mismo muestra la importancia de las estadísticas de morbilidad en Medicina y Salud Pública. Es necesario utilizar las medidas de morbilidad para conocer el comportamiento de determinada enfermedad, detectar el ó los grupos poblacionales más afectados y determinar en que grupos poblacionales son más frecuentes los desenlaces fatales.

Indicadores que miden los recursos y servicios disponibles y su grado de utilización. Médicos por mil habitantes. Camas por mil habitantes. Embarazadas captadas en primer trimestre. Mortalidad hospitalaria bruta y neta.

Estos indicadores miden la disponibilidad y utilización de los recursos y servicios de salud que se tienen para la atención de la población.

Los indicadores se expresan de diferente manera:

- frecuencia absoluta de una característica dada
- porcentajes, índices, razones, proporciones
- una característica por n habitantes

A manera de ejemplo mencionemos algunos de los más utilizados:

- Cantidad de unidades asistenciales por tipo de unidad
- Cantidad de personal de la salud por especialización
- Cantidad de camas asistenciales por habitantes
- Cantidad de camas de asistencia social por habitantes
- Cantidad de consultas por habitantes
- Cantidad de médicos por habitantes
- Cantidad de enfermeras por habitantes
- Cantidad de técnicos medios de la salud por habitantes
- Razón de habitantes por médico
- Ingresos hospitalarios por habitantes
- Cantidad de consultas médicas
- Porcentaje de consultas externas
- Porcentaje de consultas en cuerpo de guardia
- Indice ocupacional en hospitales
- Embarazadas captadas en el primer trimestre
- Mortalidad hospitalaria bruta y neta

Médicos por mil habitantes

Mide la cantidad de médicos por cada mil habitantes, o sea la disponibilidad de médicos para la atención de la población.

- Modo de cálculo

$$\text{Médicos por mil habitantes} = \frac{\text{Total de médicos}}{\text{Población media total}} \times 1000$$

Camas por mil habitantes
Mide la cantidad de camas asistenciales ó de asistencia social ó de las dos por cada mil habitantes. Es un indicador de las posibilidades que tiene la población de atención médica especializada ó de asistencia social en centros asistenciales.

- Modo de cálculo

$$\text{Camas por mil habitantes: } \frac{\text{Número de camas}}{\text{Población media total}} \times 1000$$

Consultas médicas por mil habitantes
Mide la cantidad de consultas realizadas a la población por cada mil habitantes. Expresa las veces que la población ha acudido a ver al facultativo a una institución de salud por cada mil habitantes.

- Modo de cálculo

$$\text{Consultas médicas por mil habitantes: } \frac{\text{Número de consultas}}{\text{Población media total}} \times 1000$$

Embarazadas captadas en el primer trimestre
Mide la cantidad de embarazadas que tienen una captación precoz por el médico de la familia.

- Modo de cálculo

Es el número de embarazadas con tres meses o menos de gestación que se encuentran bajo control del médico de la familia.

Mortalidad hospitalaria bruta y neta.
Mortalidad Bruta
Mide la mortalidad general que se ha presentado en un hospital.

- Modo de cálculo

$$\text{Mortalidad bruta} = \frac{\text{Todos los fallecidos del hospital en un periodo de tiempo}}{\text{Cantidad de egresos del hospital en el mismo periodo de tiempo}} \times 100$$

Mortalidad Neta
Mide la mortalidad después de 48 horas de ingresado. Es un indicador de la eficiencia hospitalaria.

- Modo de cálculo

$$\text{Mortalidad neta} = \frac{\text{Los fallecidos del hospital con más de 48 horas de ingreso en un periodo de tiempo}}{\text{Cantidad de egresos del hospital en el mismo periodo de tiempo}} \times 100$$

Las cifras aceptadas como normales son hasta 4% para la mortalidad bruta y entre 2.5% y 3% para la mortalidad neta.

Indicadores que miden los aspectos de higiene y saneamiento.
Estos indicadores miden la situación higiénica y de saneamiento en el entorno físico y social de la población. (Muchos de los indicadores que se utilizan en este aspecto son indicadores socioeconómicos).

Por lo general se expresan en porcentajes

Entre ellos encontramos los siguientes:
- Porcentaje de población con abasto de agua potable
- Porcentaje de población con abasto de agua potable por conexión domiciliaria
- Porcentaje de población urbana con abasto de agua potable por conexión domiciliaria
- Porcentaje de población rural con abasto de agua potable por conexión domiciliaria
- Porcentaje de población urbana con letrinas
- Porcentaje de población rural con letrinas

INDICE

Capítulo 8
INTRODUCCION A LA ESTADISTICA INFERENCIAL

Autor: Dr. Emilio Morales Jiménez

SUMARIO

- *Conceptos de variable aleatoria y probabilidad.*
- *Modelo teórico de distribución o ley de una variable aleatoria.*
- *Modelo de la distribución normal. Parámetros de esta distribución. Propiedades.*
- *La distribución normal estándar. Ejemplos de aplicación.*
- *Tabla de Z.*

INTRODUCCION

La estadística inferencial es la parte de la Estadística que permite tomar decisiones con relación al universo o población basándose en valores calculados con datos de muestras de ese universo o población. En Medicina, al igual que en otras ciencias, las técnicas de la estadística inferencial son muy útiles. Por lo general, el médico atiende a un grupo de pacientes, no a todo el universo de pacientes. Las conclusiones que pueda obtener estudiando a esos pacientes no se pueden generalizar al universo de pacientes a menos que utilice las técnicas de la estadística inferencial.

Para introducirnos en el campo de la estadística inferencial se requieren conocer algunos conceptos y entender algunos aspectos relacionados con las diferentes técnicas que se aplican. Estos son los conceptos de variable aleatoria, probabilidad, modelo teórico de distribución de una variable aleatoria, la distribución normal y la distribución normal estándar.

Conceptos de variable aleatoria y probabilidad

La Medicina es una ciencia probabilística, no es una ciencia determinística. Aunque el médico no sea un matemático, en su diario quehacer utiliza las probabilidades. Por lo general el médico dice, es ***"probable"*** que el paciente presente ***"determinada enfermedad"*** y es ***"muy probable"*** que se cure siguiendo un tratamiento determinado. Intuitivamente, los médicos utilizan las probabilidades basándose en sus propios conocimientos, experiencias previas y lo que han aprendido en libros, revistas y experiencias de otros médicos.

Algunos ejemplos nos ayudarán a entender por que la Medicina no es una ciencia determinística.

1. Suponga que un paciente que padece cefalea migrañosa se alivia el dolor con dos aspirinas. Teniendo en cuenta este paciente, no se puede decir que todo paciente que padece migraña aliviará su dolor con aspirina.

2. Si un longevo de 105 años ha fumado durante toda su vida y esta obeso; ¿Se puede decir que la obesidad y el hábito de fumar son buenos para vivir una larga vida? Por supuesto que no.

3. Si un paciente viene con dolor en hipocondrio derecho que irradia a la espalda y tiene intolerancia a las grasas, chocolate y huevos; ¿Se puede asegurar que padece de litiasis vesicular? Por supuesto que no, ese diagnóstico es una posibilidad, pero otras enfermedades pueden tener síntomas similares.

Luego, en Medicina, hechos similares no determinan el mismo resultado en todos los casos. Por esa razón se dice que en Medicina, a veces $2 + 2 \neq 4$, porque no es una ciencia determinista. Medicina es una ciencia probabilística. Esto se debe a muchos factores entre los cuales están:

- La Medicina se refiere a seres humanos (ser bio-psico-social).
- No existen dos seres humanos exactamente iguales.
- Para establecer un diagnóstico es necesario recoger información (síntomas y signos) que varían de una persona a otra.
- Los síntomas y signos (variables) valorados y evaluados no tienen el mismo valor en todos los pacientes.

De hecho, no se puede conocer de antemano el valor exacto de una variable cualquiera en un paciente determinado aunque las condiciones de ese paciente sean similares a la de otros pacientes. Por ejemplo, supongamos que tenemos un grupo de 10 personas sanas de la misma edad, sexo y condiciones de vida. A todos le hacemos un examen de hemoglobina y veremos que cada cual tiene un valor diferente.

Tomemos otro ejemplo. Si medimos la talla de 20 estudiantes sanos de la misma edad y sexo, cuyos padres tengan talla similar y el mismo estado socioeconómico, veremos que los estudiantes tendrán tallas diferentes.

La causa de esas diferencias individuales es porque las variables que evaluamos en Medicina son variables aleatorias. Considerando las cuestiones ya señaladas una **VARIABLE ALEATORIA es aquella variable que no se puede predecir su valor aunque conozcamos las condiciones que la rodean y que no siempre toma el mismo valor a pesar de que esas condiciones se mantengan estables.**

Los diferentes valores individuales de las variables aleatorias pueden parecer incomprensibles o anárquicos, pero cuando mediante métodos apropiados se analiza una gran cantidad de datos nos damos cuenta que no es así observándose regularidades que se expresan por leyes estadísticas. La incertidumbre del valor individual se encuentra enmarcado en un rango de posibles valores. Estos rangos son muy importantes y útiles en Medicina. Cada variable, por ejemplo, glicemia, colesterol, urea, hemoglobina entre muchas otras, tienen un rango de valores posibles que indican la normalidad de ese valor. Si el valor de la variable está por debajo o por encima del rango de normalidad consideramos entonces que existe una anormalidad, o sea un valor patológico.

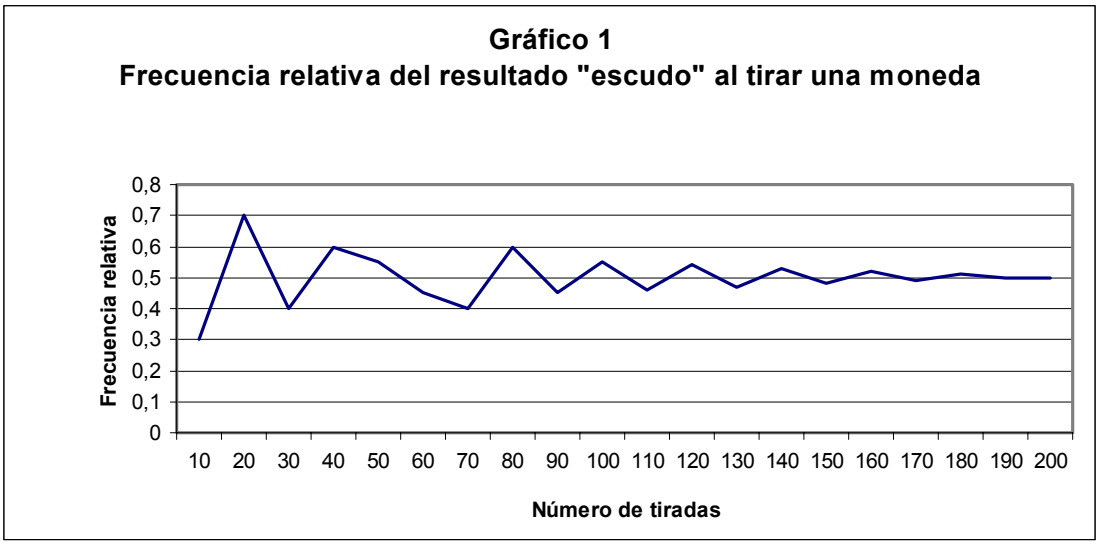
Resumiendo, se puede asegurar que un valor individual de una variable aleatoria no se puede predecir, que es imposible de conocer realmente a menos que se haga la medición correspondiente. Sin embargo, por otro lado, si podemos predecir valores colectivos. Para analizar esta cuestión veamos algunos ejemplos.

El primer ejemplo es muy sencillo y lo puede realizar cualquiera si tiene la suficiente paciencia. Nos referimos a tirar una moneda balanceada al aire y ver por que cara cae, escudo o estrella. En una sola tirada no se puede predecir que cara estará hacia arriba, el resultado es aleatorio. No obstante, si hacemos un experimento consistente en tirar una moneda balanceada 10 veces anotando cuantas veces la cara "escudo" cae hacia arriba y lo repetimos incrementando 10 tiradas cada vez hasta unas 200 tiradas y hacemos una tabla de los resultados tendremos algo parecido a lo siguiente.

TABLA 1

NUMERO DE TIRADAS	FRECUENCIA ABSOLUTA DE "ESCUDO"	FRECUENCIA RELATIVA DE "ESCUDO"
10	3	.3
20	14	.7
30	12	.4
40	24	.6
50	28	.55
60	27	.45
70	28	.4
80	48	.6
90	41	.45
100	55	.55
110	51	.46
120	65	.54
130	61	.47
140	74	.53
150	72	.48
160	83	.52
170	83	.49
180	92	.51
190	95	.5
200	100	.5

Si hacemos un gráfico colocando en el eje de la "Y" el valor de la frecuencia relativa y en el eje de la "X" el número de tiradas tendremos algo parecido al gráfico 1.



Es fácil de ver que a medida que la cantidad de tiradas aumenta, la frecuencia relativa de que la cara “escudo” caiga hacia arriba tiende a estabilizarse alrededor de un valor específico, en efecto, 0.5.

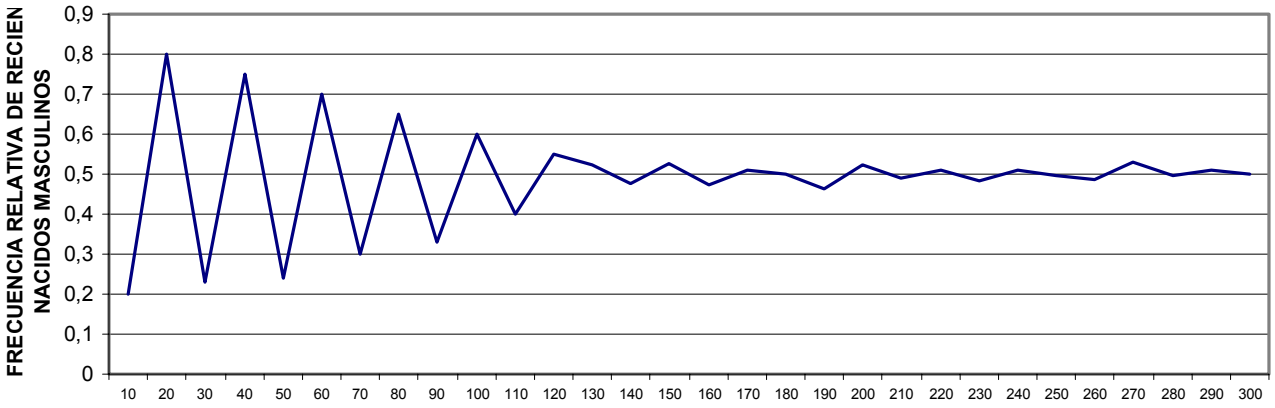
Otro ejemplo, más relacionado con la Medicina, se puede hacer con la variable sexo en recién nacidos. Si se hace una investigación en un Hospital Materno para conocer la proporción de recién nacidos masculinos tomando grupos de nacimientos e incrementando 10 en cada grupo hasta llegar a 300 recién nacidos tendremos un resultado parecido al que se presenta en la tabla 2.

TABLA 2

GRUPO	NACIMIENTOS	RECIEN NACIDOS MASCULINOS	FRECUENCIA RELATIVA DE MASCULINOS	GRUPO	NACIMIENTOS	RECIEN NACIDOS MASCULINOS	FRECUENCIA RELATIVA DE MASCULINOS
1	10	2	0.2	16	160	76	0.475
2	20	16	0.8	17	170	87	0.511
3	30	7	0.23	18	180	90	0.5
4	40	30	0.75	19	190	88	0.463
5	50	12	0.24	20	200	105	0.525
6	60	42	0.7	21	210	103	0.490
7	70	21	0.3	22	220	112	0.51
8	80	52	0.65	23	230	111	0.482
9	90	30	0.33	24	240	122	0.51
10	100	60	0.6	25	250	124	0.496
11	110	44	0.4	26	260	127	0.488
12	120	66	0.55	27	270	143	0.530
13	130	68	0.523	28	280	139	0.496
14	140	67	0.478	29	290	148	0.51
15	150	79	0.527	30	300	151	0.5

Con estos datos se puede hacer un gráfico al igual que se hizo anteriormente.

GRAFICO 2
Recién Nacidos Masculinos



En el gráfico se aprecia que a medida que el número de nacimientos incrementa la frecuencia relativa de recién nacidos masculinos tiende a estabilizarse alrededor de cierto valor, en efecto, 0.5.

La tendencia a estabilizarse de la frecuencia relativa alrededor de un valor determinado cuando el número de observaciones de la variable aleatoria se incrementa y es lo suficientemente grande se conoce como la **"Ley de Regularidad Estadística"**. Debido a esta ley podemos predecir un resultado colectivo. En ambos ejemplos podemos decir: la cara **"escudo"** al tirar una moneda debe aparecer con una frecuencia relativa de 0.5 y cuando nace un niño la frecuencia relativa del sexo masculino es 0.5. **Sin embargo, es importante conocer que a pesar de poder predecir el resultado colectivo de una variable aleatoria, es imposible predecir el resultado de un caso individual.** En efecto, si se tira una moneda una vez es imposible predecir si caerá la cara **"escudo"** hacia arriba en esa tirada y si una mujer está embarazada no se puede asegurar que el sexo del feto sea masculino.

La predicción del resultado colectivo utilizando la frecuencia relativa de una variable aleatoria es la base de lo que llamamos probabilidad. Luego, una definición de probabilidad sería: **El valor al cual se acerca la frecuencia relativa a medida que el número de observaciones aumenta y que es único para cada resultado de una variable aleatoria se le llama probabilidad del resultado.**

Las características de las probabilidades son:

- Como la frecuencia relativa es un número real entre 0 y 1, la probabilidad es también un número entre 0 y 1, luego, **$0 \leq p \leq 1$**, donde **p = probabilidad**.
- Cuando la probabilidad es 0, **(p=0),** el resultado nunca ocurrirá; cuando la probabilidad es 1, **(p=1)**, el resultado siempre ocurrirá. Sin embargo, esos valores son muy poco comunes.
- Lo que nos encontramos en la práctica diaria es que cuando la probabilidad está cerca de 1 (por ejemplo, p=0.8; p=0.91, etc.) se considera que el resultado tendrá alta probabilidad de ocurrencia y por el contrario cuando está cerca de 0 (por ejemplo, p=0.2; p=0.05; etc.) el resultado tiene poca probabilidad de ocurrencia.

MODELO TEÓRICO DE DISTRIBUCIÓN O LEY DE UNA VARIABLE ALEATORIA.

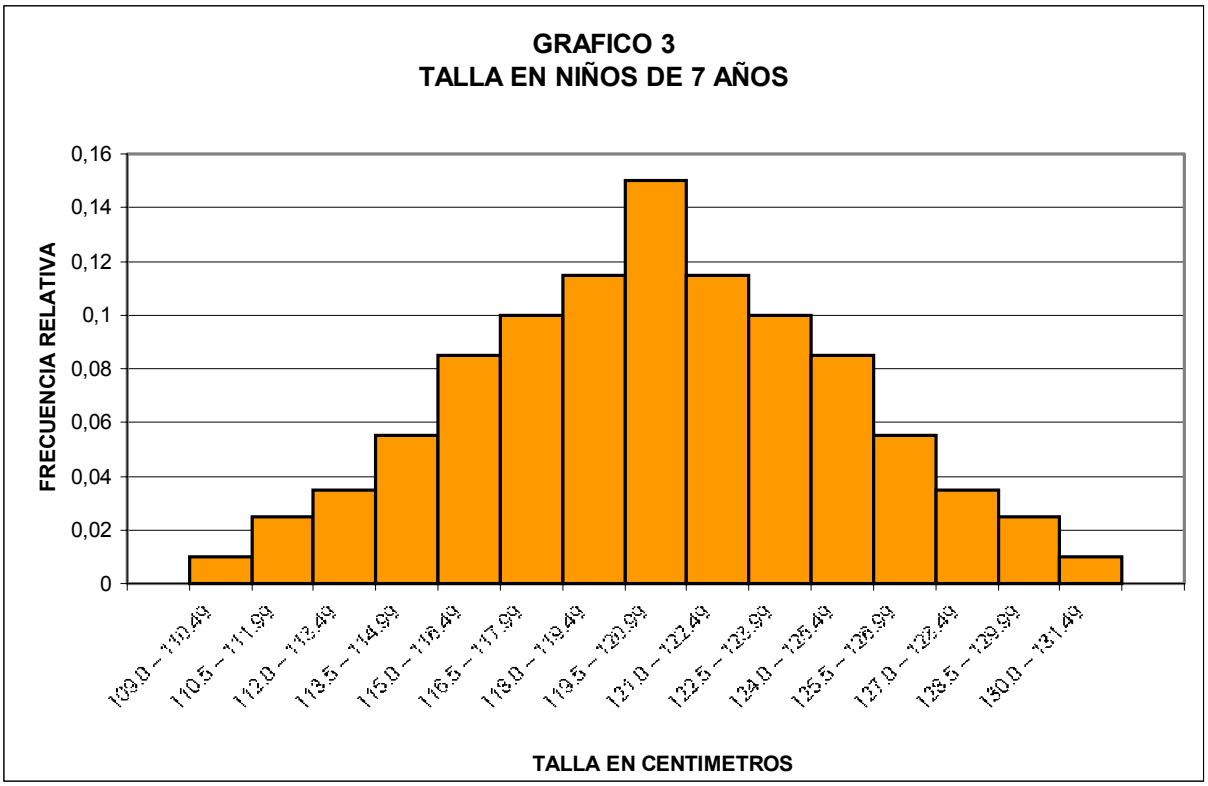
Antes de desarrollar este punto es necesario conocer la simbología que utilizaremos y su significado en las explicaciones subsiguientes y en las técnicas de estadística inferencial. De ahora en adelante nos referiremos a muestras y poblaciones o universos existiendo diferentes símbolos y nombres para cada una de ellas como se muestra en la tabla siguiente:

TABLA 3

ITEM	MUESTRA	UNIVERSO O POBLACION
Distribución de frecuencias	Distribución de frecuencias empíricas	Distribución de frecuencias teóricas
Cualquier medida de estadística descriptiva (media aritmética, mediana, moda, etc.)	Estadígrafo	Parámetro poblacional
Media aritmética	$\overline{X}$	μ **(MIU)**
Varianza	s^2	σ^2 **(SIGMA)**
Desviación estándar	s	σ
Porcentaje o proporción	p	P

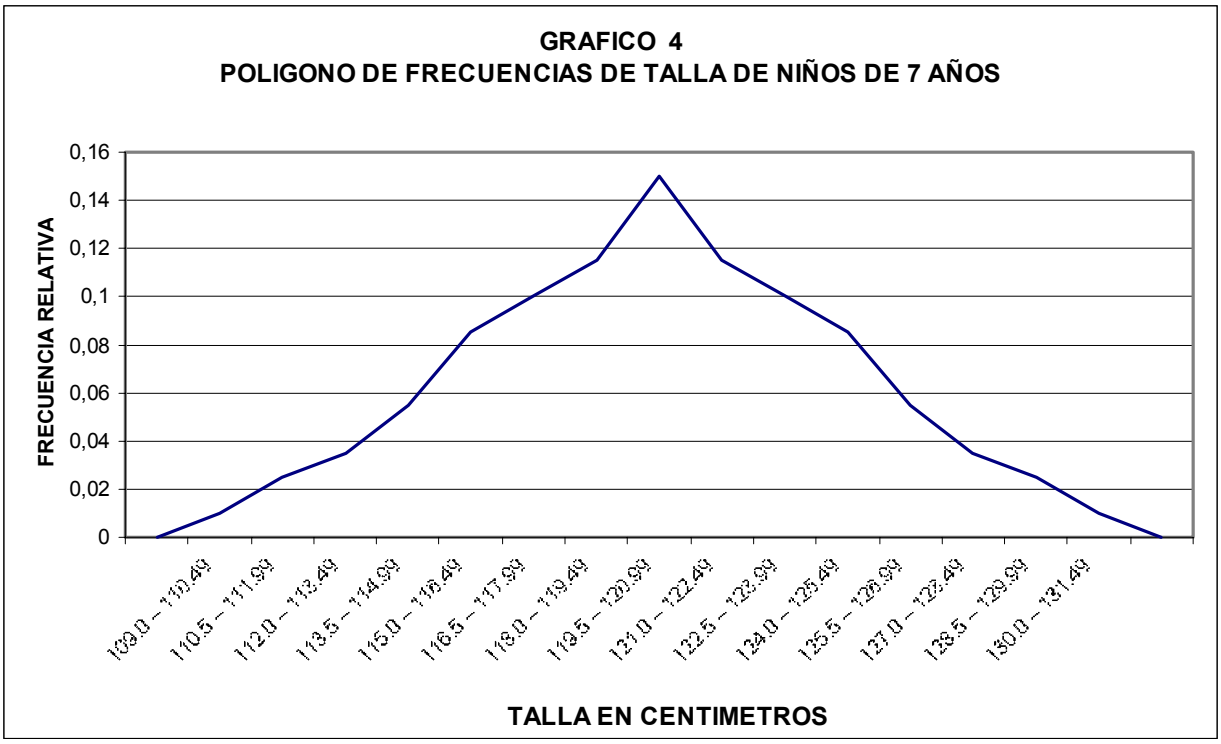
Para entender el modelo teórico de distribución o ley de una variable aleatoria veamos un ejemplo con una variable aleatoria, la talla. La talla es una variable numérica continua. Supongamos que de una población de niños varones de 7 años de edad se toma una muestra aleatoria de 200 niños (n=200). A cada niño se le mide la talla. La distribución de frecuencias de la talla se muestra en la tabla 4.

TABLA 4		
TALLA (CENTIMETROS)	**FRECUENCIA ABSOLUTA**	**FRECUENCIA RELATIVA**
109.0 – 110.49	2	0.010
110.5 – 111.99	5	0.025
112.0 – 113.49	7	0.035
113.5 – 114.99	11	0.055
115.0 – 116.49	17	0.085
116.5 – 117.99	20	0.100
118.0 – 119.49	23	0.115
119.5 – 120.99	30	0.150
121.0 – 122.49	23	0.115
122.5 – 123.99	20	0.100
124.0 – 125.49	17	0.085
125.5 – 126.99	11	0.055
127.0 – 128.49	7	0.035
128.5 – 129.99	5	0.025
130.0 – 131.49	2	0.010
TOTAL	200	1.000



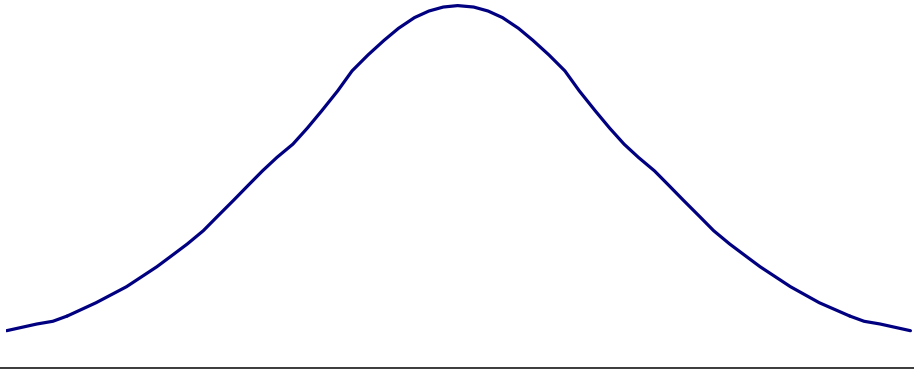
Para mostrar la distribución de frecuencias en forma gráfica hagamos un histograma con las frecuencias relativas. (La forma del histograma fuera la misma se hiciera con las frecuencias absolutas).

Si en lugar de un histograma hacemos un polígono de frecuencias tendríamos el siguiente gráfico.



Supongamos que la variable aleatoria talla la podemos medir con la precisión que deseemos, de forma tal que los intervalos de clase se puedan hacer tan pequeños como quisiéramos. Entonces, a medida que el tamaño muestral aumenta y la amplitud de los intervalos es lo suficientemente pequeña, la forma del polígono de frecuencias correspondiente al histograma, se hará más suave, más regular, sin escalones, aproximándose a la curva mostrada en el gráfico 5.

GRAFICO 5
DISTRIBUCION DE FRECUENCIAS RELATIVAS DE LA VARIABLE ALEATORIA



Esta curva es una idealización o el modelo de distribución de las frecuencias relativas con la cual se espera encontrar los diferentes valores de la variable aleatoria. La distribución de frecuencias dado por esta curva **"ideal"** es el **modelo teórico de distribución o ley de una variable aleatoria**.

Considerando que la curva fue construida a partir de un histograma de una distribución de frecuencias, se puede considerar como el límite del polígono de frecuencia del histograma dado y por tanto las frecuencias tienen las mismas propiedades del histograma. Basándonos en esto y recordando que el histograma fue hecho con la frecuencia relativa tenemos que:

- El área bajo la curva = al área del histograma = 1 (recuerde que el histograma representa todas las frecuencias relativas de una distribución de frecuencias y la suma de las frecuencias relativas siempre es 1).

- En el histograma, el área del rectángulo = la frecuencia relativa del intervalo, luego el área entre dos puntos bajo la curva = la frecuencia relativa del intervalo formado por los dos puntos.

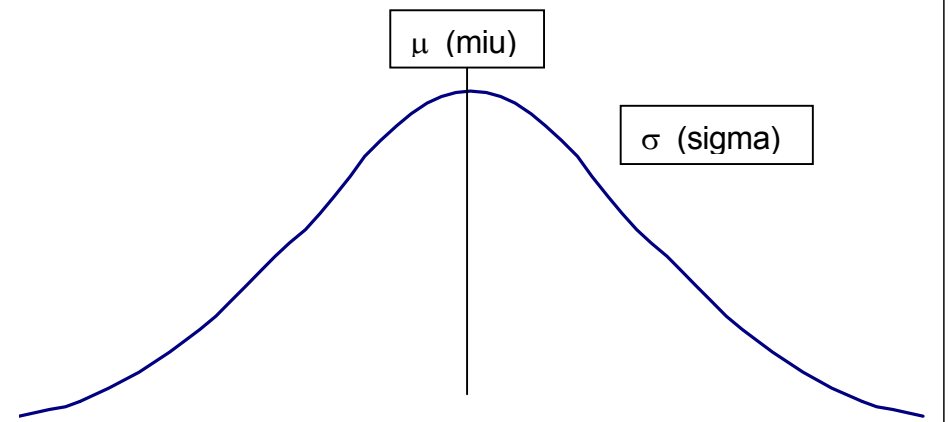
MODELO DE LA DISTRIBUCIÓN NORMAL. PARÁMETROS DE ESTA DISTRIBUCIÓN. PROPIEDADES.

Existen muchas variables aleatorias, incluyendo muchas de las variables con las que normalmente trabajamos en Medicina, que presentan un modelo teórico de distribución llamado Modelo Normal de Distribución, Distribución Normal, Curva Normal o Curva de Gauss, (diferentes nombres que existen para la Distribución

Normal). Esta distribución o curva es simétrica, acampanada, disminuyendo rápidamente su grosor hacia ambos lados, (colas). Vea el Gráfico 6.

La talla, componentes de plasma y sangre, el error de medición y otras variables utilizadas en Medicina siguen este modelo de distribución. El centro de simetría de la curva se denomina μ (miu), mientras σ (sigma) da la medida de la amplitud o ancho de la campana. Es sigma quien determina la forma o concavidad de la curva. Estas

GRAFICO 6
MODELO NORMAL DE DISTRIBUCION
DISTRIBUCION NORMAL



dos características (μ,σ), miu y sigma, se conocen como los **parámetros de la distribución normal.**

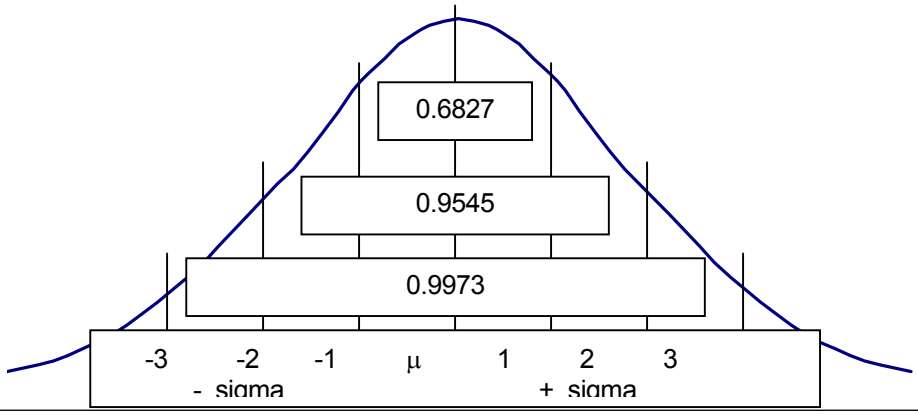
La distribución normal tiene las siguientes propiedades:

- La curva es simétrica respecto al eje que pasa por μ (miu).
- La forma de la curva depende de los valores de (μ,σ) miu y sigma.
- Existe un número infinito de distribuciones normales, una para cada valor de (μ,σ), miu y sigma.
- El área bajo la curva = 1.
- Para cualquier distribución normal siempre se cumple que el área de los intervalos:
  - $\mu \pm \mathbf{1}\sigma$ **= 0.6827 ó 68.27%**
  - $\mu \pm \mathbf{2}\sigma$ **= 0.9545 ó 95.45%** (Vea Gráfico 7)
  - $\mu \pm \mathbf{3}\sigma$ **= 0.9973 ó 99.73%**

Está última propiedad es muy importante y será utilizada frecuentemente en Estadística Inferencial. Se puede expresar de otra forma, más comprensible para el personal de la salud. En los intervalos formados por la media (μ) y la desviación estándar (σ) de una población hallaremos que:

- $\mu \pm 1\sigma$ **= 68.27% de los elementos**
- $\mu \pm 2\sigma$ **= 95.45% de los elementos**
- $\mu \pm 3\sigma$ **= 99.73% de los elementos**

GRAFICO 7
DISTRIBUCION NORMAL



En Medicina y en las investigaciones, por lo general se utilizan los siguientes intervalos:

- $\mu \pm 1.96\sigma$ **= 0.9500 = 95.0 % de los elementos**
- $\mu \pm 2.58\sigma$ **= 0.9900 = 99.0 % de los elementos**

En Medicina, la aplicación práctica es que cuando una variable aleatoria sigue la distribución normal, el rango de normalidad esta dado por $\mu \pm 1.96\sigma$. Ello significa que el 95 % de los elementos tendrán un valor dentro de ese intervalo. Aquellos elementos, cuyos valores se encuentren por fuera del intervalo se consideran patológicos.

LA DISTRIBUCIÓN NORMAL ESTÁNDAR. EJEMPLOS DE APLICACIÓN.

La distribución normal estándar es una distribución normal con fines eminentemente prácticos. En efecto, es una distribución con μ **= 0 y** σ **= 1, N(0,1),** habiéndose calculado el área en cualquier punto posible bajo la curva. Al punto bajo la curva se denomina **Z** y existen tablas para determinar el área bajo la curva para cualquier punto Z. Un ejemplo de tabla de Z se puede ver en la última página de este texto.

Existen diferentes tablas de Z. La que se presenta en este material se utiliza desde el punto central (MIU) hacia cualquiera de las dos mitades de la curva normal estándar. Se reconoce que su empleo es así por ser el valor del área = ,0000 para una Z = 0,00. Otra tabla de Z se emplea en sentido contrario, o sea de la periferia hasta el punto central y entonces para Z = 0,00 el valor del área fuera 0,4998 o 0,5000.

Para determinar un área bajo la curva, utilizando la tabla, se procede de la siguiente manera:

- los valores de Z se encuentran en la primera columna y en la primera fila;
- la columna es para enteros y un decimal;
- la fila es para la centésima;
- el valor del área bajo la curva desde el centro (MIU) hasta el valor de Z deseado está en la cuadrícula donde se cruza el valor de Z por la primera columna y el valor de la centésima por la primera fila.

Por ejemplo, para un valor de **Z = 2.36**; en la primera columna hallará el valor **2.3** y en la primera fila el valor **.06. En la intersección de la fila 2.3 y la columna .06 encontrará el área que corresponde a partir de MIU, esto es 0,4909.** Como la tabla da los valores de la mitad de la curva, los valores para un valor negativo de Z fueran los mismos, sólo que para el lado izquierdo de la curva.

La tabla de Z o de la curva normal estándar se utiliza para hacer cálculos de cualquier variable aleatoria que siga una distribución normal en la población. Lo que se hace es transformar un valor cualquiera **X** de la curva normal de la variable aleatoria al correspondiente valor de **Z** de la curva normal estándar. Para ello se utiliza la fórmula siguiente:

$$Z = \frac{X - \mu}{\sigma}$$

donde: X = valor de la variable aleatoria
μ = media poblacional de la variable aleatoria
σ = desviación estándar poblacional de la variable aleatoria
Z = valor del punto correspondiente en la curva normal estándar

Una vez realizada la transformación, el área bajo la curva se determina por la tabla de Z. Veamos algunos ejemplos.

1. Suponga que una variable aleatoria que sigue la distribución normal tiene un μ = 230 y σ = 20. Calcule el área bajo la curva entre los valores 220 y 280. Para simplificar el esquema tracemos una línea que nos represente la base de la curva normal.

μ

220---------230----------280

Para hallar el área entre 220 y 280 tenemos que hallar el área entre 220 y 230 y el área entre 220 y 280 para luego sumarlas. (**Recuerde que la tabla sólo permite valores del centro a la periferia).**

μ = 230

σ = 20

X_1 = 220

X_2 = 280

Area en la tabla de Z

$$Z_1 = \frac{220 - 230}{20} = -.50 \text{ ---------- } 0.1915$$

(El signo negativo en el valor de Z solamente indica que está a la izquierda de MIU)

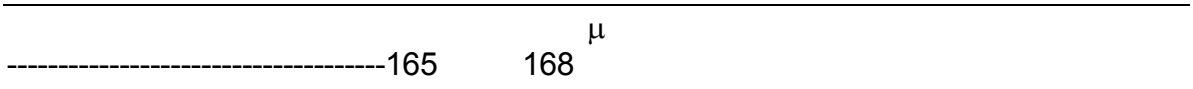
$$Z_2 = \frac{280 - 230}{20} = 2.50 \text{ ---------- } 0.4938$$

El área entre los dos valores fuera entonces: 0.1915 + 0.4938 = 0.6853

2. La talla en adultos masculinos sigue una distribución normal en determinada población con μ = 168 cms. y σ = 8 cms. Calcule:
   a) La probabilidad que una persona tenga una talla de 165 cms. ó menos.
   b) El porcentaje de personas con una talla de 170 cms o mayor.
   c) El porcentaje de personas cuya talla este entre 170 y 173 cms.
   d) Si la población total de adultos masculinos es de 20,000 personas, cuantos tuvieran una talla de 166 cms o menos.

Antes de entrar a solucionar los incisos y ver como se puede utilizar la tabla de la curva normal estándar, es bueno recordar que el área total bajo la curva es igual a 1 y por tanto la mitad de la curva, del centro (MIU) a la periferia, es igual a 0,5 hacia ambos lados.

a) Para calcular la probabilidad de una talla de 165 cms o menos tenemos que hallar el área bajo la curva. Recuerde que el área bajo la curva es también la probabilidad de ocurrencia de un resultado determinado. Debemos hallar el área representada en el siguiente esquema, o sea de 165 hacia la periferia:



Para hallar esa área es necesario encontrar el área entre 165 y 168 y después restar esa área de 0.5 que es la mitad del área de la curva normal.

X = 165
μ = 168
σ = 8

$Z = \frac{X - \mu}{\sigma}$ $\quad Z = \frac{165 - 168}{8} = \frac{-3}{8} = -0.375 \approx -0.38$

Z = -0.38; el área en la tabla entre 165 y 168 es = 0.1480

El área bajo la curva igual a 165 cms o menos sería: 0.5000 – 0.1480 = 0.352. Luego, la probabilidad que la talla de un adulto masculino fuera de 165 cms o menos es de 0.352.

b) Para calcular el porcentaje de personas con una talla de 170 cms o mayor tenemos que hallar el área bajo la curva. Debemos hallar el área representada en el siguiente esquema, o sea de 170 hacia la periferia:

μ

168 170----------

Para hallar esa área es necesario encontrar el área entre 168 y 170 y después restar esa área de 0.5 que es la mitad del área de la curva normal.

$X = 170$

$\mu = 168$

$\sigma = 8$

$Z = \frac{X - \mu}{\sigma}$ $Z = \frac{170 - 168}{8} = \frac{2}{8} = 0.25$

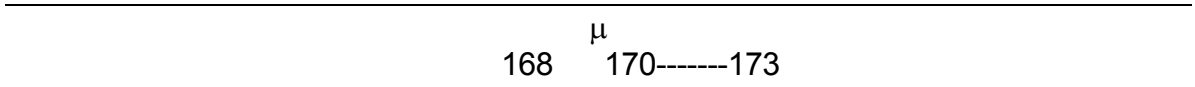
Z = -0.25; el área en la tabla entre 168 y 170 es = 0.0987

El área bajo la curva igual a 170 cms o más sería: 0.5000 – 0.0987 = 0.4013.

Luego, el porcentaje de adultos masculinos con talla de 170 cms o más es 0.4013 X 100 = 40.13%.

(Recuerde que la frecuencia relativa multiplicada por 100 es el porcentaje.)

c) Para calcular el porcentaje de personas cuya talla esté entre 170 y 173 cms tenemos que hallar el área bajo la curva. Debemos hallar el área representada en el siguiente esquema, o sea entre 170 y 173:



Para hallar esa área es necesario encontrar el área entre 168 y 173 y el área entre 168 y 170. El área entre 170 y 173 cms será la resta del área de 168 a 173 cms menos el área de 168 a 170 cms.

$X_1 = 173$
$X_2 = 170$
$\mu = 168$
$\sigma = 8$

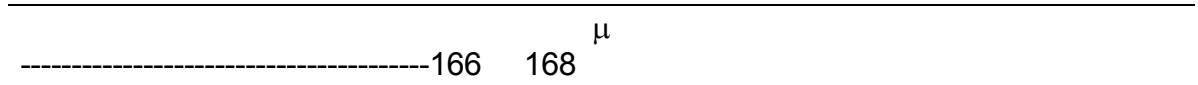
$$Z_1 = \frac{X_1 - \mu}{\sigma} \qquad Z_1 = \frac{173 - 168}{8} = \frac{5}{8} = 0.625 \approx 0.63$$

Para $Z_1 = 0.63$; el área es 0.2537

Para Z_2 tenemos el cálculo del ejercicio anterior; $Z_2 = 0.25$; el área es 0.0987

El área bajo la curva entre 170 cms y 173 cms sería: 0.2537 – 0.0987 = 0.137. Luego, el porcentaje de adultos masculinos con talla entre 170 y 173 cms es 0.137 X 100 = 13.7 %.

d) Para calcular cuantos adultos masculinos tuvieran una talla de 166 cms o menos, si la población total de adultos masculinos es de 20,000 personas, tenemos que hallar el área bajo la curva. Debemos hallar el área representada en el siguiente esquema, o sea de 166 cms hacia la periferia:



Para hallar esa área es necesario encontrar el área entre 168 y 166 y después restar de 0.5.

X = 166
μ = 168
σ = 8

$Z = \frac{X - \mu}{\sigma}$ $\quad Z = \frac{166 - 168}{8} = \frac{-2}{8} = -0.25$

Z = -0.25; el área es 0.0987

El área bajo la curva igual a 166 cms o menos sería: 0.5000 – 0.0987 = 0.4013.

Luego, el porcentaje de adultos masculinos con talla de 166 cms o menos es 0.4013 X 100 = 40.13 %.

Ahora se puede calcular cuantos adultos masculinos deben tener esa talla:

20,000 es al 100 % como
X es al 40.13 %

Por lo tanto, $X = \frac{20{,}000 \times 40.13}{100} = 8{,}026$

Los adultos masculinos con talla de 166 cms o menos deben ser 8,026.

CURVA NORMAL ESTANDAR

VALOR DE Z

Z	,00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0,0	,0000	,0040	,0080	,0120	,0160	,0199	,0239	,0279	,0319	,0359
0,1	,0398	,0438	,0478	,0517	,0557	,0596	,0636	,0675	,0714	,0753
0,2	,0793	,0832	,0871	,0910	,0948	,0987	,1026	,1064	,1103	,1141
0,3	,1179	,1217	,1255	,1293	,1331	,1368	,1406	,1443	,1480	,1517
0,4	,1554	,1591	,1628	,1664	,1700	,1736	,1772	,1808	,1844	,1879
0,5	,1915	,1950	,1985	,2019	,2054	,2088	,2123	,2157	,2190	,2224
0,6	,2257	,2291	,2324	,2357	,2389	,2422	,2454	,2486	,2517	,2549
0,7	,2580	,2611	,2642	,2673	,2703	,2734	,2764	,2794	,2823	,2852
0,8	,2881	,2910	,2939	,2967	,2995	,3023	,3051	,3078	,3106	,3133
0,9	,3159	,3186	,3212	,3238	,3264	,3289	,3315	,3340	,3365	,3389
1,0	,3413	,3438	,3461	,3485	,3508	,3531	,3554	,3577	,3599	,3621
1,1	,3643	,3665	,3686	,3708	,3729	,3749	,3770	,3790	,3810	,3830
1,2	,3849	,3869	,3888	,3907	,3925	,3944	,3962	,3980	,3997	,4015
1,3	,4032	,4049	,4066	,4082	,4099	,4115	,4131	,4147	,4162	,4177
1,4	,4192	,4207	,4222	,4236	,4251	,4265	,4279	,4292	,4306	,4319
1,5	,4332	,4345	,4357	,4370	,4382	,4394	,4406	,4418	,4429	,4441
1,6	,4452	,4463	,4474	,4484	,4495	,4505	,4515	,4525	,4535	,4545
1,7	,4554	,4564	,4573	,4582	,4591	,4599	,4608	,4616	,4625	,4633
1,8	,4641	,4649	,4656	,4664	,4671	,4678	,4686	,4693	,4699	,4706
1,9	,4713	,4719	,4726	,4732	,4738	,4744	,4750	,4756	,4761	,4767
2,0	,4772	,4778	,4783	,4788	,4793	,4798	,4803	,4808	,4812	,4817
2,1	,4821	,4826	,4830	,4834	,4838	,4842	,4846	,4850	,4854	,4857
2,2	,4861	,4864	,4868	,4871	,4875	,4878	,4881	,4884	,4887	,4890
2,3	,4893	,4896	,4898	,4901	,4904	,4906	,4909	,4911	,4913	,4916
2,4	,4918	,4920	,4922	,4925	,4927	,4929	,4931	,4932	,4934	,4936
2,5	,4938	,4940	,4941	,4943	,4945	,4946	,4948	,4949	,4951	,4952
2,6	,4953	,4955	,4956	,4957	,4959	,4960	,4961	,4962	,4963	,4964
2,7	,4965	,4966	,4967	,4968	,4969	,4970	,4971	,4972	,4973	,4974
2,8	,4974	,4975	,4976	,4977	,4977	,4978	,4979	,4979	,4980	,4981
2,9	,4981	,4982	,4982	,4983	,4984	,4984	,4985	,4985	,4986	,4986
3,0	,4987	,4987	,4987	,4988	,4988	,4989	,4989	,4989	,4990	,4990
3,1	,4990	,4991	,4991	,4991	,4992	,4992	,4992	,4992	,4993	,4993
3,2	,4993	,4993	,4994	,4994	,4994	,4994	,4994	,4995	,4995	,4995
3,3	,4995	,4995	,4995	,4996	,4996	,4996	,4996	,4996	,4996	,4997
3,4	,4997	,4997	,4997	,4997	,4997	,4997	,4997	,4997	,4997	,4998

INDICE

Capítulo 9

MUESTREO. DISTRIBUCION MUESTRAL DE LA MEDIA ARITMETICA CUANDO LA DISTRIBUCION DE LA VARIABLE ORIGINAL ES NORMAL. ERROR ESTANDAR.

Autor: Dr. Emilio Morales Jiménez

SUMARIO

- *Muestreo. Ventajas y desventajas. Error de muestreo.*
- *Muestreo probabilístico y no probabilístico.*
- *Muestra representativa. Condiciones de una buena muestra.*
- *Esquemas de muestreo: muestreo aleatorio simple y muestreo estratificado.*
- *Distribucion muestral de la media aritmetica cuando la distribucion de la variable original es normal. Error estandar. .*

Muestreo. Ventajas y desventajas. Error de muestreo.

En Medicina, como en muchos otros campos del saber, es imposible trabajar con todo el universo o población. Por ejemplo:

- para conocer la talla y el peso de una población de 10 millones de habitantes es imposible pesar y medir a toda la población;
- para conocer la efectividad de un nuevo medicamento es imposible tratar a toda la población de pacientes;
- para conocer los síntomas de una enfermedad es imposible estudiar a toda la población que la padece.

Lo que usualmente se hace es estudiar una parte de la población o sea una muestra de la población.

Recordemos dos conceptos básicos: población y muestra.

- Población o universo es el número total ó conjunto de elementos (personas, animales ó cosas) que tienen una o más características, bien definidas y medibles, en común.
- Muestra es cualquier subconjunto de la población ó universo.

La acción de seleccionar una muestra de una población se denomina **muestreo**. Existen libros dedicados a explicar como se realizan y utilizan diferentes métodos de muestreo para aquel interesado en el mismo.

Las ventajas del muestreo son:

- Ahorro de tiempo, esfuerzos y recursos.
- Se pueden obtener resultados más exactos y confiables estudiando una muestra que estudiando a toda la población. Se debe a que el personal altamente calificado y bien entrenado para la recolección y el procesamiento de los datos está generalmente disponible para estudiar una muestra, siendo prácticamente imposible disponer de ese personal en cantidad suficiente para estudiar a toda la población.

- Algunos estudios implican la destrucción de los elementos. Por ejemplo, el efecto de un nuevo medicamento en animales de experimentación donde el animal debe ser sacrificado.

Como desventaja se ha señalado una característica especial del muestreo, **el error de muestreo**. El error de muestreo se debe a la variabilidad de los datos y a los elementos que se incluyen en una muestra determinada y no es un error ni una equivocación. El error de muestreo no se puede evitar aunque puede ser reducido incrementando el tamaño de la muestra. También puede ser estimado si se trabaja con métodos de muestreo probabilístico. **Hay que diferenciar el error de muestreo del error sistemático que se puede producir durante la observación científica al momento de recolectar la información y que tiene que ser evitado.**

Veamos el concepto de error de muestreo con un ejemplo sencillo. Supóngase que tenemos una población de cuatro personas cuyas tallas en cms. son las siguientes: 168, 170, 176 y 182 cms. **La media poblacional (μ) es 174 cms.** De esta población obtengamos dos muestras de tamaño 2 y calculemos la media muestral:

- muestra 1 168, 170 media = 169 cms.
- muestra 2 170, 182 media = 176 cms.

Tomemos también dos muestras de tamaño 3 y realicemos el mismo cálculo:

- muestra 1 168, 170, 182 media = 173.3 cms
- muestra 2 168, 176, 182 media = 175.3 cms

Cuando la media poblacional (μ) se compara con las medias muestrales ($\bar{x}$) se ve claramente que existe una diferencia entre ellas y que la misma decrece a medida que el tamaño de la muestra aumenta. **A la diferencia entre el parámetro poblacional y el estadígrafo de la muestra se le llama error de muestreo.**

Muestreo probabilístico y no probabilístico.

Existen dos tipos de muestreo: probabilístico y no probabilístico.

El muestreo probabilístico es aquel donde se aplica la teoría de probabilidades y en consecuencia los elementos de la muestra son seleccionados en forma aleatoria. Selección en forma aleatoria significa que es la suerte, la casualidad o la oportunidad lo que determina que se seleccione un elemento. Permite estimar el error de muestreo.

El muestreo no probabilístico no aplica la teoría de las probabilidades y los elementos de la muestra no se eligen aleatoriamente. De hecho, el muestreo no probabilístico es muy frecuente en medicina. Se utiliza cuando un grupo específico de pacientes se selecciona para un estudio pero tiene el inconveniente que los

resultados obtenidos no se pueden inferir a la población de pacientes de donde se ha tomado la muestra.

Muestra representativa. Condiciones de una buena muestra.

El principal objetivo del muestreo es obtener una **buena muestra** de la población. Una muestra es buena cuando refleja las características de la población permitiendo inferencias válidas hacia ella. Existen dos factores que influencian en que una muestra sea buena:

- **el tamaño de la muestra y**
- **las condiciones de selección de la muestra.**

El tamaño de la muestra va a depender de la variabilidad o dispersión de la variable que se estudie. A medida que la variabilidad se incremente será mayor el tamaño de la muestra. Cuando veamos la estimación y la forma de calcular el tamaño muestral para una estimación se verá esto en detalle.

Las condiciones de selección de la muestra se refiere a las diferentes características que existen en la población y están involucradas directamente con la variable que se está estudiando. Por ejemplo, si la variable está influenciada por el sexo, la edad y el nivel cultural, entonces en la muestra, estas características deben estar presentes en la misma composición o proporción que ellas tienen en la población.

Cuando en una muestra se cumplen esos dos factores estamos en presencia de una **muestra representativa** de la población ya que reproduce a pequeña escala las características de la población.

Esquemas de muestreo: muestreo aleatorio simple y muestreo estratificado.

Existen diferentes muestreos probabilísticos. Ellos brindan la misma posibilidad a todos los elementos de la población de ser seleccionados para la muestra. Indiscutiblemente, esto solamente puede suceder si la manera de seleccionar los elementos se basa en la suerte ó la casualidad lo que significa que se realiza de una manera aleatoria. Estudiaremos solamente el muestreo simple aleatorio y el muestreo estratificado.

Muestreo simple aleatorio

Este método es el más comúnmente utilizado, bien solo ó en combinación con otros métodos. Se usa cuando la población es **homogénea**, esto es, cuando la variabilidad de la variable no es influenciada por las características de la población.

Para aplicar el método es necesario conocer y disponer del marco muestral, o sea la posibilidad de enumerar a toda la población. La población se enumera del 0 a n, y después se seleccionan los elementos de la muestra por una técnica aleatoria como una tabla de números aleatorios ó una rifa.

Por ejemplo, se requiere una muestra de 30 elementos (n = 30) para hacer un trabajo de investigación con las historias clínicas de pacientes de apendicitis aguda que fueron tratados en un determinado hospital durante dos años.

- Primero, se deben identificar todas las historias clínicas de los pacientes tratados por apendicitis aguda en el periodo de tiempo determinado; supongamos que fueron 500.
- Segundo, se enumeran todos del 0 al 499 (500 números).
- Tercero, realizamos una rifa. Se colocan en una caja 500 pedazos de papel iguales numerados del 0 al 499 y se mezclan bien.
- Cuarto, sin mirar se extrae un papel y se selecciona la historia clínica correspondiente. En el ejemplo se repite la operación 30 veces. para seleccionar 30 historias clínicas.

Muestreo estratificado

Este método se utiliza cuando la variable es influenciada o varía según una o más características de la población. Estas características dividen a la población en diferentes estratos. Cada estrato es diferente del otro siendo **heterogéneos** entre si pero la variable en estudio es **homogénea dentro de cada estrato**. En este caso la muestra debe reflejar estas condiciones.

Por ejemplo, se requiere una muestra de 40 elementos (n = 40) para conocer la talla de los estudiantes de una escuela primaria que tiene 600 estudiantes. La variable edad es influenciada por la edad del escolar, pero asumamos que en cada grado escolar la edad es homogénea. Los estratos estarán entonces formados por el grado escolar de la manera siguiente:

(los grados escolares son heterogéneos entre sí)

150 – escolares en primer grado
100 – escolares en segundo grado
110 – escolares en tercer grado
90 – escolares en cuarto grado
70 – escolares en quinto grado
80 – escolares en sexto grado
600 – Total de escolares en la escuela

Para seleccionar la muestra primero hay que conocer la proporción de cada estrato en la población, lo que estadísticamente se conoce como el peso relativo del estrato. Para calcularlo se divide el tamaño del estrato entre la población:

Estrato	Cálculo	Proporción o peso del estrato
Primer grado	150/600	0.25
Segundo grado	100/600	0.17
Tercer grado	110/600	0.18
Cuarto grado	90/600	0.15
Quinto grado	70/600	0.12
Sexto grado	80/600	0.13
	Suma de las proporciones	**1.00**

(Fíjese que la suma de las proporciones de los estratos es 1.00, ya que es una frecuencia relativa)

La cantidad de elementos de cada estrato que se incluirá en la muestra de manera de reflejar ó reproducir las características de la población se calcula de la siguiente forma:

Estrato	Cálculo	Proporción o peso del estrato
Primer grado	0.25 x 40	10----------10
Segundo grado	0.17 x 40	6.8 ≈ 7-------7
Tercer grado	0.18 x 40	7.2 ≈ 7-------7
Cuarto grado	0.15 x 40	6----------6
Quinto grado	0.12 x 40	4.8 ≈ 5-------5
Sexto grado	0.13 x 40	5.2 ≈ 5-------5
	Tamaño muestral	**40**

Luego en el primer grado se seleccionan 10 escolares, 7 en el segundo grado y así sucesivamente como se muestra en la tabla precedente obteniéndose una muestra de 40 escolares. Para escoger los escolares dentro de cada estrato se aplica un muestreo simple aleatorio como se describió anteriormente.

Distribucion muestral de la media aritmetica cuando la distribucion de la variable original es normal. Error estándar

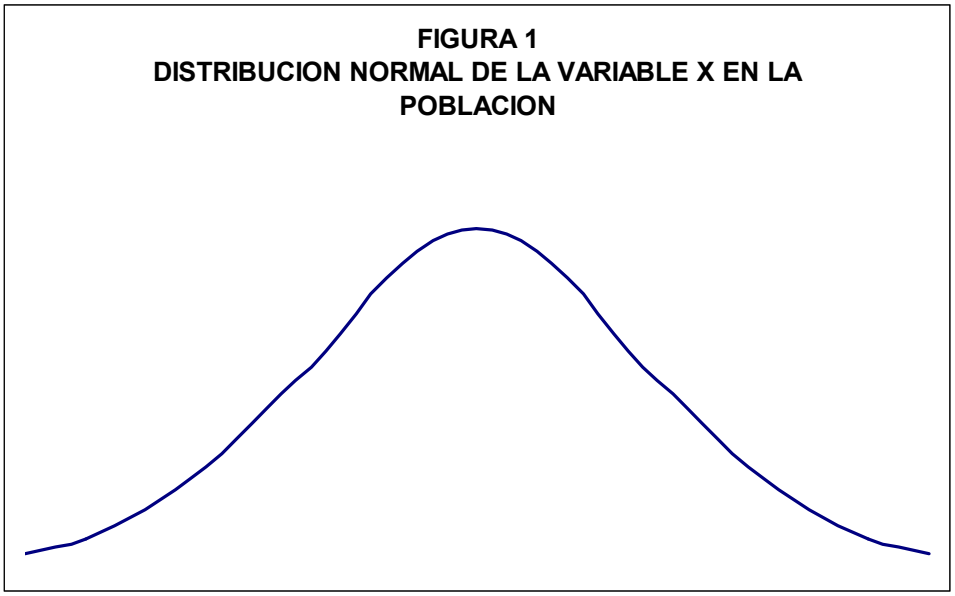
Los estadígrafos calculados en una muestra (por ejemplo la media aritmética), cuando se calculan de una variable aleatoria también son valores aleatorios. Si de una población tomamos diferentes muestras y de cada una de ellas calculamos una media aritmética vamos a tener un conjunto de medias aritméticas y todas ellas son valores aleatorios. Ese conjunto de medias constituye de por sí una variable aleatoria y en este caso se le conoce como la media muestral.

La media muestral tiene también un modelo teórico de distribución cuando la variable de origen tiene una distribución normal. Ese modelo teórico de la población de medias muestrales tiene dos parámetros poblacionales:

- La media poblacional de las medias muestrales también llamado **media de las medias** denotado por $\mu_{\bar{x}}$ y
- La desviación estándar de las medias muestrales también llamado **desviación estándar de la media** denotado por $\sigma_{\bar{x}}$**.**

Veamos estos conceptos con un ejemplo teórico. Supongamos que tenemos una población en la cual se está midiendo una variable **X** que sigue una distribución normal. Esa variable tendrá una media poblacional (μ**x**) y una desviación estándar poblacional (σ**X**).

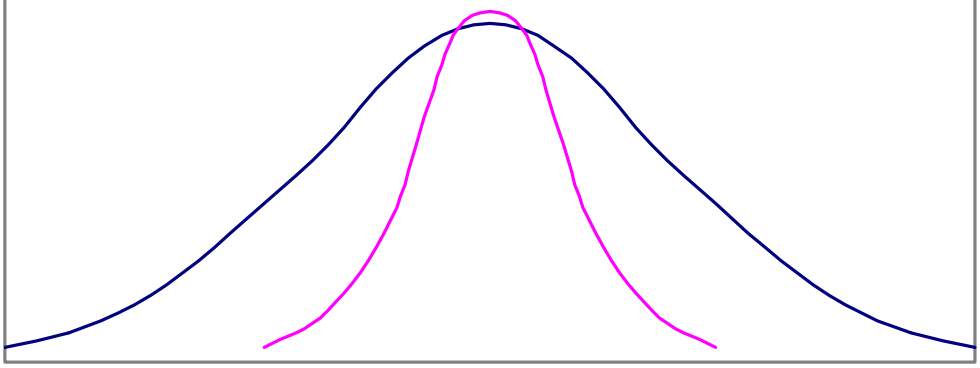
FIGURA 1
DISTRIBUCION NORMAL DE LA VARIABLE X EN LA POBLACION



Con una media (μ**X**) y una desviación estándar (σ**X**).

Si de esa población tomamos diferentes muestras de igual tamaño **n** y de cada una obtenemos una media aritmética ($\bar{x}$) tendremos un conjunto de medias. Esas medias por definición deben estar muy cerca de la media poblacional y como provienen de una variable con una distribución normal también van a tener una distribución normal. Como es lógico la curva de distribución normal de las medias muestrales va a resultar mucho más estrecha que la curva de la variable original ya que todos sus elementos son medias muestrales de la variable original.

GRAFICO 2
DISTRIBUCION NORMAL DE LA MEDIA MUESTRAL SUPERPUESTA A LA DISTRIBUCION NORMAL DE LA VARIABLE X



Es facil comprender que:

- la media de la población de medias muestrales será la misma media poblacional, o sea $\mu_{\bar{x}}$ **= μx.**
- Sin embargo, la desviación estándar de la media ($\sigma_{\bar{x}}$) tiene que ser mucho más pequeña que la desviación estándar de la variable (σ**X**) ya que los valores de las medias muestrales están muy cerca de μ**x** con menor dispersión.
- Ha sido probado matemáticamente que

$\sigma_{\bar{x}}$ **=** σ**x/**$\sqrt{}$**n**, donde **n** es el tamaño muestral.

La desviación estándar de la media **(**$\sigma\bar{x}$) es la desviación teórica de las medias muestrales ($\bar{x}$) con relación a la media poblacional (μ**x**) y también se le conoce como el **error estándar de la media**, aunque no sea un error. Se denota por **EE**$\bar{x}$. Podemos resumir que $\sigma\bar{x}$ **=** σ**x/**$\sqrt{}$**n = EE**$\bar{x}$**.**

Cuando este parámetro no se conoce puede ser estimado con una muestra y entonces se le conoce como el **error estándar muestral** ó **error estándar de la muestra** y se denota por **ee**$\bar{x}$. Se calcula con la desviación estándar de la muestra, luego **ee**$\bar{x}$ **= s/**$\sqrt{}$**n.**

La importancia del error estándar de la media es que mide la variabilidad de las medias muestrales.

INDICE

Capítulo 10
INFERENCIA ESTADISTICA. ESTIMACION

Autores: Dr. Emilio Morales Jiménez

SUMARIO

- *Estimación de parámetros poblacionales. Estimación puntual y por intervalos de confianza.*
- *Estimación por intervalo de confianza de la media poblacional (μ) con σ conocida y desconocida.*
- *Estimación por intervalo de confianza de una proporción poblacional (P).*
- *Precisión y confiabilidad de una estimación por intervalo.*
- *El tamaño de la muestra en función de la precisión (máximo error admisible) y la confiabilidad de la estimación.*

INTRODUCCION

La inferencia estadística es la parte de la Estadística que nos permite tomar decisiones acerca de los valores que se calculan con los datos de una muestra en relación con la población o universo de donde se extrajo la muestra. En Medicina así como en otras ciencias, las técnicas de la inferencia estadística son muy útiles. Por lo general el médico está en contacto con un grupo limitado de pacientes, no con la población o universo de pacientes, y las conclusiones que el pueda obtener estudiando a estos pacientes no pueden ser generalizadas a toda la población de pacientes al menos que se apliquen las técnicas de la inferencia estadística.

En la práctica diaria del médico está tratando pacientes continuamente y dándose cuenta que no todos los pacientes son iguales. Un tratamiento que es bueno para uno, puede no ser el adecuado para otro aunque padezcan la misma enfermedad. Muchas veces estará motivado en averiguar porque sucede esta situación. Para ello es necesario realizar una investigación científica para obtener algunas conclusiones sobre este problema. Si no se aplican las técnicas de la estadística inferencial, el trabajo se limita a una descripción de lo que se ha encontrado en los pacientes y no se pueden plantear conclusiones válidas para el universo de pacientes.

Durante el curso estudiaremos dos aspectos diferentes de la estadística inferencial: la estimación y la prueba de hipótesis.

- La estimación permite dar valores de un evento o fenómeno (variables) en la población mediante el estudio de una muestra representativa de ésta. Dicho de otra manera, hacer válido para la población los valores que se han determinado en la muestra representativa de ella.
- La prueba de hipótesis permite conocer si dos poblaciones son semejantes ó difieren con relación a un evento ó fenómeno (variables) mediante el estudio de una muestra representativa de cada una; ó conocer si un valor hipotético ó

teórico es válido para ser aplicado a una población determinada. Permite por ende la toma de decisiones.

Estimación de parámetros poblacionales. Estimación puntual y por intervalos de confianza.

Como se planteó en el párrafo anterior, la estimación, como dice su nombre, nos permite **estimar el valor de un parámetro poblacional** mediante el estudio de una muestra representativa de la población. Ahora bien, es necesario apuntar que siempre la **estimación permite conocer en forma aproximada los valores de los parámetros poblacionales**, a partir de una muestra representativa de la población.

Observe que, estimación es sinónimo de aproximación.

Existen dos **tipos de estimación**:

- **Estimación puntual** (Mediante estadígrafos)
- **Estimación por intervalos** (Por medio de intervalos de confianza).

Estimación puntual

La estimación puntual es muy sencilla. Se asume que el valor del estadígrafo calculado con los datos de la muestra se utiliza como una aproximación al valor del parámetro poblacional. Por ejemplo, si en una muestra representativa de una población se calcula la media aritmética ($\bar{x}$), entonces el valor de la media poblacional (μ) será similar. Por tanto, un **estimador puntual** es todo estadígrafo que sirve como aproximación de un parámetro poblacional.

La principal ventaja de un estimador puntual es su sencillez, sin embargo, su utilización directa conduce solamente a resultados más bien groseros. En efecto, saltan a la vista las siguientes **insuficiencias** de todo estimador puntual:

1. No permite conocer cuán buena es la aproximación.

2. No toma en cuenta ninguna medida del error muestral, el cual es, recuerde, la diferencia entre un parámetro y el estadígrafo que lo aproxima, es ineludible, y se debe a la variabilidad de los datos.

Estimación por intervalos

Para entender como se calcula la estimación por intervalo, nos basaremos en las propiedades de la distribución normal aplicadas al modelo teórico de la media muestral. En el modelo teórico de la media muestral teníamos que:

$$\mu_{\bar{x}} = \mu x$$
$$\text{y} \quad \sigma_{\bar{x}} = \sigma x/\sqrt{n} = EE_{\bar{x}}.$$

Luego podemos decir que:

$\mu_{\bar{x}} \pm \sigma_{\bar{x}}$ ó $\mu x \pm \sigma x/\sqrt{n}$ incluye al 68.27% de las medias muestrales $\bar{x}$.
$\mu_{\bar{x}} \pm 2\sigma_{\bar{x}}$ ó $\mu x \pm 2\sigma x/\sqrt{n}$ incluye al 95.45% de las medias muestrales $\bar{x}$.
$\mu_{\bar{x}} \pm 3\sigma_{\bar{x}}$ ó $\mu x \pm 3\sigma x/\sqrt{n}$ incluye al 99.73% de las medias muestrales $\bar{x}$.

Esto lo podemos expresar de otra forma, por ejemplo:

$$\mu x - 2\frac{\sigma x}{\sqrt{n}} \leq \bar{x} \leq \mu x + 2\frac{\sigma x}{\sqrt{n}}$$

con una probabilidad de 0.9545 o con una certeza del 95.45%

Si despejamos μx en las dos partes de la ecuación tendremos que:

$$\mu x - 2\frac{\sigma x}{\sqrt{n}} \leq \bar{x} \longrightarrow \mu x \leq \bar{x} + 2\frac{\sigma x}{\sqrt{n}}$$

y

$$\bar{x} \leq \mu x + 2\frac{\sigma x}{\sqrt{n}} \longrightarrow \bar{x} - 2\frac{\sigma x}{\sqrt{n}} \leq \mu x$$

entonces se puede escribir que:

$$\bar{x} - 2\frac{\sigma x}{\sqrt{n}} \leq \mu x \leq \bar{x} + 2\frac{\sigma x}{\sqrt{n}}$$

con una probabilidad de 0.9545 o con una certeza del 95.45%

Si expresamos la ecuación en palabras tendremos que: la media poblacional está incluida en el intervalo formado por la media muestral $\bar{x} \pm 2$ errores estándar de la media con una probabilidad de 0.9545 o con una certeza (confiabilidad) de 95.45%.

Esto implica que la media poblacional tiene la posibilidad de estar incluida en el intervalo en el 95.45 % de los casos. Estadísticamente se expresa que: **μx esta incluida en el intervalo con una confiabilidad de 95.45%**. Este intervalo se conoce como **intervalo de confianza.**

Para generalizar la fórmula, el número de desviaciones estándar se sustituye por la letra **K** siendo entonces:

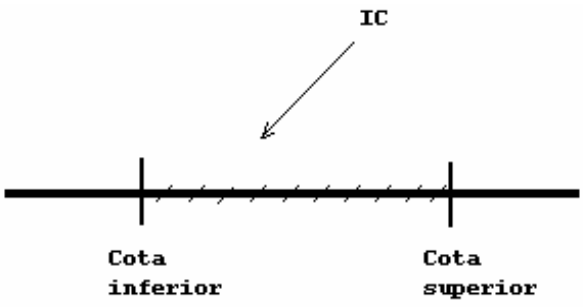
$$\bar{x} - K\frac{\sigma x}{\sqrt{n}} \leq \mu x \leq \bar{x} + K\frac{\sigma x}{\sqrt{n}} \quad o \quad \mu x \in \bar{x} \pm K\frac{\sigma x}{\sqrt{n}}$$

Como ya se ha dicho, en Medicina, por lo general, se trabaja con una confiabilidad de 95% o 99% donde la probabilidad es 0.95 y 0.99 respectivamente. **En estos casos para una confiabilidad de 95%, K = 1.96 y para una confiabilidad de 99%, K = 2.58**. **K** se considera como el coeficiente de confiabilidad.

Resumiendo:
En Medicina resulta necesaria una vía más refinada de obtención de las aproximaciones de los parámetros; ella es precisamente la **estimación por intervalos de confianza**, la cual posee las siguientes **características**:

1. Un **intervalo de confianza (IC)** está definido por dos cotas, una inferior y otra superior, que se obtienen a partir de los datos muestrales.
2. El IC se calcula especificando previamente el **nivel de confianza** (son sinónimos **confiabilidad, certeza** y **certidumbre**), el cual es la probabilidad de que el intervalo contenga al parámetro estimado. Los niveles de confianza más utilizados son 95% y 99%. El nivel de confianza se expresa por los coeficientes de confiabilidad **(K).** Son 1.96 para 95% y 2.58 para 99%, expresando las cantidades de desviación estándar poblacional (error estándar) debajo de la curva normal que se utilizan para formar el IC.
3. En el cálculo de un IC interviene el llamado **error estándar**, que constituye una medida del error muestral.

El cálculo se realiza hallando el estadígrafo en la muestra ($\bar{x}$) ó (p) y luego sumando y restando el producto del coeficiente de confiabilidad por el error estándar del estadígrafo.



Tenga presente que, un IC no es simplemente un intervalo numérico, pues él tiene intrínsecamente asociada una medida de la verosimilitud de incluir al parámetro estimado.

Estudiaremos la estimación por intervalos de la media poblacional y de la proporción poblacional.

<u>Estimación por intervalo de confianza de la media poblacional (μ) con σ conocida y desconocida.</u>

El parámetro a estimar es μ = media poblacional desconocida de cierta variable de interés que es aleatoria, cuantitativa, continua y con distribución normal ó gaussiana.

Recordemos que el error estándar de la media,

$$\mathbf{EE}_{\bar{x}} = \frac{\sigma x}{\sqrt{n}}$$

donde:

- σx = Desviación estándar poblacional
- n = tamaño de la muestra

Se utiliza cuando se conoce el valor de la desviación estándar poblacional, cuestión teórica y muy poco probable.

En la práctica, es muy poco probable conocer el valor de **σ** y se utiliza el estimador del $EE_{\bar{x}}$ que es el error estándar muestral, $\mathbf{ee}_{\bar{x}} = \mathbf{s}/\sqrt{\mathbf{n}}$, donde s = desviación estándar de la muestra.

Intervalo de confianza para σ conocida

$$\mu x \in \bar{x} \pm K * EE_{\bar{x}} \text{ ó } \mu x \in \bar{x} \pm K \frac{\sigma x}{\sqrt{n}}$$

Cota inferior = $\bar{x} - K * EE_{\bar{x}} = \bar{x} - K \frac{\sigma x}{\sqrt{n}}$

Cota superior = $\bar{x} + K * EE_{\bar{x}} = \bar{x} + K \frac{\sigma x}{\sqrt{n}}$

Donde:

- $\bar{x}$ = Media aritmética de la muestra
- K = Coeficiente de confiabilidad, 1.96 si confiabilidad de 95% y 2.58 si de 99%
- σx = Desviación estándar poblacional
- n = Tamaño de la muestra

Intervalo de confianza para σ desconocida

En este caso, que es el que sucede en la práctica diaria, el coeficiente de confiabilidad se halla, por lo general, con la distribución **t de Student**, muy similar a la distribución normal pero más exacta cuando la muestra es menor ó igual a 30 elementos. Como se desconoce **σX** no se puede utilizar el **EE**$_{\bar{x}}$ y entonces se utiliza su estimador, **ee**$_{\bar{x}}$ que se calcula con la desviación estándar de la muestra.

Intervalo de confianza

$$\mu x \in \bar{x} \pm t * ee_{\bar{x}} \text{ ó } \mu x \in \bar{x} \pm t \frac{s}{\sqrt{n}}$$

Cota inferior = $\bar{x} - t * ee_{\bar{x}} = \bar{x} - t \frac{s}{\sqrt{n}}$

Cota superior = $\bar{x} + t * ee_{\bar{x}} = \bar{x} + t \frac{s}{\sqrt{n}}$

siendo:

- n = Tamaño muestral
- $\bar{x}$ = Media muestral
- s = Desviación estándar muestral
- t = Coeficiente de confiabilidad hallado por el valor denominado “t de Student”, el cual se encuentra tabulado, o sea calculado en una tabla que se muestra al final.

Para hallar la “t” en la tabla correspondiente, debe tenerse en cuenta que ella depende de:

1. El nivel de confianza (confiabilidad)
2. **(n – 1)**, esto es, el tamaño de la muestra que se escoja rebajado en una unidad. **(n – 1)** recibe el nombre de **grados de libertad (g.l.)**.

Cuando **(n – 1) < 30** es imprescindible buscar la **“t”** en su tabla, no obstante cuando **(n – 1)** ≥ **30** sucede que **t** ≈ **K** donde K = 1.96 para 95% de confiabilidad y K = 2.58 para 99% de confiabilidad, con lo cual la tabla se hace innecesaria (observe que, en esta situación los grados de libertad (g.l.) no influyen en el valor de t).

Ejemplo. Un conjunto de 12 animales experimentales recibe su alimentación con una dieta especial durante tres semanas. Al cabo de este tiempo, se registran los siguientes aumentos de peso (gramos): 30, 22, 32, 26, 24, 40, 34, 36, 32, 33, 28, 30. ¿Con una confianza del 95% cuales serían los límites para la ganancia de peso media en la población correspondiente?

Solución.
La fórmula a utilizar es:

$$\mu x \in \bar{x} \pm t * ee_{\bar{x}} \text{ ó } \mu x \in \bar{x} \pm t \frac{s.}{\sqrt{n}}$$

para:
- n = 12
- Calculada la media aritmética del aumento de peso de los 12 animales; $\bar{x}$ = 30.58
- Calculada la desviación estándar; s = 5.09
- t(95% de confianza, 11 g.l.) = 2.20

Sustituyendo estos valores en la expresión del IC, resulta:

$$30.58 \pm 2.20.\frac{5.09}{\sqrt{12}} = 30.58 \pm 2.20.(1.47) = 30.58 \pm 3.23$$

de donde:

Límite inferior del IC = 30.58 – 3.23 = 27.3 gm

Límite superior del IC = 30.58 + 3.23 = 33.8 gm

(No parece tener mucho sentido dar estas respuestas con aproximación hasta la centésima de gramo.)

Luego, la ganancia media de peso en la población en cuestión se ubica entre 27.3 y 33.8 gramos, con una certeza (confiabilidad) de 95%.

Estimación por intervalo de confianza de una proporción poblacional (P).
El parámetro a estimar es la **proporción poblacional (P)** desconocida de cierta característica de interés. **P** puede expresarse como un número entre 0 y 1, o bien como un porcentaje: valor comprendido entre 0% y 100%; es esta segunda opción la que se utiliza en el presente curso.

Recordemos como se calcula el porcentaje: **p = a/n x 100**

Esta estimación solamente se puede realizar cuando la muestra es grande. Para la proporción una muestra se considera grande cuando el tamaño muestral n es

igual o mayor de 100 elementos **(n ≥ 100)**. En estos casos la proporción tiene una distribución semejante a una distribución normal.

De la misma forma que en las variables numéricas o cuantitativas se considera la desviación estándar como una medida de la dispersión de los datos y por ende de como varían los mismos, en las variables categóricas donde calculamos proporciones esa variación se determina por **(p * q),** donde:

- p = Proporción muestral de presencia de la cualidad estudiada = a/n x 100
- q = Proporción muestral de ausencia de la cualidad estudiada = 100 - p

El **error estándar del porcentaje muestral EE(p)** se calcula de la siguiente forma:

$$EE(p) = \sqrt{\frac{p.q}{n}}$$

donde:

- p = Proporción muestral de presencia de la cualidad estudiada = a/n x 100
- q = Proporción muestral de ausencia de la cualidad estudiada = 100 - p
- n = Tamaño muestral

Intervalo de confianza

$$\mathbf{P \in p \pm K * EE(p) \quad ó \quad P \in p \pm K * \sqrt{\frac{p.q}{n}}}$$

Cota inferior = $p - K.EE(p) = p - K.\sqrt{\frac{p.q}{n}}$

Cota superior = $p + K.EE(p) = p + K.\sqrt{\frac{p.q}{n}}$

siendo:

- n = Tamaño muestral
- p = Proporción muestral de presencia de la cualidad estudiada = a/n x 100
- q = Proporción muestral de ausencia de la cualidad estudiada = 100 – p
- K = 1.96 si nivel de confianza = 95%
- K = 2.58 si nivel de confianza = 99%

Ejemplo. En una muestra de 800 personas hay 40 asmáticos. Se desea una estimación por intervalo para el porciento de asmáticos en la población correspondiente, con una certidumbre de 99%.

Solución.
La fórmula a emplear es:

$$\mathbf{P \in p \pm K*EE(p) \quad ó \quad P \in p \pm K*\sqrt{\frac{p.q}{n}}}$$

Como se solicita una certidumbre ó confiabilidad de 99% será:

$$p \pm 2.58.\sqrt{\frac{p.q}{n}}$$

para:

n = 800
a = 40

el porcentaje de la muestra será: $p = \frac{40}{800} x 100 = 5\%$ y $q = 100 - 5 = 95\%$

Sustituyendo estos valores en la expresión escrita para el IC, resulta:

$$5 \pm 2.58.\sqrt{\frac{5.(95)}{800}} = 5 \pm 2.58.(0.770) = 5 \pm 1.99 \cong 5 \pm 2$$

de donde:

Cota inferior del IC = 5 – 2 = 3%;

Cota superior del IC = 5 + 2 = 7%

De esta manera, el porcentaje de asmáticos en la población se encuentra entre 3%
y 7%, con una certeza de 99%.

<u>Precisión y confiabilidad de una estimación por intervalo.</u>

La precisión y la confiabilidad que se desea en una estimación por intervalo de confianza son fundamentales para la determinación del tamaño de la muestra que se debe tomar de la población. Esto es una cuestión que normalmente preocupa a todo investigador.

Si el tamaño de la muestra no es el adecuado pueden suceder dos situaciones. Si es muy grande ocurre un malgasto de recursos y si es muy pequeña el resultado puede no ser válido.

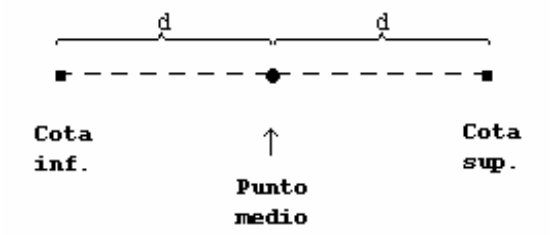
El tamaño de la muestra depende de tres factores:

- **Variabilidad de la variable de estudio en el universo o población.** Mientras mayor sea la variación de los datos, mayor ha de ser la muestra. La variación se determina por la varianza poblacional y en su defecto por la varianza de una muestra.
- **Precisión que se quiere con los resultados**, o sea el máximo error permisible en los resultados. A mayor precisión hará falta una muestra mayor.
- **Confiabilidad o margen de certeza que se desea obtener (95% ó 99%)**. A mayor certeza para determinada precisión la muestra será más grande.

La variación de la variable no depende del investigador sino de las características intrínsecas de la propia variable aleatoria, sin embargo tanto la precisión como la confiabilidad dependen del investigador. De ahí su importancia para lograr una adecuada estimación.

El tamaño de la muestra en función de la precisión (máximo error admisible) y la confiabilidad de la estimación.

El IC estudiado tiene la estructura siguiente:



"**d**" es el **máximo error admisible** (%).

Para calcular este IC se disponía de una muestra de tamaño conocido. Este valor, conjuntamente con la confiabilidad escogida de antemano y el estadígrafo calculado para el atributo de interés definían las cotas del intervalo y, por ende, el máximo error de estimación que con él se cometía. Ahora, se planteará la interrogante inversa, según se enuncia a continuación.

Problema. Hallar el tamaño de muestra necesario para calcular un IC con una confiabilidad dada y un error de estimación máximo **(d)** seleccionados previamente.

Tamaño de la muestra para la estimación de una proporción (porcentaje)

Se distinguirán dos casos.

Primer caso: Se conoce un valor de la proporción muestral **p** a través de un estudio preliminar.
La fórmula adecuada es:

$$n = \frac{pqK^2}{d^2}$$

donde, una vez más:

p = Proporción muestral disponible
q = 100 - p
K = 1.96 ó 2.58 según que la confiabilidad = 95% ó 99%, respectivamente
d = máximo error admisible en la estimación (%).

Segundo caso: No se dispone de un valor preliminar de la proporción muestral **p**.
La fórmula apropiada es ahora:

$$n = \frac{2500K^2}{d^2}$$

El valor de 2500 sustituyendo a **pq** surge por ser este el máximo valor posible de este producto. En efecto, si p = 50%, q = 50% y 50 x 50 = 2500. Cualquier otra combinación de **p** y **q** darán productos menores. En este segundo caso se toma la mayor variación posible.

Este segundo caso es más realista que el primero pero, tiene la desventaja de proporcionar tamaños de muestra mayores y, por consiguiente, más difíciles de lograr en la práctica.

En ambos casos, el tamaño muestral resultante se incrementa con la elevación de la confiabilidad de la estimación y, disminuye con el aumento de su máximo error admisible.

Ejemplo. Se desea conocer el tamaño de muestra necesario para estimar con una confianza del 95% y un error de estimación no mayor que 3%, el porciento de sujetos con disnea en una población de enfermos de cáncer del pulmón.

Solución.
Se debe aplicar la expresión:

$$n = \frac{2500K^2}{d^2}$$

con K = 1.96 y d = 3%.
Entonces:

$$n = \frac{2500(1.96)^2}{3^2} = 1067.1 \approx 1068$$

Así, el tamaño de muestra necesario es de 1068 enfermos.
Debido a que el máximo error de estimación es más bien pequeño, este tamaño muestral resulta algo elevado.

Tamaño de la muestra para la estimación de una media
Se distinguirán dos casos.
Primer caso: Se conoce el valor de la varianza poblacional **(σ^2)**.
La fórmula adecuada es:

$$n = \frac{K^2 \sigma^2}{d^2}$$

donde:
σ^2 = Varianza poblacional
K = 1.96 ó 2.58 según que la confiabilidad = 95% ó 99%, respectivamente
d = máximo error admisible en la estimación

Segundo caso: No se conoce el valor de la varianza poblacional, luego es necesario utilizar un estimador de ella que sería la varianza muestral **(S^2).** Para ello se toma una muestra de la población y se determina la misma.
La fórmula apropiada es ahora:

$$n = \frac{K^2 S^2}{d^2}$$

Como en la proporción, este segundo caso es más realista que el primero pero, tiene la desventaja de haberse realizado con una muestra cuyo tamaño puede no haber sido el adecuado.

En ambos casos, el tamaño muestral resultante se incrementa con la elevación de la confiabilidad de la estimación y, disminuye con el aumento de su máximo error admisible.

Ejemplo. Se desea conocer el tamaño de muestra necesario para estimar con una confianza del 95% y un error de estimación no mayor de 0.5 el nivel de la hemoglobina en una población de la misma edad y sexo que por un estudio piloto se determinó una desviación estándar muestral de la hemoglobina de 2.0

Solución.
Se debe aplicar la expresión:

$$n = \frac{K^2 S^2}{d^2}$$

con:
K = 1.96
S = 2.0
d = 0.5

Entonces:

$$n = \frac{(1.96)^2\,(2)^2}{(0.5)^2} = \frac{3.84 \times 4}{0.25} = \frac{15.36}{0.25} = 61.4 \approx 62$$

Así, el tamaño de muestra necesario es de 62 personas.

Tabla de la t de Student

g.l.	Conf. 95% α = 0.05	Conf. 99% α = 0.01	g.l.	Conf. 95% α = 0.05	Conf. 99% α = 0.01
1	12.71	63.66	20	2.09	2.84
2	4.30	9.92	21	2.08	2.83
3	3.18	5.84	22	2.07	2.82
4	2.78	4.60	23	2.07	2.80
5	2.57	4.03	24	2.06	2.79
6	2.45	3.71	25	2.06	2.78
7	2.36	3.50	26	2.06	2.78
8	2.31	3.35	27	2.05	2.77
9	2.26	3.25	28	2.05	2.76
10	2.23	3.17	29	2.05	2.76
11	2.20	3.11	30	2.04	2.75
12	2.18	3.05	40	2.02	2.70
13	2.16	3.01	50	2.00	2.68
14	2.14	2.98	60	2.00	2.66
15	2.13	2.95	70	1.99	2.65
16	2.12	2.92	80	1.99	2.64
17	2.11	2.90	90	1.99	2.63
18	2.10	2.88	100	1.98	2.63
19	2.09	2.86	∞	1.96	2.58

INDICE

Capítulo 11
INFERENCIA ESTADISTICA. PRUEBA DE HIPOTESIS

Autores: Dr. Emilio Morales Jiménez

SUMARIO

- *Conceptos generales: Hipótesis nula e hipótesis alternativa. Prueba de hipótesis de una cola y de dos colas. Errores de Tipo I (α) y de Tipo II (β). Nivel de significación. Región crítica de una prueba de hipótesis.*
- *Prueba de hipótesis acerca de la media de una distribución normal cuando σ es conocida y desconocida.*
- *Prueba de hipótesis acerca de una Proporción Poblacional.*
- *Prueba de hipótesis sobre Diferencia de Medias.*
- *Prueba de hipótesis sobre diferencias de Proporciones Poblacionales.*

Conceptos generales: Hipótesis nula e hipótesis alternativa. Prueba de hipótesis de una cola y de dos colas. Errores de Tipo I (α) y de Tipo II (β). Nivel de significación. Región crítica de una prueba de hipótesis.

Introducción e importancia

Otro aspecto de la inferencia estadística muy útil para el personal de la salud, en especial médicos, estomatólogos, enfermeros y tecnólogos de la salud son las pruebas de hipótesis. Los métodos de pruebas de hipótesis permiten al profesional de la salud tomar decisiones en relación con diferentes aspectos muy necesarios en Medicina y en Salud Pública. Estos métodos permiten modelar estadísticamente y resolver importantes problemas de naturaleza muy diversa, que de otra forma no tendrían una solución rigurosamente científica. Para ello los diferentes métodos suministran procedimientos para la toma de decisiones concernientes a:

⇒ Comparaciones de un parámetro **(por ejemplo, media o porcentaje)** de una población con un valor teórico dado o con el parámetro homólogo de otra población.
⇒ Asociaciones entre variables cualitativas.
⇒ Homogeneidad de poblaciones.
⇒ Correlaciones entre variables cuantitativas.

Ejemplos de la utilidad en el campo de la salud pueden ser las siguientes situaciones problemáticas entre otras muchas y diversas:

- Decidir si un método terapéutico nuevo es superior o no al método utilizado habitualmente.
- Corroborar si una prueba diagnóstica es más eficiente que otra.
- Establecer si una enfermedad tiene comportamiento similar en dos locaciones diferentes.
- Saber si los pacientes tratados por uno mismo tienen una evolución similar a la descrita por otros especialistas.

En todas estas cuestiones y otras de gran interés en Medicina, la metodología de las pruebas de hipótesis aporta conclusiones de valor práctico. Durante la práctica diaria, en trabajos científicos, temas de investigación y muchos otros aspectos de las actividades competentes del profesional de la salud se requiere comparar valores que se han calculado de una o más variables. Con frecuencia el profesional limita su análisis al valor hallado en una muestra (media, porcentaje u otro), e incluso si hay varios valores los analiza sin realizar pruebas de hipótesis. Este error muy común tiene como consecuencia arribar a conclusiones erróneas o no científicas fácilmente cuestionables. Para asegurar que dos porcentajes o dos medias difieren entre sí, especialmente cuando sus valores no difieren mucho, es necesario demostrarlo por métodos estadísticos. La diferencia de los valores por si sola, sin la prueba estadística no permite arribar a una conclusión y tomar una decisión a partir de ella. De ahí la gran importancia de las pruebas de hipótesis en el campo de la salud. De hecho, en los estudios médicos, las pruebas de hipótesis coadyuvan a respaldar o impugnar la convicción de que, las conclusiones novedosas rebasan el marco de las meras anécdotas clínicas o epidemiológicas debidas a la casualidad.

Las pruebas de hipótesis que abordaremos en este capítulo se les llama **pruebas paramétricas ya que las comparaciones siempre se referirán a parámetros poblacionales. Sin embargo, la decisión de aceptar o rechazar una hipótesis se realizará con los valores observados y calculados de una muestra aleatoria.** Esto significa que siempre existirá un cierto margen de error. Con esto en mente, una posible definición de las pruebas de hipótesis es: **las técnicas y procedimientos estadísticos que tienen como principal objetivo el establecer conclusiones válidas para una población mediante el estudio de una muestra representativa de la misma, midiendo el margen de error en términos probabilísticos.**

Hipótesis nula e hipótesis alternativa

Una **hipótesis estadística** es una afirmación concerniente a una o más poblaciones.

Para **someter a prueba (contrastar, docimar) una afirmación tal** se utilizan las **pruebas de hipótesis estadísticas**, las cuales se denominan también **pruebas de significación** y **dócimas de hipótesis**.

Las pruebas de hipótesis se estructuran sobre la base de una **hipótesis nula y una hipótesis alterna o alternativa:**

- La **hipótesis nula** es un postulado de igualdad, de no-diferencia o no-asociación y se expresa por $\mathbf{H_0}$.
- La **hipótesis alternativa** expresa desigualdad, diferencia, asociación y se expresa por $\mathbf{H_A}$.

Estas hipótesis conforman una pareja y, se requiere de una clase específica de prueba para cada tipo de dupla. La decisión estadística se efectúa siempre a favor de una de las hipótesis del par.

Prueba de hipótesis de una cola y de dos colas

En el momento que en una hipótesis nula se comparan dos magnitudes numéricas, (una ignorada y la otra conocida, o bien, ambas desconocidas), ella asume la forma de una igualdad; entonces, la hipótesis alternativa puede ser de los siguientes tipos:

- "distinto de" (**alternativa bilateral**)
- "mayor que" o "menor que" (**alternativas unilaterales**)

La primera de estas tres modalidades requiere de una prueba de hipótesis **llamada "de dos colas"**. En ella lo que interesa es demostrar estadísticamente que las dos magnitudes son distintas ya sea por defecto o por exceso.

Las otras dos modalidades necesitan una prueba de hipótesis llamada **"de una cola"**. En ellas interesa demostrar que una magnitud es mayor que la otra o lo contrario, que una magnitud es menor que la otra.

Ejemplo. Si **P** es una proporción poblacional (desconocida) y $\mathbf{P_0}$ es un valor dado, pueden escribirse las tres parejas siguientes de hipótesis:

- $\mathbf{H_0: P = P_0 \; ; \; H_A: P \neq P_0}$
- $\mathbf{H_0: P \leq P_0 \; ; \; H_A: P > P_0}$
- $\mathbf{H_0: P \geq P_0 \; ; \; H_A: P < P_0}$

La primera de estas parejas tiene una alternativa bilateral y requiere pues de una prueba de dos colas. Las parejas segunda y tercera contienen alternativas unilaterales, luego, cada una se resuelve a través de una prueba de una cola.

Para destacar que la prueba de hipótesis favorecerá siempre la adopción de una de las hipótesis en dejación de la otra, se usa con frecuencia una escritura del tipo

$$\mathbf{H_0: P = P_0 \; VS \; H_A: P \neq P_0}$$

Lo mismo se aplica a cada una de las otras dos parejas.

Solamente estudiaremos las alternativas bilaterales y las pruebas desarrolladas serán por consiguiente de dos colas.

Errores de Tipo I (α) y de Tipo II (β)

En las pruebas de hipótesis se rechaza o se acepta la hipótesis nula. Es lógico pensar que si se rechaza la hipótesis nula, se acepta la alternativa. Los criterios de decisión son de naturaleza probabilística. Así, puede ser rechazada una

hipótesis nula verdadera (**error de tipo I (α)**) o, aceptada una que sea falsa (**error de tipo II (β)**). La tabla siguiente informa con más detalle acerca de esta cuestión.

Tabla de Decisión

Decisión	H_0 es cierta	H_0 es falsa
Aceptar H_0	Decisión correcta	Error de tipo II
Rechazar H_0	Error de tipo I	Decisión correcta

Estos tipos de error están presentes siempre en una prueba de hipótesis y cuando se intenta disminuir uno de ellos el otro se incrementa. Solamente disminuyen ambos cuando se incrementa el tamaño muestral. Usualmente, se adopta la conducta de limitar la probabilidad de cometer un error de tipo I.

Nivel de significación

El **nivel de significación estadística** es la máxima probabilidad admisible de cometer un error de tipo I ó **(α)**, la cual es fijada por el propio investigador. Este número debe ser "pequeño" y en este texto se utilizará siempre un nivel de significación de 5% ó α = 0.05 y un nivel de significación de 1% ó α = 0.01; (en la literatura se utilizan también 10% y valores incluso menores que 1%).

El nivel de significación está relacionado con la certeza o confiabilidad, así como al coeficiente de certeza o confiabilidad visto anteriormente en la estimación. La relación se muestra en la siguiente tabla:

NIVEL DE SIGNIFICACION (α)	CERTEZA O CONFIABILIDAD (100 - α)	COEFICIENTE DE CONFIABILIDAD (k) – estimación (Z) – prueba de hipótesis
5 % 1 %	95 % 99 %	1.96 2.58
(Expresado en probabilidades)		
0.05 0.01	0.95 0.99	1.96 2.58

Región crítica de una prueba de hipótesis

Una vez adoptado el valor que ha de tener el nivel de significación, con él se construyen las regiones subsiguientes:

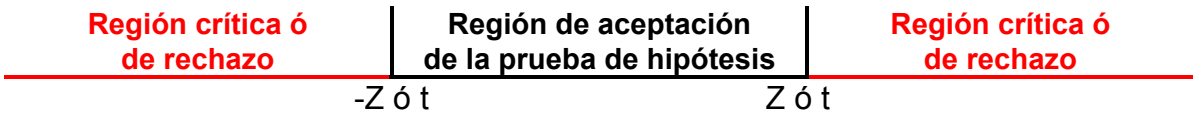
1) La **región crítica** o **región de rechazo** de H_0.
2) La **región de aceptación** de H_0.

Estas regiones son complementarias una de la otra.

Regiones

Las regiones dependen del nivel de significación fijado por el investigador antes de realizar la prueba de hipótesis. Depende, por tanto, del valor de α y del valor de Z ó de t correspondiente, según se utilice la distribución normal o la distribución de Student respectivamente. Se obtienen estableciendo un intervalo

de confianza (también llamado intervalo de aceptación) para el parámetro que está siendo probado y las áreas que quedan fuera del intervalo constituyen la región crítica como se muestra en la figura siguiente.



El nivel de significación se denomina también **tamaño de la región crítica**.

Si en la prueba de hipótesis se utiliza el valor de Z la región de aceptación se construye con los siguientes valores:
Z = 1.96 si nivel de significación = 5% (α = 0.05);
Z = 2.58 si nivel de significación = 1% (α = 0.01)

Si en la prueba de hipótesis se utiliza el valor de t hay que buscar en la tabla de la "t de Student" el valor correspondiente para el nivel de significación y los grados de libertad de la prueba como se verá mas adelante.
Nota: Si los grados de libertad son < 30 se toma el valor de la tabla de la t de Student. Si los grados de libertad son ≥ 30 no se recurre a la tabla y se asume que el valor de t ≈ Z y por ende se toma el valor de esta última para el nivel de significación correspondiente.

Toda prueba de hipótesis descansa en última instancia en el empleo de un estadígrafo propio de ella: el **estadígrafo de la prueba**, el cual es una magnitud adimensional que mide en una escala probabilística el grado de desviación de la hipótesis nula que un juego de datos específico exhibe. Se calcula por tanto un valor de Z ó de t y se analiza si cae en la región de aceptación o en la región crítica o de rechazo.

Observación. En cualquier prueba de hipótesis aparecen dos valores de Z o de t: uno de ellos es el tabulado y el otro es el calculado. La decisión estadística resulta, en última instancia, de la comparación de estos valores.

- Si el valor encontrado para este estadígrafo se encuentra en la región crítica, $\mathbf{H_0}$ se rechaza y se acepta $\mathbf{H_A}$; se dice entonces que, la prueba produjo un **resultado estadísticamente significativo**.

- Si, por el contrario, el valor mencionado se halla en la región de aceptación de $\mathbf{H_0}$, esta hipótesis no se rechaza, o sea se acepta, en detrimento de $\mathbf{H_A}$; se declara en este caso que, la prueba tuvo un **desenlace estadísticamente no significativo**.

Estas reglas integran el **criterio de decisión** de la prueba de hipótesis.

El nivel de significación es la frontera entre lo significativo y aquello que no lo es, vale decir, entre las desigualdades, diferencias o asociaciones esenciales y las

casuales. Al mismo tiempo, él expresa la máxima probabilidad admisible de rechazar una hipótesis nula que sea cierta, esto es, de cometer un error de tipo I. En definitiva, un resultado significativo indica que, las desigualdades, diferencias o asociaciones halladas en las muestras reunidas no son producto del azar y subsisten, pues, para las poblaciones estudiadas, siempre con una probabilidad de error, medible pero no eliminable: esta es precisamente la **característica básica de la inferencia** que aporta una prueba.

Para aplicar el aparato matemático correspondiente a una prueba de hipótesis deben tomarse aleatoriamente tantas muestras como poblaciones se citen en las hipótesis: una muestra para cada población.

Un aspecto que hay que tener en cuenta en el análisis estadístico en el campo de la salud es que significación estadística **no es sinónima de importancia o** relevancia biológica o clínica**.** El no demostrar significación estadística no elimina la existencia de una diferencia clínica o biológica importante entre dos poblaciones.

Es importante recalcar que c**ualquier diferencia, por pequeña que sea, puede ser estadísticamente significativa,** si el tamaño de la muestra es lo suficientemente grande**.** Si **n** aumenta, el poder aumenta y cualquier efecto, por pequeño que sea, puede ser detectado. Sin embargo una diferencia de tal pequeña magnitud, a pesar de ser estadísticamente significativa, **puede no ser clínicamente importante**.

El valor de **p** se asocia siempre al resultado de una prueba estadística y en muchas ocasiones solamente se utiliza este valor como indicador de la significación estadística. Un valor pequeño de **p** brinda evidencia de que existe alguna diferencia entre dos poblaciones, pero no necesariamente implica que la diferencia sea grande. Es importante destacar que el valor de **p** esta influenciado por el tamaño muestral **n** y la desviación estándar **σ**.

- A medida que **n se incrementa p disminuye.**
- A medida que **σ aumenta también aumenta p**.

Por lo general, cuando la muestra es pequeña o cuando la desviación estándar de los datos es grande es muy probable que las pruebas estadísticas no permitan demostrar significación estadística en las diferencias observadas. En estos casos, si se considera que **la diferencia es clínicamente importante,** es apropiado plantear:

"Los resultados no son estadísticamente significativos; los datos no mostraron suficiente evidencia para demostrar la diferencia existente" o

"El tamaño muestral resultó insuficiente para demostrar significación estadística en las diferencias observadas"

Pasos a seguir para desarrollar una prueba de hipótesis

1) Identificar los parámetros, valores teóricos de interés y los otros valores necesarios para el cálculo.

2) Enunciar las hipótesis apropiadas.

3) Escoger el nivel de significación

4) Calcular el estadígrafo de la prueba

5) Construir las regiones crítica y de aceptación de H_0 y comparar el valor del estadígrafo calculado con el valor tabulado

6) Aplicar el criterio de decisión

7) Redactar la conclusión

Prueba de hipótesis acerca de la media de una distribución normal cuando σ es conocida y desconocida.

En Medicina en muchas ocasiones trabajamos con variables numéricas o cuantitativas y lo que nos interesa valorar y comparar es la media aritmética como medida de tendencia central de los datos de esa variable. Variables tales como el nivel de hemoglobina, glicemia, temperatura, tensión arterial, frecuencia cardiaca, días de estadía, número de sesiones de una aplicación fisioterapéutica entre muchos otros son ejemplos de ello. Para tomar una decisión adecuada sobre los resultados es necesario muchas veces aplicar una prueba de hipótesis.

Veremos ahora el caso particular de realizar la prueba de hipótesis de una media poblacional. Un requisito para la prueba es que la variable siga una distribución normal en la población. En este caso se compara el valor de una media aritmética obtenida de una muestra representativa de la población con un valor teórico o supuesto para la media de la característica en estudio en esa población.

Parámetro y valor teórico

- μ = Media poblacional desconocida de cierta variable de interés que es aleatoria, cuantitativa, continua y con distribución normal o gaussiana.
- μ_0 = Valor (hipotético) especificado.
  El valor μ_0 tiene su origen en la experiencia, en un estudio preliminar o piloto, una reflexión teórica o una simple elucubración entre otras formas de obtenerse.

Hipótesis

Hipótesis nula $H_0: \mu = \mu_0$

Hipótesis alternativa $H_A: \mu \neq \mu_0$

De este modo, se trata de decidir si la media poblacional toma cierto valor dado o difiere significativamente de él.

Estadígrafo de la prueba

Cuando σ es conocido, caso teórico, pues muy pocas veces se da en la práctica, el estadígrafo de la prueba es Z y se calcula por la siguiente formula:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}.$$

(En este caso las regiones se hacen con el valor de Z; 1.96 si $\alpha = 0.05$ y, 2.58 si $\alpha = 0.01$)

Cuando σ no se conoce, que es la generalidad de los casos, el estadígrafo de la prueba es la t y se utiliza la siguiente fórmula:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

siendo:

n = Tamaño muestral
$\bar{x}$ = Media aritmética de los valores de la muestra
s = Desviación estándar muestral.
μ_0 = Valor (hipotético) especificado.

Ejemplo. Se afirma que la estatura media de los estudiantes de cierta edad en un municipio es 163.5 cm. De la población formada por dichos estudiantes se extrae una muestra representativa de 23 estudiantes, la cual da una estatura promedio de 165 cm con una desviación estándar de 3.8 cm. Determinar si la afirmación manifestada debe aceptarse o rechazarse con un nivel de significación del 5%.

Solución.

1. **Identificar los parámetros, valores teóricos de interés y los otros valores necesarios para el cálculo.**
   - Parámetro: μ = Media poblacional (desconocida) de la estatura de los estudiantes.
   - El valor teórico (μ_0) es 163.5 cm, puesto que así lo expresa el enunciado del problema.
   - n = 23 estudiantes
   - $\bar{x}$ = 165 cm ; s = 3.8 cm

2. **Enunciar las hipótesis apropiadas.**
   H_0: $\mu = \mu_0$ ó H_0: μ = 163.5 cm
   H_A: $\mu \neq \mu_0$ ó H_A: $\mu \neq$ 163.5 cm
   Luego, hay que establecer si μ = 163.5 o existe una diferencia estadísticamente significativa entre el parámetro μ y el valor 163.5.

3. **Escoger el nivel de significación**
   Nivel de significación = 5%

4. **Calcular el estadígrafo de la prueba**

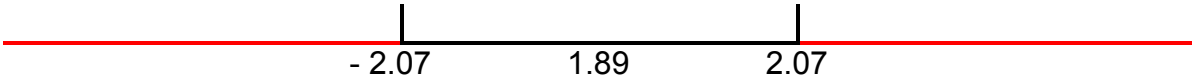
$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

para:
n = 23, $\bar{x}$ = 165, s = 3.8
o sea:

$$t = \frac{165 - 163.5}{3.8 / \sqrt{23}} = 1.89$$

5. **Construir las regiones crítica y de aceptación de H_0 y comparar el valor del estadígrafo calculado con el valor tabulado**
   Los grados de libertad de la t en esta prueba son: (n – 1).

   Para construir las regiones necesarias se busca la t de Student tabulada correspondiente a (n – 1) = 22 g.l. y nivel de significación 5%.
   Resulta que, t = 2.07. En consecuencia, las regiones son las que siguen.



   El valor del estadígrafo se encuentra en la región de aceptación de la hipótesis nula.

6. **Aplicar el criterio de decisión**
   El valor del estadígrafo se encuentra en la región de aceptación. Se acepta la hipótesis nula con el nivel de significación 5%, esto es, la prueba de hipótesis arroja un resultado estadísticamente no significativo con este nivel.

7. **Redactar la conclusión**
   La estatura media de la población de estudiantes estudiada no difiere significativamente de 163.5 cm, por lo cual se acepta el criterio existente, con el nivel de significación 5%.

Prueba de hipótesis acerca de una Proporción Poblacional.

En Medicina la mayoría de los trabajos presentados en eventos científicos o publicados en revistas médicas, así como las tesis de grado, se realizan con variables tales como el diagnóstico, complicaciones más frecuentes, evolución de pacientes, síntomas y otras variables categóricas o cualitativas. Los datos se resumen y presentan en porcentajes (proporción). Para tomar una decisión adecuada sobre los resultados es necesario muchas veces aplicar una prueba de hipótesis.

Veremos ahora el caso particular de realizar la prueba de hipótesis de una proporción poblacional. En este caso se compara el valor de un porcentaje obtenido de una muestra representativa de la población con un valor teórico o supuesto para el porcentaje de la característica en estudio en esa población.

Parámetro y valores teóricos

- **P =** Proporción poblacional desconocida de cierta característica de interés.
  P, puede expresarse como un número entre 0 y 1, o bien como un porcentaje: valor comprendido entre 0% y 100%, y es esta segunda opción la que se utiliza en el presente curso.

- $\mathbf{P_0}$ = Porcentaje (hipotético) especificado, $0\% < P_0 < 100\%$
- $\mathbf{Q_0}$ = 100 - $\mathbf{P_0}$

El valor $\mathbf{P_0}$ tiene su origen en la experiencia, en un estudio preliminar o piloto, una reflexión teórica o una simple elucubración entre otras formas de obtenerse.

Hipótesis

Hipótesis nula $\mathbf{H_0: P = P_0}$

Hipótesis alternativa $\mathbf{H_A: P \neq P_0}$

Así, se trata de decidir si la proporción poblacional asume cierto valor dado o difiere significativamente de él.

Estadígrafo de la prueba

$$Z = \frac{p - P_0}{\sqrt{[P_0 Q_0 / n]}}$$

siendo:

p = Porcentaje muestral de la cualidad o categoría estudiada
n = Tamaño muestral
P_0 = Porcentaje (hipotético) especificado
$Q_0 = 100 - P_0$

Recuerde que el porcentaje se calcula de la siguiente forma: **p = f/n x 100**; donde f es la frecuencia de aparición en la muestra de una característica determinada.

Ejemplo. En una provincia de cierto país se desarrolla un estudio acerca de la prevalencia de amebiasis en niños. Existe el criterio de que la enfermedad está presente en el 12% de estos. Una muestra aleatoria de 150 niños informa de 27 niños con amebas. Averigüe si el criterio existente es válido con un nivel de significación del 5%.

Solución.

1. **Identificar los parámetros, valores teóricos de interés y los otros valores necesarios para el cálculo.**

- Parámetro: P = Porcentaje poblacional (desconocido) de niños con amebiasis.
- El valor teórico ($\mathbf{P_0}$) es 12%, puesto que así lo expresa el enunciado del problema.
- n = 150 niños
- f = 27 niños con amebas

2. **Enunciar las hipótesis apropiadas.**

$\mathbf{H_0: P = P_0}$ ó $\mathbf{H_0: P = 12\%}$

$\mathbf{H_A: P \neq P_0}$ ó $\mathbf{H_A: P \neq 12\%}$

3. **Escoger el nivel de significación**

Nivel de significación = 5%

4. **Calcular el estadígrafo de la prueba**

$$Z = \frac{p - P_0}{\sqrt{[P_0 Q_0 / n]}}$$

para:

$\mathbf{n}$ = 150, $\mathbf{f}$ = 27, $\mathbf{p} = \frac{27}{150} \times 100 = 18.0$

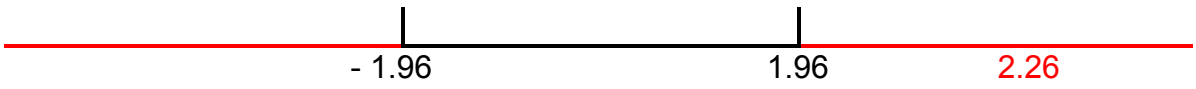
$\mathbf{P_0}$ = 12, $\mathbf{Q_0}$ = 100 - P_0 = 88

o sea:

$$\mathbf{Z = \frac{18 - 12}{\sqrt{[12 \times 88/150]}} = 2.26}$$

5. **Construir las regiones crítica y de aceptación de H_0 y comparar el valor del estadígrafo calculado con el valor tabulado**

Como nivel de significación es 5% el valor de Z que se toma es 1.96



El valor del estadígrafo se encuentra en la región de rechazo.

6. **Aplicar el criterio de decisión**

El valor del estadígrafo se encuentra en la región crítica. Se rechaza la hipótesis nula con el nivel de significación 5%, esto es, la prueba de hipótesis produce un resultado estadísticamente significativo con este nivel.

7. Redactar la conclusión

La prevalencia de amebiasis en la población infantil estudiada difiere significativamente del 12%, por lo cual se refuta el criterio existente acerca de dicha prevalencia, con el nivel de significación 5%.

Prueba de hipótesis sobre Diferencia de Medias.

Con bastante frecuencia el profesional de la salud tiene que discriminar entre dos medias de una variable numérica aleatoria, o sea, comparar las medias de dos poblaciones. Por ejemplo, tenemos dos poblaciones de pacientes que padecen de una misma enfermedad. La población A recibe tratamiento con Cefalexina y la población B con Amikacina. Se quiere saber si el tiempo promedio de curación es igual o diferente en las dos poblaciones. Para averiguar esto es necesario obtener una muestra aleatoria de cada población y calcular el tiempo promedio de curación en cada muestra y con esas medias hacer el procedimiento de la prueba de hipótesis de diferencia de medias.

Es importante recalcar que esta prueba se realiza entre dos medias solamente. Si hay que comparar mas de dos medias hay que utilizar otro procedimiento llamado análisis de varianza (ANOVA).

Parámetros

- μ_1 = Media poblacional desconocida, en población 1, de cierta variable de interés que es aleatoria, cuantitativa, continua y con distribución normal o gaussiana.
- μ_2 = Media poblacional desconocida, en población 2, de cierta variable de interés que es aleatoria, cuantitativa, continua y con distribución normal o gaussiana.

Hipótesis

Hipótesis nula $H_0: \mu_1 = \mu_2$

Hipótesis alternativa $H_A: \mu_1 \neq \mu_2$

De este modo, se trata de decidir si las medias poblacionales son iguales o difieren significativamente.

Estadígrafo de la prueba

Cuando σ es conocido, caso teórico, pues muy pocas veces se da en la práctica, el estadígrafo de la prueba es Z y se calcula por la siguiente formula:

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{[(\sigma_1)^2/n_1 + (\sigma_2)^2/n_2]}}.$$

(En este caso las regiones se hacen con el valor de Z; 1.96 si $\alpha = 0.05$ y, 2.58 si $\alpha = 0.01$)

Cuando σ no se conoce, que es la generalidad de los casos, el estadígrafo de la prueba es la t y se utiliza la siguiente fórmula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{[(Sp)^2/n_1 + (Sp)^2/n_2]}}.$$

donde:

n_1 = Tamaño de la muestra de la población 1

n_2 = Tamaño de la muestra de la población 2

$\bar{x}_1$ = Media de la variable estudiada en la muestra de la población 1.

$\bar{x}_2$ = Media de la variable estudiada en la muestra de la población 2.

s_p^2 = Pool de varianza o varianza muestral **ponderada o combinada** de las dos muestras (es una varianza calculada sobre la base de la unión de ambas muestras)

El Pool de varianza se calcula de la siguiente forma:

$$Sp^2 = \frac{S_1^2 (n_1 - 1) + S_2^2 (n_2 - 1)}{n_1 + n_2 - 2}$$

donde: s_1^2 y s_2^2 = Varianzas muestrales.

Ejemplo. En un estudio de la relación entre la diverticulosis y el hábito dietético, intervinieron 27 vegetarianos. El contenido total de fibra dietética fue una de las variables de interés. Se obtuvo la información siguiente para un grupo de sujetos sanos y otro de enfermos:

CONTENIDO DE FIBRA DIETETICA EN LA DIETA

Medidas descriptivas	Sin diverticulosis	Con diverticulosis
Tamaño muestral	18	9
Media	42.7	27.7
Desviación estándar	9.9	9.5

Diga si es admisible que el promedio del contenido total de fibra dietética en la dieta de los sujetos sanos es mayor que en la de los enfermos con un nivel de significación 5%.

Solución.

1. **Identificar los parámetros, valores teóricos de interés y los otros valores necesarios para el cálculo.**
   - Parámetro: μ_1 = Media poblacional (desconocida) del contenido de fibra dietética en la dieta de los pacientes sin diverticulosis.
   - Parámetro: μ_2 = Media poblacional (desconocida) del contenido de fibra dietética en la dieta de los pacientes con diverticulosis.
   - n_1 = 18 pacientes
   - $\bar{x}_1$ = 42.7
   - s_1 = 9.9
   - n_2 = 9 pacientes
   - $\bar{x}_2$ = 27.7
   - s_2 = 9.5

2. **Enunciar las hipótesis apropiadas.**
   H_0: $\mu_1 = \mu_2$
   H_A: $\mu_1 \neq \mu_2$

Luego, hay que establecer si $\mu_1 = \mu_2$ o existe una diferencia estadísticamente significativa entre ellas. Se trata de precisar si los contenidos totales de fibra dietética en ambas dietas son promedialmente iguales o difieren significativamente.

3. **Escoger el nivel de significación**
   Nivel de significación = 5%

4. **Calcular el estadígrafo de la prueba**

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{[(Sp)^2/n_1 + (Sp)^2/n_2]}}.$$

con:
- $n_1 = 18$ pacientes
- $\bar{x}_1 = 42.7$
- $s_1 = 9.9$
- $n_2 = 9$ pacientes
- $\bar{x}_2 = 27.7$
- $s_2 = 9.5$

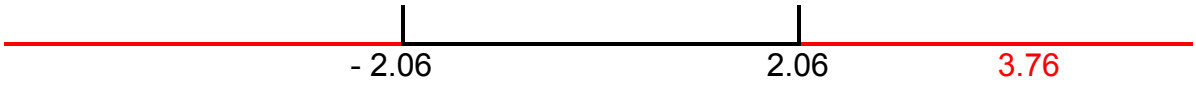
$$Sp^2 = \frac{S_1^2 (n_1 - 1) + S_2^2 (n_2 - 1)}{n_1 + n_2 - 2} = \frac{[(9.9)^2 \times (17)] + [(9.5)^2 (8)]}{18 + 9 - 2} = 95.53$$

de modo que:

$$t = \frac{42.7 - 27.7}{\sqrt{[95.53/18 + (95.53/9]}} = 3.76$$

5. **Construir las regiones crítica y de aceptación de H_0 y comparar el valor del estadígrafo calculado con el valor tabulado**
   Los grados de libertad de la t en esta prueba son: $n_1 + n_2 - 2$. Como tiene dos muestras esa formula se obtiene de $(n_1 - 1) + (n_2 - 1)$.
   Para construir las regiones necesarias se busca la t de Student tabulada correspondiente a $n_1 + n_2 - 2 = 25$ g.l. y nivel de significación 5%.

   Resulta que, t = 2.06. En consecuencia, las regiones son las que siguen.



El valor del estadígrafo se encuentra en la región de rechazo.

6. Aplicar el criterio de decisión

El valor del estadígrafo se encuentra en la región de rechazo. Se rechaza la hipótesis nula con el nivel de significación 5%, esto es, la prueba de hipótesis ofrece un resultado estadísticamente significativo con este nivel.

Para lograr la conclusión final tenga presente además que, la media muestral de la variable de interés es 42.7 en los sujetos sanos y, 27.7 en los enfermos.

7. Redactar la conclusión

Se admite que el promedio del contenido total de fibra dietética en la dieta de los sujetos sanos es significativamente mayor que en la de los enfermos, con el nivel de significación 5%.

Prueba de hipótesis sobre diferencias de Proporciones Poblacionales.

Esta prueba es una de las más utilizadas en el campo de la Medicina y la Salud Pública. Ella permite comparar dos porcentajes y ver si la diferencia entre ellos es estadísticamente significativa o no.

Parámetro y valores teóricos

- P_1 = Proporción poblacional desconocida en la población 1 de cierta característica de interés.
- P_2 = Proporción poblacional desconocida en la población 2 de la misma característica de interés

Hipótesis

Hipótesis nula $H_0: P_1 = P_2$

Hipótesis alternativa $H_A: P_1 \neq P_2$

Así, se trata de decidir si las proporciones poblacionales son iguales o difieren significativamente.

Estadígrafo de la prueba

$$Z = \frac{p_1 - p_2}{\sqrt{[P_0Q_0/n_1 + P_0Q_0/n_2]}}.$$

donde:

n_1: Tamaño de muestra de población 1

n_2 :Tamaño de muestra de población 2

p_1 :Porcentaje muestral de la cualidad estudiada en la muestra de la población 1

p_2 :Porcentaje muestral de la cualidad estudiada en la muestra de la población 2

P_0 : Porcentaje muestral **ponderada o combinada** de las dos muestras (es un porcentaje calculado sobre la base de la unión de ambas muestras)

$$P_0 = \frac{f_1 + f_2}{n_1 + n_2} \text{ x } 100$$

f_1 y f_2 : Frecuencias respectivas de aparición en las muestras de una característica determinada

Q_0 : $100 - P_0$

Ejemplo.
En el departamento de Fisioterapia de un hospital, se ha propuesto un nuevo procedimiento terapéutico para los pacientes con secuelas de politraumas. Durante dos años se efectuó un estudio para comparar el nuevo tratamiento con el utilizado tradicionalmente, obteniéndose los siguientes resultados:

Tratamiento	Pacientes tratados	Pacientes rehabilitados
Nuevo	200	186
Tradicional	800	592

Determine si los tratamientos producen resultados similares o significativamente distintos. Utilice el nivel de significación 1%.

Solución.

1. **Identificar los parámetros, valores teóricos de interés y los otros valores necesarios para el cálculo.**

- Parámetros: P1 y P2 = Porcentajes poblacionales (desconocidos) de pacientes rehabilitados respectivamente con el tratamiento nuevo y el tradicional.
- n_1 = 200 pacientes; f_1 = 186 pacientes
- n_2 = 800 pacientes; f_2 = 592 pacientes

2. **Enunciar las hipótesis apropiadas.**
   H_0: $P_1 = P_2$
   H_A: $P_1 \neq P_2$
   De este modo, hay que definir si los porcentajes en cuestión son iguales o significativamente distintos.

3. **Escoger el nivel de significación**
   Nivel de significación = 1%

4. **Calcular el estadígrafo de la prueba**

$$Z = \frac{p_1 - p_2}{\sqrt{[P_0Q_0/n_1 + P_0Q_0/n_2]}}.$$

para: n_1 = 200; n_2 = 800

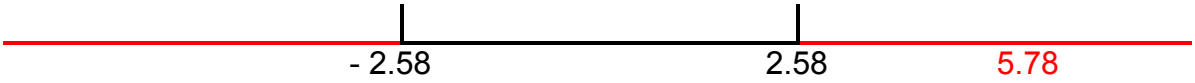
$p_1 = \frac{186}{200} \times 100 = 93\%$ $\qquad$ $p_2 = \frac{592}{800} \times 100 = 74\%$

$P_0 = \frac{186 + 592}{200 + 800} \times 100 = 77.8\%$ $\qquad$ $Q_0 = 100 - 77.8 = 22.2\%$

$$Z = \frac{93 - 74}{\sqrt{[77.8 \times 22.2/200 + 77.8 \times 22.2/800]}} = 5.78$$

5. **Construir las regiones crítica y de aceptación de H_0 y comparar el valor del estadígrafo calculado con el valor tabulado**

   Como nivel de significación es 1% el valor de Z que se toma es 2.58



El valor del estadígrafo se encuentra en la región de rechazo.

6. **Aplicar el criterio de decisión**

   El valor del estadígrafo se encuentra en la región crítica. Se rechaza la hipótesis nula con el nivel de significación 1%, esto es, la prueba de hipótesis produce un resultado estadísticamente significativo con este nivel.

   Para llegar a la conclusión final note además que, el porcentaje muestral de pacientes rehabilitados con el nuevo tratamiento es 93.0 y, con el clásico, 74.0.

7. **Redactar la conclusión**

   El nuevo tratamiento ofrece resultados significativamente mejores que el tratamiento tradicional, con el nivel de significación 1%.

Tabla de la t de Student

g.l.	Conf. 95% α = 0.05	Conf. 99% α = 0.01	g.l.	Conf. 95% α = 0.05	Conf. 99% α = 0.01
1	12.71	63.66	**20**	2.09	2.84
2	4.30	9.92	**21**	2.08	2.83
3	3.18	5.84	**22**	2.07	2.82
4	2.78	4.60	**23**	2.07	2.80
5	2.57	4.03	**24**	2.06	2.79
6	2.45	3.71	**25**	2.06	2.78
7	2.36	3.50	**26**	2.06	2.78
8	2.31	3.35	**27**	2.05	2.77
9	2.26	3.25	**28**	2.05	2.76
10	2.23	3.17	**29**	2.05	2.76
11	2.20	3.11	**30**	2.04	2.75
12	2.18	3.05	**40**	2.02	2.70
13	2.16	3.01	**50**	2.00	2.68
14	2.14	2.98	**60**	2.00	2.66
15	2.13	2.95	**70**	1.99	2.65
16	2.12	2.92	**80**	1.99	2.64
17	2.11	2.90	**90**	1.99	2.63
18	2.10	2.88	**100**	1.98	2.63
19	2.09	2.86	∞	1.96	2.58

INDICE

Capítulo 12
INFERENCIA ESTADISTICA. METODO DE CHI-CUADRADO

Autores: Dr. Emilio Morales Jiménez

SUMARIO

- *Relación entre variables.*
- *Tablas de contingencia.*
- *Estadígrafo Chi-cuadrado.*
- *Prueba no-paramétrica de independencia.*
- *Prueba no-paramétrica de homogeneidad.*
- *Tabla de Chi-cuadrado*

Relación entre variables

Lo que hemos visto hasta ahora de la estadística inferencial se refiere a una estimación y prueba de hipótesis para comparar dos valores pero de una sola variable. En Medicina y en Salud Pública no es suficiente ya que muy frecuentemente se requiere conocer si dos variables se encuentran relacionadas entre sí; por ejemplo, obesidad e hipertensión; stress e hipertensión; personalidad y úlcera gastro-duodenal; tratamiento y evolución del paciente e innumerables relaciones más. De hecho, para determinar la magnitud de factores de riesgo de determinada enfermedad es menester conocer previamente si hay relación entre el factor y la enfermedad. De ahí la importancia de que el profesional de la salud conozca el método del Chi-cuadrado que brinda la posibilidad de conocer si entre dos variables cualitativas existe o no relación.

La estimación y prueba de hipótesis de parámetros poblacionales son pruebas paramétricas porque están relacionadas con los parámetros poblacionales y sus resultados implican una decisión relativa a ellos. **Ahora bien, el método del Chi-cuadrado (también llamado Ji-cuadrado) se refiere a relación entre variables y no a parámetros estando incluida en las pruebas conocidas con el nombre de no-paramétricas.** Sus resultados se infieren a la población pero en correspondencia con la relación entre las variables analizadas.

El método del Chi-cuadrado, entre otras cosas, permite precisar:

(a) si existe alguna correspondencia entre dos variables, vale decir, si el comportamiento de una de ellas influye en la manera de conducirse la otra; o sea aplicar la **prueba no-paramétrica de independencia**.

(b) si varios grupos se distribuyen desigualmente a lo largo de las categorías de clasificación de una misma variable, ocasión en la cual esta variable establece una distinción entre los grupos; o sea aplicar la **prueba no-paramétrica de homogeneidad**.

Destaquemos que, en todo problema del segundo tipo también existen dos variables: la variable de grupo y la variable clasificadora de los elementos de cada grupo.

Se mencionó en el primer párrafo que la relación a que nos referimos es entre variables categóricas o cualitativas. Vale decir que se puede aplicar también con variables numéricas o cuantitativas con escala de medición discreta. En el caso de variables numéricas continuas se puede aplicar pero haciendo una transformación de la variable utilizando una escala categórica ordinal para medirlas. Por ejemplo, supongamos que tenemos el valor de la hemoglobina que es numérica continua y se desea aplicar este método con esa variable y otra cualquiera. Con el valor de la hemoglobina se pudiera utilizar la siguiente escala:

Hemoglobina < 11---------- bajo nivel de hemoglobina
Hemoglobina ≥ 11 a ≤ 13 ---------- nivel normal de hemoglobina
Hemoglobina > 13 ---------- nivel alto de hemoglobina

Para aplicar el método del Chi-cuadrado en las pruebas no-paramétricas de independencia y homogeneidad hay que elaborar primero que todo una **tabla de contingencia**.

Tablas de contingencia

Una tabla de contingencia (también llamada **tabla de r x k** o **tabla de n x m**) se utiliza para precisar si hay relación entre dos variables. Es una tabla de doble entrada realizada con los datos de dos variables.

Por ejemplo: Se estudia la posible relación entre el grupo sanguíneo y la severidad de cierta enfermedad en una población.

Severidad de la enfermedad	Grupo sanguíneo				Totales
	A	B	AB	O	
Leve	543	211	90	476	**1320**
Moderada	44	22	8	31	**105**
Grave	28	9	7	31	**75**
Totales	**615**	**242**	**105**	**538**	**1500**

La tabla anterior es una **tabla de contingencia (3 x 4)**, esto es, de tres filas y cuatro columnas, en la cual se ha recogido la información de 1500 sujetos. Tiene **tres filas** pues solamente se consideran las categorías de la **variable "severidad de la enfermedad" (leve, moderada y grave)** y **cuatro columnas** pues son los **grupos sanguíneos correspondientes (A; B; AB; O)**. La fila de totales y la columna de totales no se consideran para clasificar la tabla. Dicho de otro modo, **ni la fila ni la columna de totales figuran en el cómputo del tamaño de la tabla**.

Fíjese que la tabla tiene una primera columna de títulos y una primera o más filas de títulos. Las celdas de datos que en el ejemplo son 12, ya que es una tabla de (3 x 4), están ocupadas por la frecuencia absoluta de los datos que aquí se llaman **"frecuencias observadas"**. Las frecuencias observadas se llaman así por ser las que se determinaron u observaron durante la investigación o el estudio realizado. Las frecuencias observadas se denotan por la letra **"Oij"** donde i es el número de la fila y j el número de la columna.

Así, en el ejemplo el valor 543 está en la celda O_{11} (fila 1 y columna 1) y el valor 211 en la celda O_{12} (fila 1 y columna 2).

Los totales de las columnas se encuentran en la última fila y los totales de las filas se encuentran en la última columna y se les llama **"totales marginales"** de las columnas y de las filas respectivamente. Así, 1320 es el total marginal de la fila "leve" y 615 es el total marginal de la columna del "grupo sanguíneo A". Se denotan por **"n_i"** los totales marginales de las filas y **"n_j"** los totales marginales de las columnas

La última celda de la tabla (extremo inferior derecho, ocupada por 1500) se conoce como **"gran total"** y es igual al número de elementos que intervienen en el estudio, o sea, al tamaño de la muestra. El gran total se denota por **"N"**.

Estadígrafo Chi-cuadrado.

Una vez que se tiene la tabla de contingencia se puede calcular el valor del estadígrafo Chi-cuadrado que se denota por **"X^2".**

Lo primero a ejecutar es obtener para cada celda de datos la **"frecuencia esperada"**, o sea, **el valor que debiera estar en la celda si las dos variables no se relacionan entre sí.** El valor esperado se denota por **"Eij",** donde i es el número de la fila y j el número de la columna. El valor esperado de cada celda se obtiene de la siguiente forma:

$$Eij = \frac{(ni)(nj)}{N} = \frac{\text{Total marginal de la fila x Total marginal de la columna}}{\text{Gran total o tamaño de la muestra}}$$

donde: n_i = total marginal de la fila
n_j = total marginal de la columna
N = gran total o tamaño de la muestra

Las frecuencias esperadas se pueden definir como las frecuencias teóricas que hipotéticamente debieran aparecer en las celdas de datos de una tabla de contingencia si las variables no están relacionadas entre sí.

Veamos un ejemplo. Una muestra de 200 pacientes se extrae de una población y se quiere averiguar si la hipertensión arterial está relacionada con la obesidad. Una vez clasificados los pacientes por las dos variables se obtiene la siguiente tabla de contingencia.

FRECUENCIAS OBSERVADAS DE OBESIDAD E HIPERTENSION ARTERIAL EN LOS PACIENTES			
	OBESIDAD	**NO OBESIDAD**	**TOTALES**
HIPERTENSION	32	28	**60**
NO HIPERTENSION	18	122	**140**
TOTALES	**50**	**150**	**200**

Las frecuencias observadas son las siguientes: $\mathbf{O_{11}}$= 32; $\mathbf{O_{12}}$= 28; $\mathbf{O_{21}}$= 18 y $\mathbf{O_{22}}$= 122.

Ahora se pueden calcular las frecuencias esperadas o sea, los valores que debieran estar en la tabla si la obesidad y la hipertensión no estuvieran relacionadas entre sí.
Las frecuencias esperadas serían las siguientes:

$$\mathbf{E_{11}}= \frac{60 \times 50}{200} = 15; \quad \mathbf{E_{12}}= \frac{60 \times 150}{200} = 45; \quad \mathbf{E_{21}}= \frac{140 \times 50}{200} = 35; \quad \mathbf{E_{22}}= \frac{140 \times 150}{200} = 105$$

La tabla de contingencia con las frecuencias esperadas fuera la siguiente:

FRECUENCIAS ESPERADAS DE OBESIDAD E HIPERTENSION ARTERIAL EN LOS PACIENTES			
	OBESIDAD	**NO OBESIDAD**	**TOTALES**
HIPERTENSION	15	45	**60**
NO HIPERTENSION	35	105	**140**
TOTALES	**50**	**150**	**200**

Fíjese que los totales marginales en la tabla de contingencia con las frecuencias esperadas son los mismos que en la tabla de contingencia con las frecuencias observadas. Los únicos que cambian son los valores en las celdas de datos.

Una vez conocidas las frecuencias observadas que son resultado de los datos obtenidos con la muestra y calculadas las frecuencias esperadas que son las frecuencias que debieran estar en las celdas si las variables no se relacionan entre sí, es fácil entender que **una medida de la diferencia entre las frecuencias observadas y las frecuencias esperadas mostrará si las variables se relacionan entre sí o no se relacionan**. Si la diferencia es pequeña, las frecuencias observadas y esperadas se pueden considerar similares y las variables no se relacionan entre sí. Si la diferencia es grande, entonces las frecuencias observadas y esperadas se consideran distintas y existe cierta relación entre las variables. Esa medida está dada por el estadígrafo Chi-cuadrado.

El estadígrafo Chi-cuadrado se denota por "X^2" y se define como la medida de la diferencia entre las frecuencias observadas y las frecuencias esperadas en una tabla de contingencia.

Este estadígrafo tiene su propio modelo de distribución que comienza en 0 y tiende al infinito.

Se calcula por la siguiente fórmula:

$$X^2 = \Sigma \{ (O_{ij} - E_{ij})^2 / E_{ij} \}$$

ó sea, la **sumatoria del resultado de dividir el cuadrado de la diferencia de la frecuencia observada menos la frecuencia esperada entre la frecuencia esperada para cada celda de datos de la tabla de contingencia**.

Por ejemplo, calculemos el Chi-cuadrado para la tabla de contingencia de la obesidad e hipertensión arterial que se expuso anteriormente:

	Frecuencia observada	Frecuencia esperada
Pacientes con hipertensión y obesidad	O_{11} = 32	E_{11} = 15
Pacientes con hipertensión y no obesidad	O_{12} = 28	E_{12} = 45
Pacientes sin hipertensión y obesidad	O_{21} = 18	E_{21} = 35
Pacientes sin hipertensión y no obesidad	O_{22} = 122	E_{22} = 105

Para calcular el estadígrafo Chi-cuadrado tenemos:

$$X^2 = \frac{(32 - 15)^2}{15} + \frac{(28 - 45)^2}{45} + \frac{(18 - 35)^2}{35} + \frac{(122 - 105)^2}{105} =$$

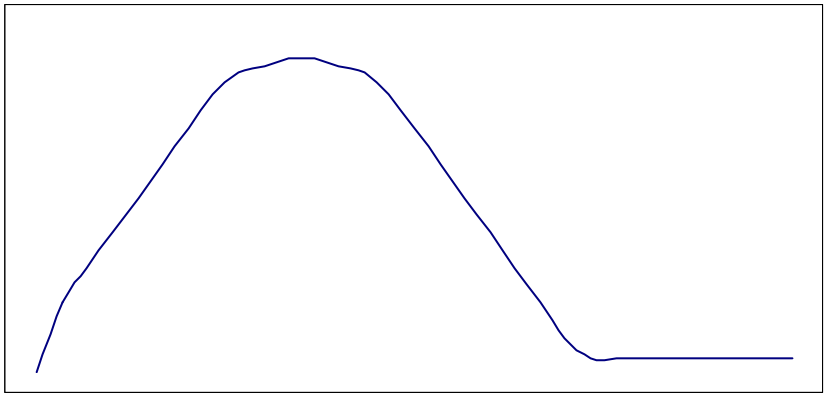
$$X^2 = 19.26 + 6.42 + 8.25 + 2.75 =$$

$$X^2 = 36.88$$

Propiedades del estadígrafo Chi-cuadrado

1. El Chi-cuadrado calculado para una tabla de contingencia constituye una expresión cuantitativa de la discrepancia entre las frecuencias observadas y las frecuencias esperadas.
2. $X^2 \geq 0$
3. $X^2 = 0$ indica una coincidencia total entre las frecuencias observadas y las esperadas. En consecuencia, se tiene en esta situación una total ausencia de relación entre las variables.
4. El valor de X^2 se incrementa cuando aumenta la diferencia entre las frecuencias observadas y las esperadas.

Distribución Chi-cuadrado
La distribución Chi-cuadrado es una curva que comienza en valor 0 (a la izquierda), sube en forma de semicírculo, hace como una campana y baja pero no llega a la línea base y se prolonga hasta el infinito.



Prueba no-paramétrica de independencia
Se parte de una tabla de contingencia originada por dos variables de clasificación (atributos, criterios, factores) aplicadas a una misma muestra representativa de una población. Esta muestra, clasificada de acuerdo con las categorías de las variables en cuestión, produce la tabla de contingencia que sirve de base para el análisis.

El interés radica en conocer si las dos variables se comportan de manera independiente o existe dependencia entre ellas. **Si las variables tienen un comportamiento independiente significa que no hay asociación o relación entre ellas**, por el contrario, **si existe dependencia entre las variables es que se asocian o relacionan significativamente.**

Una asociación estadísticamente significativa no asegura necesariamente una relación causa-efecto, ella expresa, sencillamente, que las dos variables se vinculan de alguna manera esencial.

Hipótesis
Se plantean las hipótesis estadísticas que en este caso son las siguientes:

H_0: Las variables de clasificación son independientes.
H_A: Las variables de clasificación no son independientes.

También se pudieron haber escrito de la siguiente forma:
H_0: Las variables de clasificación no están asociadas.
H_A: Las variables de clasificación están asociadas significativamente.

H_0: Las variables de clasificación no se relacionan.
H_A: Las variables de clasificación se relacionan significativamente.

Se trata de decidir si las variables de clasificación tienen un comportamiento independiente o, por el contrario, están asociadas significativamente en la población correspondiente.

Toma de decisión

Para tomar la decisión de aceptar o rechazar la hipótesis nula hay que tomar en consideración el modelo teórico de distribución del Chi-cuadrado y las propiedades del estadígrafo $\mathbf{X^2}$**.** Se dijo antes que si $\mathbf{X^2 = 0}$ indicaba total independencia de las variables y también que si el valor del $\mathbf{X^2}$ incrementaba, indicaba que las diferencias entre las frecuencias observadas y las esperadas también incrementaban.

El problema radica en conocer cuando el valor del $\mathbf{X^2}$ calculado es lo suficientemente grande para rechazar la hipótesis nula. Como observamos en el gráfico, la distribución Chi-cuadrado comienza en 0 y a medida que aumenta se mueve hacia la derecha de la distribución tendiendo al infinito, mostrando cada vez menor independencia entre las variables. La cuestión radica en determinar un **"punto crítico o frontera"** debajo de la curva que permita determinar si el $\mathbf{X^2}$ calculado se puede considerar lo suficientemente grande para rechazar $\mathbf{H_0}$.

Ese punto crítico que llamaremos **"$\mathbf{X^2}$ tabulado"** se determina por el nivel de significación que hayamos escogido para la prueba (5% ó 1%) y los grados de libertad de la tabla de contingencia buscándose en la tabla de la distribución Chi-cuadrado. (Ver tabla al final). Los grados de libertad se calculan por la fórmula:

$$\mathbf{G.L. = (n - 1)(m - 1)}$$

donde n = número de filas de la tabla de contingencia
m = número de columnas de la tabla de contingencia

(Recuerde que solamente se consideran las filas y columnas de las frecuencias observadas.)

Por ejemplo: Para un nivel de significación del 5% y 2 grados de libertad, el valor del $\mathbf{X^2}$ **tabulado = 5.99** (ver tabla). Cuando en la tabla hay un salto en los grados de libertad es porque se utiliza el valor precedente.

La decisión de rechazar o aceptar $\mathbf{H_0}$ se toma por la comparación del $\mathbf{X^2}$ **calculado con el** $\mathbf{X^2}$ **tabulado.** Ello se puede esquematizar de la siguiente forma:



Si el valor del **X^2 calculado cae a la derecha del X^2 tabulado**, cae en el área crítica y se rechaza **H_0,** aceptándose **H_A**. Dicho de otra forma si el **X^2 calculado > X^2 tabulado** se rechaza **H_0.**

Si el valor del **X^2 calculado cae a la izquierda del X^2 tabulado**, cae en el área de aceptación y se acepta **H_0**. Dicho de otra forma si el **X^2 calculado ≤ X^2 tabulado** se acepta **H_0.**

Ejemplo. Se desea conocer con un nivel de significación del 5%, si la ingestión de grasa animal y el padecimiento de alguna enfermedad hepática son características independientes o están relacionadas. Con este propósito se examinan 237 individuos, los cuales quedan clasificados como sigue:

Ingestión de grasa animal	Enfermedad Hepática		Totales
	Si	No	
Superior al máximo aceptable	13	72	85
Igual o menor que el máximo aceptable	4	148	152
Totales	17	220	237

Solución.

Hipótesis

H_0: La ingestión de grasa animal y el padecimiento de alguna enfermedad hepática **son independientes.**

H_A: La ingestión de grasa animal y el padecimiento de alguna enfermedad hepática **no son independientes.**

Cálculo del X^2

- **Se calculan las frecuencias esperadas:**

$$\mathbf{E_{11}} = \frac{85 \times 17}{237} = 6.1 \quad \mathbf{E_{12}} = \frac{85 \times 220}{237} = 78.9; \quad \mathbf{E_{21}} = \frac{152 \times 17}{237} = 10.9; \quad \mathbf{E_{22}} = \frac{152 \times 220}{237} = 141.1$$

- **Se calcula el X^2**

$$X^2 = \Sigma \{ (O_{ij} - E_{ij})^2 / E_{ij} \}$$

$$\mathbf{X^2} = \frac{(13 - 6.1)^2}{6.1} + \frac{(72 - 78.9)^2}{78.9} + \frac{(4 - 10.9)^2}{10.9} + \frac{(148 - 141.1)^2}{141.1} =$$

$$\mathbf{X^2} = 7.80 + 0.60 + 4.36 + 0.33 =$$

$$\mathbf{X^2} = 13.09$$

- Se determina el X^2 tabulado (valor crítico)

Nivel de significación = 5% o sea, $\alpha = 0.05$;

G.L.= (n–1)(m–1) = (2 – 1)(2 – 1) = (1)(1) = 1

El valor del **X^2 tabulado** = 3.84

Toma de decisión

Area de aceptación de H_0	Area crítica o de rechazo de H_0
0	3.84
	X^2 tabulado
	13.09
	X^2 calculado

X^2 calculado > X^2 tabulado, se rechaza la hipótesis nula y se acepta la hipótesis alternativa con el nivel de significación 5%, esto es, la prueba de hipótesis produce un resultado estadísticamente significativo con este nivel.

Para llegar a la conclusión final tenga presente además que:

El % de enfermos en aquellos con consumo elevado de grasa = 13/85 x 100 = 15.3

El % de enfermos en aquellos con consumo adecuado de grasa = 4/152 x 100 = 2.6

Conclusión

Existe asociación entre la ingestión de grasa animal y el padecimiento de alguna enfermedad hepática, de manera que, el porcentaje de enfermos hepáticos entre los sujetos consumidores exagerados de grasa es significativamente superior que entre aquellos cuyo consumo es adecuado, con el nivel de significación 5%.

Prueba no-paramétrica de homogeneidad

Se estudiará ahora el problema de saber si dos o más grupos (poblaciones) son homogéneos respecto a cierta variable (atributo, criterio, factor), esto es, conocer si se distribuyen de manera similar o significativamente distinta en las distintas categorías de la variable. Para esto, se escoge una muestra representativa de cada uno de los grupos. Estas muestras, clasificadas con arreglo a las categorías en cuestión, originan la tabla de contingencia que sirve de soporte para el estudio.

Hipótesis

H_0: Los grupos se distribuyen de manera homogénea respecto a la variable.
H_A: Los grupos no se distribuyen de manera homogénea respecto a la variable.

De este modo, se trata de decidir si los grupos tienen distribuciones semejantes o, por el contrario, exhiben distribuciones significativamente distintas.

Si se rechaza **H_0** aceptándose **H_A**, los grupos no tienen una distribución homogénea con relación a la variable, luego existe algún tipo de asociación entre la variable y el criterio por el cual se formaron los grupos.

El cálculo del Chi-cuadrado es igual al ya visto para la prueba de independencia así como los criterios para la toma de decisión. Lo único que diferencia ambas pruebas son las hipótesis. La tabla de contingencia es igual pero con la diferencia de que aquí se saben de antemano los totales de los grupos que integran el estudio, que no se sabían en la

prueba de independencia hasta que no se clasificaran por las dos variables todos los individuos de la muestra.

Ejemplo. Se tienen tres grupos de pacientes previamente clasificados por el tamaño de las amígdalas en normales, aumentadas e hipertróficas con 516, 589 y 293 pacientes respectivamente. Se desea conocer con un nivel de significación del 1% si la presencia de estreptococo B hemolítico en orofaringe es o no homogénea entre los grupos. Una vez determinada la presencia del germen en la orofaringe de cada paciente se tiene los siguientes resultados:

Estreptococo B Hemolítico	**Tamaño de las amígdalas**			**Totales**
	Normales	**Aumentadas**	**Hipertróficas**	
Presente en orofaringe	19	29	24	**72**
No presente en orofaringe	497	560	269	**1326**
Totales	**516**	**589**	**293**	**1398**

Solución.

Hipótesis

H_0: El estreptococo B hemolítico **se distribuye homogéneamente** entre los grupos de pacientes según el tamaño de las amígdalas.

H_A: El estreptococo B hemolítico **no se distribuye homogéneamente** entre los grupos de pacientes según el tamaño de las amígdalas.

Cálculo del X^2

- **Se calculan las frecuencias esperadas:**

	Frecuencia observada	**Frecuencia esperada**
Estreptococo y amígdalas normales	$O_{11} = 19$	$E_{11} = \frac{72x516}{1398} = 26.57$
Estreptococo y amígdalas aumentadas	$O_{12} = 29$	$E_{12} = \frac{72x589}{1398} = 30.33$
Estreptococo y amígdalas hipertróficas	$O_{13} = 24$	$E_{13} = \frac{72x293}{1398} = 15.09$
No estreptococo y amígdalas normales	$O_{21} = 497$	$E_{21} = \frac{1326x516}{1398} = 489.42$
No estreptococo y amígdalas aumentadas	$O_{22} = 560$	$E_{22} = \frac{1326x589}{1398} = 558.66$
No estreptococo y amígdalas hipertróficas	$O_{23} = 269$	$E_{23} = \frac{1326x293}{1398} = 277.90$

- **Se calcula el X^2**

$$X^2 = \Sigma \{ (O_{ij} - E_{ij})^2 / E_{ij} \}$$

$$X^2 = \frac{(19-26.57)^2}{26.57} + \frac{(29-30.33)^2}{78.9} + \frac{(24-15.09)^2}{15.09} + \frac{(497-489.42)^2}{489.42} + \frac{(560-558.66)^2}{558.66} + \frac{(269-277.90)^2}{277.90} =$$

$$X^2 = 2.15 + 0.05 + 5.26 + 0.12 + 0.003 + 0.28 =$$

$$X^2 = 7.863$$

- Se determina el X^2 tabulado (valor crítico)
Nivel de significación = 1% o sea, $\alpha = 0.01$;
G.L.= (n–1)(m–1) = (2 – 1)(3 – 1) = (1)(2) = 2
El valor del **X^2 tabulado** = 9.21

Toma de decisión

Area de aceptación de H_0		Area crítica o de rechazo de H_0
0	**7.863**	**9.21**
	X^2 calculado	**X^2 tabulado**

X^2 calculado ≤ X^2 tabulado, se acepta la hipótesis nula con el nivel de significación 1%, esto es, la prueba de hipótesis no produce un resultado estadísticamente significativo con este nivel.

Conclusión
No existe asociación entre la presencia de estreptococo en la orofaringe y el tamaño de las amígdalas en los pacientes con el nivel de significación 1%.

TABLA DE CHI-CUADRADO

g.l.	α = 5%	α = 1%	g.l.	α = 5%	α = 1%
1	3.84	6.63	**16**	26.30	32.00
2	5.99	9.21	**17**	27.59	33.41
3	7.81	11.34	**18**	28.87	34.81
4	9.49	13.28	**19**	30.14	36.19
5	11.07	15.09	**20**	31.41	37.57
6	12.59	16.81	**25**	37.65	44.31
7	14.07	18.48	**30**	43.77	50.89
8	15.51	20.09	**40**	55.76	63.69
9	16.92	21.67	**50**	67.50	76.15
10	18.31	23.21	**60**	79.08	88.38
11	19.68	24.73	**70**	90.53	100.42
12	21.03	26.22	**80**	101.9	112.3
13	22.36	27.69	**90**	113.1	124.1
14	23.68	29.14	**100**	124.3	135.8
15	25.00	30.58	**120**	146.57	158.95

INDICE

Capítulo 13
INFERENCIA ESTADISTICA. CORRELACION Y REGRESION LINEAL

Autores: Dr. Emilio Morales Jiménez

SUMARIO

- *Relación entre variables. Problemas de correlación y regresión lineal.*
- *Diagrama de dispersión. Utilidad.*
- *Coeficiente de correlación lineal poblacional o de Pearson **(ρ)** y Coeficiente de correlación lineal muestral **(r).***
- *Prueba de hipótesis sobre el Coeficiente de correlación lineal poblacional.*
- *La regresión lineal.*

Relación entre variables. Problemas de correlación y regresión lineal.

Otro aspecto de la inferencia estadística es la relación entre variables cuantitativas continuas. En Medicina se trabaja con muchas variables numéricas continuas y en buena parte de ellas es menester conocer si existe relación entre ellas. Hoy veremos un tipo especial de relación entre este tipo de variables, la correlación lineal y la regresión lineal.

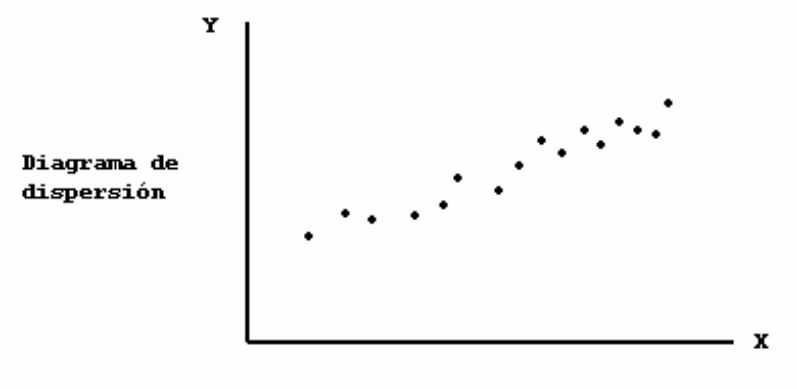
Esta relación específica se utiliza con alguna frecuencia en Medicina para poder predecir el valor de una variable **(Y)** conociendo el valor de otra variable **(X)**. Para ello, una vez establecido que existe una correlación lineal entre las dos variables se procede a realizar una regresión lineal para hallar el valor del **intercepto (a)** y la **pendiente (b)** que permiten establecer la ecuación de regresión **y = a + bX**.

La correlación y regresión lineal se utilizan frecuentemente en laboratorio clínico y en la determinación de peso y talla de pacientes, aunque se puede emplear en todas las especialidades. El coeficiente de correlación es una **medida de la relación entre dos variables numéricas continuas cuyos valores se encuentran pareados.** Se trabajará con dos variables cuantitativas continuas **X** y **Y** cuyas observaciones se hacen en parejas: **(X,Y)**. La pareja **(X,Y)** puede interpretarse como una nueva variable, de tipo bivariado o bidimensional. En resumen se dispondrá siempre de n pares ó parejas:

$$(X_1,Y_1);\ (X_2,Y_2);\ \ldots\ ;\ (X_n,Y_m).$$

Diagrama de dispersión. Utilidad

El diagrama de dispersión o nube de puntos es la representación gráfica de las duplas **(X,Y)**. Cada una de estas duplas es un punto del plano y, por tanto, puede presentarse en un sistema de dos ejes de coordenadas: el eje X o eje de las abscisas (eje horizontal) y, el eje Y o eje de las ordenadas (eje vertical). A continuación se muestra un ejemplo:



El diagrama de dispersión es conveniente realizarlo cuando pensamos en una posible correlación lineal entre dos variables. La forma que toma la distribución de los puntos en el diagrama da una idea de si puede o no existir una relación lineal entre las dos variables.

Coeficiente de correlación lineal poblacional o de Pearson (ρ) y Coeficiente de correlación lineal muestral (r).

El **grado de dispersión** de la nube de puntos en el diagrama de dispersión se determina en la población por el **coeficiente de correlación lineal de Pearson** representado por ρ **(ro).** Como los valores de los parámetros poblacionales por lo general no se conocen es necesario estimarlos con los datos de una muestra representativa de la población. Para ello se utiliza el **coeficiente de correlación muestral o real** representado por **r = r(X,Y).** En ambos casos **el coeficiente de correlación lineal** es un número que **mide el grado de dispersión** de la nube de puntos que constituye el diagrama de dispersión, **al mismo tiempo que evalúa la tendencia a estar alineados** que esos puntos tienen.

En definitiva, **r = r(X,Y)** es una magnitud que expresa el ajuste de los puntos a cierta línea recta óptima: la recta de mejor ajuste.

La recta que mejor se ajusta al diagrama de dispersión se determina mediante el llamado **criterio de mínimos cuadrados**, razón por la cual esa recta se denomina la **recta mínimo-cuadrática**. Esta recta **no cumple la propiedad de pasar por el mayor número posible de puntos** del diagrama de dispersión **sino ser la recta que tiene menor distancia entre ella y los puntos**.

Las fórmulas para el cálculo de **r** se encuentran en cualquier libro de estadística, las mismas son por lo general bastante complejas para el personal de la salud y lo que nos interesa es que el médico pueda interpretar y analizar el resultado del coeficiente de correlación lineal para su utilización en aplicaciones médicas.

Propiedades de ρ y de r

Las propiedades de ρ y de **r** son las mismas, siendo las siguientes:

1) $-1 \leq r \leq +1$; **r** es un valor entre –1 y +1
2) **r(X,Y) = r(Y,X);** el orden de las variables no influye en el valor de **r** y es el mismo para las dos posibles combinaciones.
3) El valor del coeficiente de correlación no depende de las unidades de medida utilizadas para las variables. Por ejemplo, la correlación entre estatura y peso será idéntica independientemente de si se utilizan como unidades de medida pulgadas y libras, o bien, centímetros y kilogramos.
4) Los **valores positivos** de **r** indican una **relación directa** entre las variables, esto es, cuando una aumenta la otra también lo hace; los **valores negativos** denotan una **relación inversa**, es decir, al aumentar una variable la otra disminuye.
5) Si **r = 1** expresa una correlación lineal perfecta entre las dos variables. Si es **+1** expresa una correlación **positiva perfecta**, en tanto que si es **-1** manifiesta una correlación **negativa perfecta**.
6) Si **r = 0**, no existe una correspondencia lineal entre las variables.
7) Los valores de **r** próximos a 1 (bien sea –1 ó +1) indican un alto grado de linealidad.
8) Los valores de **r** cercanos a cero señalan el alejamiento de la linealidad.

Está muy difundida una idea considerablemente errónea acerca de la interpretación del coeficiente de correlación:

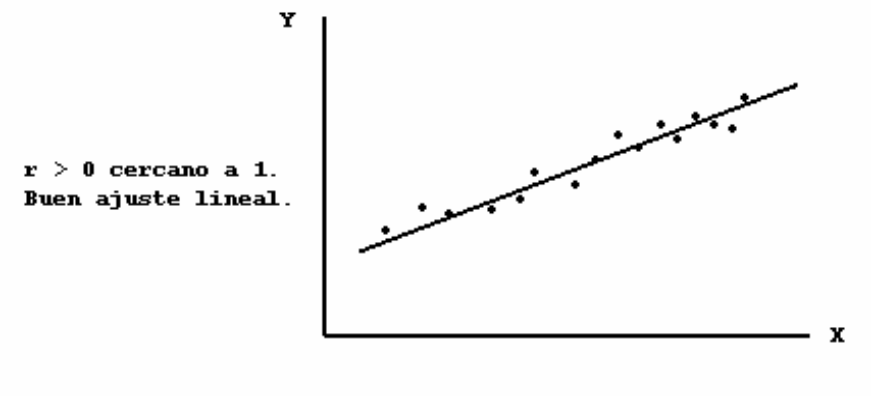
- Un valor de **r** próximo a 0 sólo indica una relación **lineal** grosera entre **X** y **Y** o la ausencia de la misma, pero, no impide la posibilidad de alguna relación **no lineal**.
- Por otra parte, no hay forma de establecer por una vía puramente algebraica un criterio objetivo que permita distinguir entre un **r** "grande" (cercano en valor absoluto a 1) y uno "pequeño" (próximo a 0),
- **Es la Estadística quien desarrolla los métodos para decidir si un coeficiente de correlación encontrado se debe al azar o a que existe una relación lineal entre las variables (resultado estadísticamente significativo), lo cual es mucho más trascendente que el mero orden de magnitud de dicho coeficiente. Esta es la prueba de hipótesis de ρ que se verá mas adelante.**

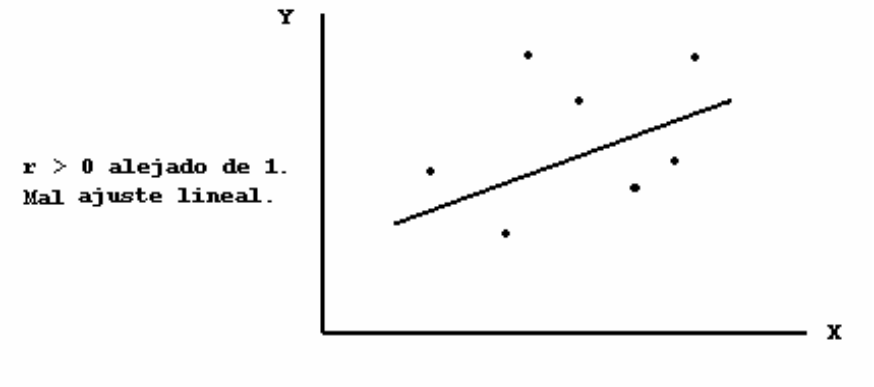
Se han propuesto diversos sistemas de intervalos para categorizar la intensidad de la posible relación lineal representada por **r**. Una escala muy razonable es la siguiente:

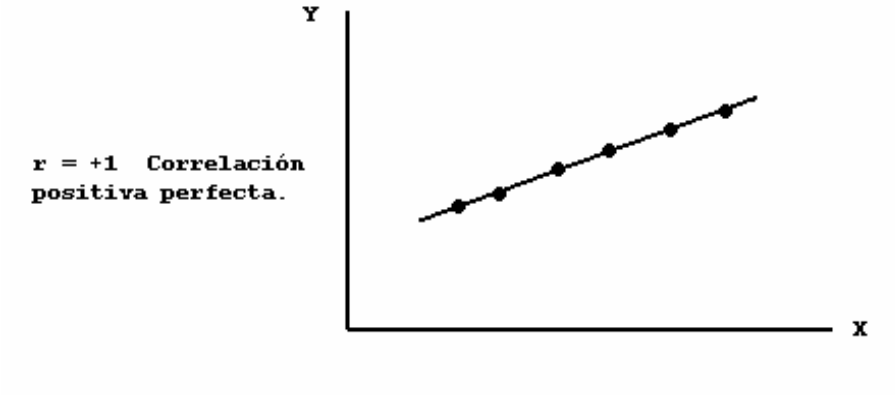
❖ De –1 a -0.9	Negativa fuerte
❖ De -0.8 a -0.5	Negativa moderada
❖ De -0.4 a 0	Negativa débil
❖ De 0 a 0.4	Positiva débil
❖ De 0.5 a 0.8	Positiva moderada
❖ De 0.9 a 1	Positiva fuerte

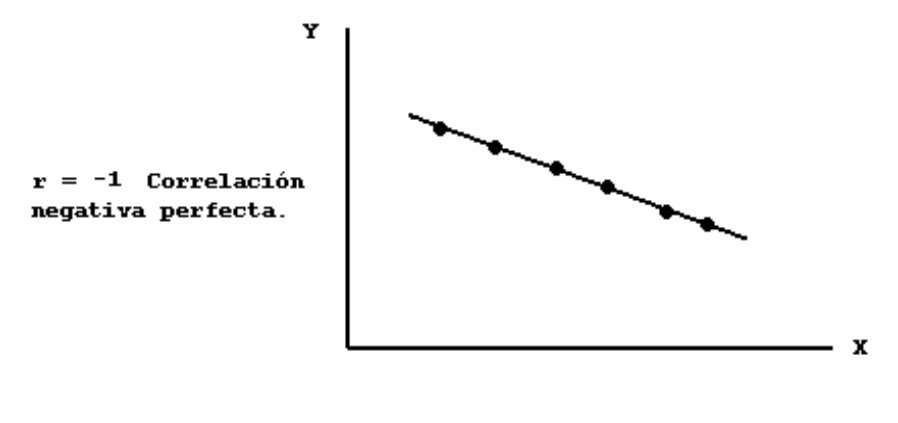
El orden de magnitud del coeficiente que puede considerarse satisfactorio depende notablemente del campo de estudio, empero, como ya se dijo, lo sustancial es su (posible) significación estadística.

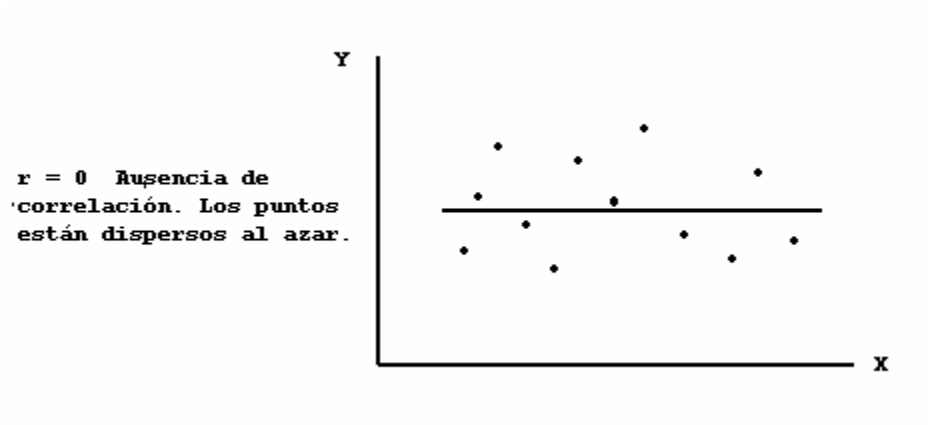
Los siguientes gráficos esclarecen las ideas presentadas hasta ahora.











Todo este análisis gráfico ilustra el sentido que tiene el coeficiente de correlación lineal de Pearson **(ρ)** o el coeficiente de correlación muestral **r = r(X,Y)**.

Ahora bien, lo más importante es la posible **significación estadística de r**, esto es, la cualidad de manifestar una característica **poblacional** de la pareja de variables **(X,Y)**. Esta problemática se examinará a continuación a través de una prueba de hipótesis paramétrica.

Prueba de hipótesis sobre el Coeficiente de correlación lineal poblacional.

Parámetro y valor teórico

Parámetro: ρ = Coeficiente de correlación poblacional desconocido de la pareja de variables (X,Y)

Valor teórico: 0 = Valor (hipotético) especificado.

Hipótesis

Hipótesis nula **H_0: $\rho = 0$** (**No hay correlación lineal**)

Hipótesis alterna **H_A : $\rho \neq 0$** (**Correlación lineal estadísticamente significativa**)

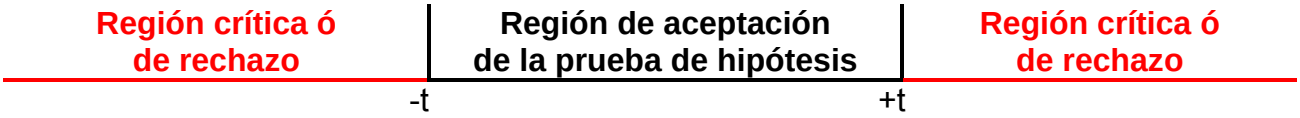
Problema de prueba de hipótesis

De este modo, se trata de decidir si el coeficiente de correlación poblacional tiene el valor nulo o su valor difiere significativamente de cero, existiendo en este caso una correlación lineal.

Nivel de significación

5% ó 1%, o sea $\alpha = 0.05$ ó 0.01 respectivamente.

Regiones



Aquí, t es un valor denominado "t de Student", el cual se halla **tabulado**, o sea, en la tabla de la curva de "t de Student".

Para hallar la "t" en la tabla correspondiente, debe tenerse en cuenta que ella depende de dos factores:

1) El nivel de significación, o sea si $\alpha = 0{,}01$ ó 0,05;
2) Los **grados de libertad (g.l)**, que en este caso es el tamaño de la muestra rebajado en dos unidades, (n – 2), **n** es la cantidad de parejas **(X,Y)** incluidas en la muestra.

Para buscar en la tabla de la t de Student se tiene que considerar que cuando **n – 2 < 30** es ineludible buscar "t" en la tabla, pero, si **n – 2** $\geq$ **30** sucede que, t = 1.96 si $\alpha = 0.05$ y, t = 2.58 si $\alpha = 0.01$, con lo cual la tabla se hace superflua (note que, en esta situación, los grados de libertad no afectan al valor de t).

Advierta ahora que, la región crítica es la unión de dos segmentos de recta, esto es, de dos intervalos, lo cual es característico de una prueba de dos colas. De acuerdo con la teoría general, la región de aceptación de la hipótesis nula es el complemento de la región crítica en la recta numérica total.

Estadígrafo de la prueba

$$t = r\sqrt{\frac{n-2}{1-r^2}}$$

siendo:

n = Tamaño muestral = Cantidad de parejas **(X,Y)** existentes
r = Coeficiente de correlación muestral.

Criterio de decisión

1. Se rechaza la hipótesis nula a favor de la hipótesis alternativa si el valor encontrado para el estadígrafo de la prueba se encuentra en la región crítica **(resultado estadísticamente significativo)**.
2. Si, por el contrario, el valor mencionado se halla en la región de aceptación de la hipótesis nula, esta no se rechaza, en detrimento de la alterna **(desenlace estadísticamente no significativo)**.

Observación. En esta prueba de hipótesis aparecen dos valores de t: uno de ellos es el tabulado y el otro es el calculado. La decisión estadística se alcanza, en definitiva, comparando estos valores.

- Si el valor de t calculado es menor o igual que el valor de t tabulado **($t_{calculado} \leq t_{tabulado}$)** se **acepta la hipótesis nula**.
- Si el valor de t calculado es mayor que el valor de t tabulado **($t_{calculado} > t_{tabulado}$)** se **rechaza la hipótesis nula y se acepta la alternativa**.

Precaución. Las variables X y Y se miden en ó mediante la misma entidad, física o conceptual, llamada la "unidad de asociación". Por ejemplo, si existe interés en estudiar la correspondencia entre estatura y peso en humanos, estas dos medidas se toman en el mismo individuo. Es posible, sin embargo, plantearse la investigación del nexo cuantitativo entre las estaturas de las madres y los pesos de los hijos, ya que aquí el parentesco constituye la unidad de asociación.

Ejemplo. Se desea comparar dos métodos de medir el gasto cardíaco en cierto tipo de animales experimentales. Al estudiar 10 de estos, resultan los siguientes datos para el gasto cardíaco en l/min.

Método I (X)	1.3	1.0	0.8	1.4	1.5	1.4	2.7	3.0	2.0	2.4
Método II (Y)	1.1	1.2	0.5	1.3	1.1	1.8	2.4	2.8	1.6	2.0

De aquí, se obtiene el valor r = 0.93. Se pide averiguar si existe o no una correlación lineal estadísticamente significativa entre los dos métodos mencionados con un nivel de significación de 5%.

Solución.
Hipótesis

Hipótesis nula **H_0: $\rho = 0$** (**No hay correlación lineal**)

Hipótesis alterna **H_A: $\rho \neq 0$** (**Correlación lineal estadísticamente significativa**)

Nivel de significación: Significación de 5%; $\alpha = 0.05$

Regiones
Para construir las regiones necesarias se busca la t de Student tabulada correspondiente a (n – 2) = 8 g.l. y nivel de significación 5%. Resulta que, t = 2.31. En consecuencia, las regiones son las que siguen.

Región crítica ó de rechazo	Región de aceptación de la prueba de hipótesis	Región crítica ó de rechazo
-2.31		2.31

Estadígrafo

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0.93\sqrt{\frac{10-2}{1-(0.93)^2}} = 7.16$$

Región crítica ó de rechazo	Región de aceptación de la prueba de hipótesis	Región crítica ó de rechazo
-2.31	2.31	7.16

El valor del estadígrafo se ubica en la región crítica.

Decisión estadística
La t calculada es mayor que la t tabulada ($t_{calculado} > t_{tabulado}$), luego se rechaza la hipótesis nula aceptándose la alternativa con el nivel de significación 5%, es decir, la prueba de hipótesis ofrece un resultado estadísticamente significativo con este nivel.

Conclusión
Existe una correlación lineal positiva estadísticamente significativa entre los dos métodos de medición del gasto cardíaco, con el nivel de significación 5%.

La regresión lineal
La **regresión lineal** es un conjunto de métodos para obtener y estudiar desde diversos ángulos la representación analítica lineal (**ecuación de regresión**) de la relación que puede existir entre dos ó más variables.

- ► Se llama **regresión lineal simple** cuando la relación es entre dos variables: **una variable dependiente (Y) y una variable independiente (X)**.
- ► Se llama **regresión lineal múltiple** cuando son más de dos variables: **una dependiente y más de una variable independiente**.

La ecuación de regresión permite predecir el valor de la variable dependiente (**respuesta** o **variable de criterio**) cuando se conocen los valores de las variables independientes (**predictoras, explicativas** o **regresores**).

El nombre "regresión" obedece a razones históricas, más concretamente, a la primera aplicación importante que se hizo de este método en las Ciencias Biológicas. "Lineal" se refiere a que el gráfico de la relación existente entre X y Y es una línea recta. "Simple" indica la existencia de una sola variable independiente X.

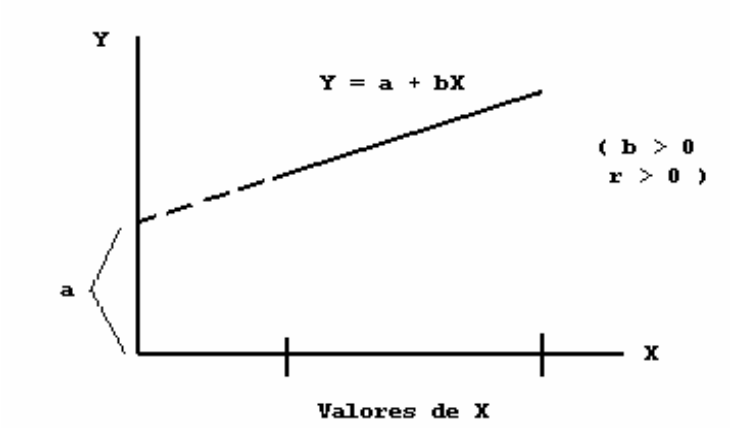
En este texto solamente nos referiremos a la regresión lineal simple. **Cuando dos variables cuantitativas continuas X y Y están correlacionadas significativamente,** es procedente calcular la **recta de regresión lineal simple** o **recta de regresión muestral:**

$$Y = a + bX$$

Con esa ecuación se pueden efectuar predicciones numéricas de valores **Y** correspondientes a valores **X no incluidos en la muestra pero, ubicados en el intervalo de valores de X de la muestra.**

Los parámetros "a" y "b" se nombran **coeficientes de regresión**.

- El parámetro **"b"** mide la **inclinación ó pendiente** de la recta. Se cumple que "b" y "r" tienen siempre el mismo signo.
- El parámetro **"a"** es el **intercepto** de la recta con el eje Y.



La recta de regresión lineal expresa la **tendencia lineal del diagrama de dispersión** y, es la que mejor se ajusta a este diagrama por ser precisamente la recta mínimo-cuadrática.

Una recta de regresión lineal **Y = a + bX** es una estimación de una **recta de regresión poblacional Y = α + βX**, válida para el universo del cual proceden las observaciones bivariadas (X_1,Y_1), (X_2,Y_2), ... , (X_n,Y_m). Precisamente, uno de los propósitos de la teoría es **lograr inferencias para α y β partiendo de "a" y "b"**.

El criterio de los mínimos cuadrados es hallar **"a"** y **"b"** de forma que se minimice la

sumatoria

$$\sum_{i=1}^{n} \left(Y_i - a - bX_i \right)^2 .$$

O sea, **"a" y "b" deben minimizar la suma** de los cuadrados de las desviaciones verticales $Y_i - \left(a + bX_i \right) = Y_i - a - bX_i$ de los Y_i respecto a los valores de la recta $Y = a + bX$ evaluada en los X_i. Observe que **no** es forzoso tener $Y_i = a + bX_i$ para todos los puntos $\left(X_i, Y_i \right)$, incluso, puede ser que esta igualdad no subsista para ninguno de estos puntos; lo único que puede garantizarse es la minimización de la suma de los cuadrados de las distancias $Y_i - \left(a + bX_i \right)$, llamadas también **residuos**.

Utilidad de la recta de regresión lineal

1. Posibilita la realización de inferencias y conocer el valor estimado de Y dado un valor de X.
2. Da en forma aproximada una relación funcional entre las variables, a partir de un criterio razonable: el de los mínimos cuadrados. Por tanto, esta recta puede ser aprovechada con fines de predicción numérica.
3. Permite resumir mediante sólo dos valores ("a" y "b") la información contenida en n pares de datos, esto es, en dos cifras, por grande que sea n. De esta manera, la recta de regresión muestral puede ser utilizada tanto para resumir datos como para estudiar la relación entre las variables.
4. Facilita precisar la tendencia más o menos oculta del diagrama de dispersión y cuantificarla.

Precauciones

1) Algebraicamente, la recta de regresión lineal simple puede calcularse para cualquier muestra bivariada, **sin embargo, ella adquiere sentido estadístico solamente cuando se ha realizado un estudio de correlación previo con una conclusión estadísticamente significativa.**
2) Las variables X y Y se miden en o mediante la misma entidad, física o conceptual, llamada la "unidad de asociación". Por ejemplo, si existe interés en estudiar la correspondencia entre estatura y peso en humanos, estas dos medidas se toman en el mismo individuo. Usualmente, no tiene sentido hablar de la relación funcional entre las estaturas de un grupo de individuos y los pesos de otro grupo arbitrario. Es posible, sin embargo, plantearse la investigación de la dependencia cuantitativa entre las estaturas de las madres y los pesos de los hijos, ya que aquí el parentesco constituye la unidad de asociación.
3) **La recta de regresión no se debe extender (extrapolar) ni emplear más allá del intervalo original o intervalo muestral de valores de X, esto es entre los valores mínino y máximo de X en la muestra. Fuera de este intervalo la relación entre las variables puede ser lineal no significativa o no lineal.**
4) La recta de regresión **no** cumple la propiedad de pasar por el mayor número posible de puntos del diagrama de dispersión. Ella no tiene tan siquiera que pasar por alguno de estos puntos. Sin embargo, ella pasa por el punto $\left(\overline{X}, \overline{Y} \right)$, integrado por las medias

aritméticas de los valores de X y Y, respectivamente. El punto $(\bar{X}, \bar{Y})$ puede pertenecer o no al susodicho diagrama.

5) Una recta de regresión puede utilizarse con fines de predicción numérica, sin embargo, **de ninguna manera ella establece una relación de causa y efecto.**

Las fórmulas para "a" y "b" se encuentran en cualquier libro de estadística y las personas interesadas se deben referir a ellos. En este solamente consideraremos como aplicar los valores de a y b.

Ejemplo. Se dispone de los datos siguientes para la concentración de una sustancia **(X)** y su densidad óptica **(Y)**.

X	80	120	160	200	240	280	320	360	400	440	480	520
Y	0.08	0.12	0.18	0.21	0.28	0.28	0.38	0.40	0.42	0.50	0.52	0.60

Se desea examinar esta información mediante correlación y regresión lineal con un nivel de significación de 5%. De ser posible, encuentre la densidad correspondiente a la concentración 220. Utilice que r = 0.9945.

Solución.

Observación preliminar. El análisis de correlación sitúa en un mismo plano a las dos variables. Por el contrario, el análisis de regresión demanda distinguir entre una variable dependiente y una independiente. La naturaleza física de las variables de este ejemplo indica claramente que, la variable dependiente es la densidad óptica de la sustancia y, la independiente es su concentración.

Análisis de correlación

Hipótesis

Hipótesis nula **H_0: $\rho = 0$** (**No hay correlación lineal**)

Hipótesis alterna **H_A : $\rho \neq 0$** (**Correlación lineal estadísticamente significativa**)

Nivel de significación: Significación de 5%; $\alpha = 0.05$

Regiones

Para definir las regiones necesarias se busca la t de Student tabulada correspondiente a n - 2 = 10 g.l. y nivel de significación 5%. Resulta que, t = 3.17. Por consiguiente, las regiones son las que siguen.

Región crítica ó de rechazo	**Región de aceptación de la prueba de hipótesis**	**Región crítica ó de rechazo**
-3.17		+3.17

Estadígrafo

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0.9945\sqrt{\frac{12-2}{1-(0.9945)^2}} = 30.03$$

Región crítica ó de rechazo	Región de aceptación de la prueba de hipótesis	Región crítica ó de rechazo
	-3.17 +3.17	30.03

El valor del estadígrafo se ubica en la región crítica.

Decisión estadística
Se rechaza la hipótesis nula con el nivel de significación 5%, es decir, la prueba de hipótesis brinda un resultado estadísticamente significativo con este nivel.

Conclusión
Existe una correlación positiva estadísticamente significativa entre la concentración de la sustancia y su densidad óptica, con el nivel de significación 5%.

Análisis de regresión

Debido a la existencia de una correlación estadísticamente significativa, es lícito emprender el análisis de regresión.

Las fórmulas para los coeficientes de regresión suministran los valores:
- a = 0.01189,
- b = 0.001142.

Luego, la recta de regresión (estadísticamente significativa) es:

$$Y = 0.01189 + 0.001142(X)$$

Debido a que el valor que se desea utilizar de X es 220 y se halla en el intervalo de valores muestrales de X: $80 \leq X \leq 520$, es admisible sustituirlo en la ecuación de la recta, con lo cual se obtiene:

$$Y = a + bX = 0.01189 + 0.001142\ (220) = 0.26$$

De esta manera, la densidad correspondiente a la concentración 220 es 0.26

INDICE

Capítulo 14
NOCIONES DE RIESGO Y SU CUANTIFICACIÓN. MÉTODO DEL VALOR PREDICTIVO.

Autor: Dr. Emilio Morales Jiménez

SUMARIO

- *Uso de la tabla de 2 x2 para el cálculo del riesgo y del método del valor predictivo.*
- *Cálculo y uso del riesgo relativo (RR) y de la razón de productos cruzados o de ventajas (Odds Ratio – OR)*
- *Cálculo y uso del Método del Valor Predictivo.*

Uso de la tabla de 2 x 2 para el cálculo del riesgo y del método del valor predictivo.

En el campo de la salud en general se requiere conocer cada día con más frecuencia los factores de riesgo que inciden en las enfermedades con vistas a su prevención, minimización o eliminación. Una vez establecido que un factor determinado tiene relación significativa con una determinada enfermedad es menester proceder a su cuantificación y precisar a quienes realmente afecta y como influye en los pacientes.

También, es cada día más utilizado el Método del valor predictivo, que aunque originalmente fue utilizado para precisar la sensibilidad y especificidad de pruebas diagnósticas, su uso se ha ido extendiendo a otras ramas de la medicina y su aplicación en trabajos científicos e investigaciones es cada día más frecuente.

Es por ello que es menester que el profesional de la salud conozca estos dos procederes para su utilización. Ambos procederes basan su cálculo en una tabla estadística de doble entrada y dicotómica, o sea que cada variable tiene solamente dos posibilidades: Positivo o Negativo; Si o No. Esa tabla se conoce por la cantidad de filas y columnas de datos que posee, o sea dos variables y cada una con dos posibilidades, luego es una tabla de dos filas por dos columnas, comúnmente llamada tabla de 2 x 2.

Un aspecto muy importante en estos dos procederes es la forma de construir la tabla. Los datos no se pueden colocar en forma indistinta en la tabla sino siguiendo un orden riguroso pre-establecido para el cálculo de ambos procederes. La tabla de 2 x 2 la podemos esquematizar de la siguiente manera:

TABLA DE 2 X 2

	Enfermo E+	No Enfermo E-	Totales
Condición +	**A**	**b**	**a + b**
Condición -	**C**	**d**	**c + d**
Totales	**a + c**	**b + d**	**N = a + b +c + d**

En las columnas se pondrá siempre cuando se está cuantificando riesgo o aplicando el método del valor predictivo la presencia o no de la enfermedad en estudio y los enfermos positivos siempre estarán en la primera columna y los no enfermos en la segunda columna.

En las filas se pondrá siempre tanto en la cuantificación del riesgo como en el método del valor predictivo, el factor de riesgo en estudio o lo que estemos analizando por el método del valor predictivo. En la primera fila serán los positivos, o sea los que presentan el factor de riesgo o tienen positivo el test en estudio o la condición que se está analizando.

Ese orden es imprescindible que se siga al llenar los datos en la tabla para poder hacer adecuadamente los cálculos estadísticos.

Cálculo y uso del riesgo relativo (RR) y de la razón de productos cruzados o de ventajas (Odds Ratio - OR)

Riesgo y causalidad

En las Ciencias Médicas no basta con determinar la frecuencia de una enfermedad, ni con determinar asociaciones entre variables, se requiere determinar relación causa-efecto con relación a las enfermedades. Es muy importante determinar condiciones o comportamientos que incrementen el riesgo de desarrollar una enfermedad en particular. Ello tiene como objetivo final descubrir relaciones causa - efecto que resulten en tratamientos efectivos y en estrategias de prevención. Esas condiciones o comportamientos que favorecen el desarrollo de un estado morboso se le conoce como **"factores de riesgo"**.

Luego, factor de riesgo para una enfermedad determinada es una condición, característica física o comportamiento que incrementa la probabilidad (riesgo) de que una persona sana desarrolle esa determinada enfermedad. Los factores de riesgo son muchos y variados. De hecho cualquier cosa puede devenir en un factor de riesgo. A continuación se listan diferentes tipos de factores de riesgo.

Tipos de factores de riesgo

- **AMBIENTALES:** aquellos que se encuentran en el medio ambiente (organismos infecciosos, agentes contaminantes, toxinas, etc.)

- **DE COMPORTAMIENTO O HABITOS DE VIDA:** tabaquismo, alcoholismo, drogadicción, sedentarismo, no cumplir con las medidas de protección en el trabajo, y muchos otros en dependencia de la enfermedad y la actividad social y laboral del sujeto.

- **SOCIALES:** divorcio, muerte de familiar cercano, pérdida de empleo y stress, entre otros.

- **GENETICOS:** sexo, raza, factores hereditarios, por ejemplo Antecedente Patológico Familiar de hipercolesterolemia incrementa el riesgo de Enfermedad Isquémica Coronaria.

La determinación de cual o cuales son las condiciones que pueden resultar en factores de riesgo para determinada enfermedad es un deseo de todos los investigadores y especialistas de cualquier especialidad médica. Ahora bien, primero que todo es necesario demostrar mediante una prueba estadística que existe una asociación o relación directa entre la condición y la presencia de la enfermedad. Para ello hay que utilizar las pruebas estadísticas que se vieron en los capítulos precedentes. Fundamentalmente la prueba de hipótesis y la prueba de Chi-cuadrado.

Una vez establecido que existe una asociación estadísticamente significativa entre la condición y la presencia de la enfermedad podemos entonces plantear que la condición es un factor de riesgo para esa enfermedad. Sin embargo, es imprescindible a continuación determinar la magnitud de ese riesgo, o sea precisar hasta que punto puede influir en el desarrollo de la enfermedad. Para ello es necesario calcular el **RIESGO RELATIVO (RR)** que compara el riesgo de desarrollar una enfermedad si se está expuesto a un factor de riesgo con el riesgo de contraerla si no se está expuesto. Ello mide la fortaleza de la asociación.

$$\text{RIESGO RELATIVO} = \frac{\text{riesgo de enfermar si expuesto al factor de riesgo}}{\text{riesgo de enfermar si no expuesto al factor de riesgo}}$$

El riesgo relativo no es posible calcularlo en todas las investigaciones ya que requiere de determinados requisitos. Entonces la fortaleza de la asociación se mide por un estimador del riesgo relativo, que es el **ODDS RATIO (OR)**, que en español se conoce como **"Prueba de productos cruzados"**. El OR compara las posibilidades (ODDS) de los expuestos y no expuestos a contraer la enfermedad.

$$\text{ODDS RATIO} = \frac{\text{posibilidad de los expuestos a contraer la enfermedad}}{\text{posibilidad de los no expuestos a contraer la enfermedad}}$$

La prueba de OR es la que más se utiliza y aparece con mucha frecuencia en las revistas y libros de Medicina. El RR se utiliza directamente en estudios experimentales y en estudios de cohorte que son los menos aunque son los más exactos. En cualquier otro estudio se utiliza el OR.

Una cuestión relevante es como interpretar el RR o el OR. Ello puede resumirse en los siguientes aspectos:

- ► A mayor el valor de RR u OR, mayor será la magnitud o fortaleza de la asociación entre el factor de riesgo y la enfermedad.

- ► Valores de RR u OR cercanos a 1 indican que la enfermedad y la exposición al riesgo no están relacionados.

- ► Valores decimales de RR u OR (< 1) indican una asociación negativa entre el factor de riesgo y la enfermedad (es decir que el factor protege de contraer la enfermedad).

Cálculo del RR

1. Hacer la tabla de 2 x 2.

	Enfermo E+	No Enfermo E-	Totales
Expuesto R+	A	B	a + b
No expuesto R-	C	d	c + d
Totales	a + c	b + d	N = a + b +c + d

2. Aplicar la formula

$$\text{RIESGO RELATIVO} = \frac{\text{riesgo de enfermar si expuesto al factor de riesgo}}{\text{riesgo de enfermar si no expuesto al factor de riesgo}}$$

$$RR = \frac{a / (a + b)}{c / (c + d)}$$

3. Interpretar el resultado.

Ejemplo:

En un estudio de cohorte se demostró que la incidencia de abortos en 100 mujeres embarazadas expuestas a un pesticida órgano-fosforado difería significativamente de la incidencia de abortos en 100 mujeres no expuestas, (X^2 dió una $p < 0.005$). Se desea conocer la magnitud del riesgo de abortar si se expone al pesticida órgano-fosforado.

	Abortos	**No abortos**	**Totales**
Expuesto al pesticida	**30 a**	**b 70**	**100 (a + b)**
No expuesto al pesticida	**10 c**	**d 90**	**100 (c + d)**
Totales	**40 (a + c)**	**160 (b + d)**	**200 (a+b+c+d)**

$$RR = \frac{a / (a + b)}{c / (c + d)} = \frac{30 / 100}{10 / 100} = 3$$

Interpretación:

El riesgo de aborto en una mujer embarazada expuesta al pesticida es tres veces superior que la no expuesta.

Calcula del ODDS RATIO (OR)

1. Hacer la tabla de 2 x 2.

	Enfermo E+	No Enfermo E-	Totales
Expuesto R+	**A**	**b**	**a + b**
No expuesto R-	**C**	**d**	**c + d**
Totales	**a + c**	**b + d**	**N = a + b +c + d**

2. Aplicar la formula

$$\text{ODDS RATIO} = \frac{\text{posibilidad de los expuestos a contraer la enfermedad}}{\text{posibilidad de los no expuestos a contraer la enfermedad}}$$

$$\mathbf{OR} = \frac{a / b}{c / d} = (ad) / (bc)$$

Note que el OR se calcula dividiendo el resultado de **(a por d)** entre el resultado de **(b por c)**. De ahí el nombre en español de prueba de productos cruzados. Como el OR es un estimador es imprescindible hacer una prueba de hipótesis para conocer si se puede o no hacer una estimación del RR.

3. Cálculo de la prueba de hipótesis para el OR.
Lo primero a ejecutar es plantear la hipótesis nula e hipótesis alternativa.

H_0: OR = 1 (No asociación)
H_A: OR $\neq$ 1 (Asociación)

Lo que se busca es determinar si el número 1 está presente en un intervalo de confianza que se hace con el valor del OR calculado. Si el 1 está incluido en el intervalo, se acepta la hipótesis nula, en detrimento de la alternativa y no hay riesgo estadísticamente significativo. Si por el contrario el 1 no está incluido en el intervalo se rechaza la hipótesis nula y se acepta la alternativa existiendo entonces un riesgo significativo.

El intervalo de confianza se calcula por el método de Woolf de la siguiente manera:

Método de Woolf para el cálculo del Intervalo de Confianza (IC) del OR

Limites del intervalo = ln OR $\pm$ (coeficiente de confiabilidad X EE del OR)

Limites del intervalo = ln OR $\pm$ K $\sqrt{}$ (1/a + 1/b + 1/c + 1/d)

Donde:
ln OR = Logaritmo natural del OR
Coeficiente de confiabilidad = K donde si α = 0.05 = 1.96 y si α = 0.01 = 2.58
EE del OR = $\sqrt{}$ (1/a + 1/b + 1/c + 1/d)

Los límites superior e inferior del IC calculados por la fórmula, se encuentran en notación logarítmica por lo que hay que hallar el antilogaritmo para conocerlos en números naturales.

4. Interpretar el resultado.

Si el número 1 está incluido en el IC se acepta H_0 y no hay riesgo significativo. Si el número 1 no está incluido en el IC se rechaza H_0, aceptándose H_A, existiendo riesgo significativo.

Ejemplo: (Utilicemos los mismos datos que en el ejemplo del RR)

En un estudio se demostró que la incidencia de abortos en 100 mujeres embarazadas expuestas a un pesticida órgano-fosforado difería significativamente de la incidencia de abortos en 100 mujeres no expuestas, (X^2 dió una $p < 0.005$). Se desea conocer la magnitud del riesgo de abortar si se expone al pesticida órgano-fosforado.

	Abortos	**No abortos**	**Totales**
Expuesto al pesticida	**30 a**	**b 70**	**100 (a + b)**
No expuesto al pesticida	**10 c**	**d 90**	**100 (c + d)**
Totales	**40 (a + c)**	**160 (b + d)**	**200 (a+b+c+d)**

Cálculo de OR:

$$OR = (ad) / (bc) = \frac{30 \times 90}{10 / 70} = 3.86$$

Prueba de hipótesis:
Ho: OR = 1 (No asociación)
H_A: OR $\neq$ 1 (Asociación)

Método de Woolf para el Intervalo de confianza:

$$\text{Limites del intervalo} = \ln(3.86) \pm 1.96 \sqrt{(1/30 + 1/70 + 1/10 + 1/90)}$$

$$= 1.35 \pm 0.78$$

Límite inferior = 0.57
Límite superior = 2.13

Aplicando el antilogaritmo tenemos que el IC para un 95% de confianza es:

Límite inferior = antilogaritmo de 0.57 = 1.8
Límite superior = antilogaritmo de 2.13 = 8.4

Se denota así: (IC95%: 1.8; 8.4 ó 1.8-8.4)

Interpretación:
Se rechaza Ho. Se puede considerar con una confiabilidad del 95% que el OR en la población de mujeres de la cual se obtuvo la muestra está entre 1.8 y 8.4. Existe una asociación estadísticamente significativa entre la exposicion al pesticida y el aborto. Las mujeres expuestas tienen 3.86 más posibilidades de abortar que las no expuestas.

También se puede expresar de la siguiente manera: Las mujeres expuestas a pesticida órgano-fosforado presentan 3.86 veces más riesgo significativo de abortar que aquellas no expuestas, **OR = 3.86; IC95%: 1.8-8.4**.

Cálculo y uso del Método del Valor Predictivo. Sensibilidad, Especificidad, Valor predictivo positivo y Valor predictivo negativo.

Como se señaló al inicio este método se está utilizando cada vez más en las investigaciones de salud. El mismo se utiliza para conocer si una nueva prueba diagnóstica es útil o no para el diagnóstico de una enfermedad y para conocer si alguna característica (síntoma o signo) ya sea sola o varias de conjunto pueden ser discriminadoras en alguna enfermedad, facilitando su diagnóstico.

Ahora bien para aplicar este método se requieren cumplir determinados requisitos. Es imprescindible disponer de una prueba llamada **"EL ESTANDAR DE ORO"** para clasificar a los pacientes en enfermos o no.

El Estándar de Oro es por lo general una prueba que tenga un alto poder de certeza en el diagnóstico. El mejor para ello es el resultado de un estudio anatomopatológico ya sea por pieza quirúrgica o por biopsia. Otros utilizados son exámenes bacteriológicos y micológicos para demostrar la presencia de determinado agente microbiano. Ahora bien, hay ocasiones en que no se pueden hacer esos estudios debido a que los procesos morbosos no lo permiten como son los casos de enfermedades psiquiátricas o neurológicas, por ejemplo. En ese caso se puede considerar que **"el criterio de expertos"** de un grupo de 3 ó 5 expertos de reconocido prestigio se convierte en el "Estándar de Oro".

El otro requisito es que se justifique su aplicación acorde al estudio o investigación que se está realizando.

Método del valor predictivo

Lo primero a realizar es la tabla de 2 x 2 con los datos de los pacientes siguiendo lo señalado al comienzo del capítulo.

	Enfermo E+	No Enfermo E-	Totales
Test T+	**a VP**	**FP b**	**a + b**
Test T-	**c FN**	**VN d**	**c + d**
Totales	**a + c**	**b + d**	**N = a + b +c + d**

Donde:
VP = Verdaderos positivos. Enfermos que presentan el test diagnóstico positivo.
FP = Falsos positivos. No enfermos que tienen el test diagnóstico positivo.
FN = Falsos negativos. Enfermos que tienen el test diagnóstico negativo.
VN = Verdaderos negativos. No enfermos con el test diagnóstico negativo.

El método calcula cuatro indicadores básicos:

SENSIBILIDAD:
Probabilidad de que un individuo enfermo tenga el test diagnóstico positivo.

$$\textbf{SENSIBILIDAD} = \frac{\textbf{ENFERMOS CON LA PRUEBA POSITIVA}}{\textbf{TOTAL DE ENFERMOS}} = \frac{a}{a + c}$$

ESPECIFICIDAD:
Probabilidad de que un individuo no enfermo tenga el test diagnóstico negativo.

$$\text{ESPECIFICIDAD} = \frac{\text{NO ENFERMOS CON LA PRUEBA NEGATIVA}}{\text{TOTAL DE NO ENFERMOS}} = \frac{d}{b + d}$$

VALOR PREDICTIVO POSITIVO (VPP):
Probabilidad de que un individuo con el test diagnóstico positivo tenga la enfermedad.

$$\text{VPP} = \frac{\text{ENFERMOS CON LA PRUEBA POSITIVA}}{\text{TOTAL CON LA PRUEBA POSITIVA}} = \frac{a}{a + b}$$

VALOR PREDICTIVO NEGATIVO (VPN):
Probabilidad de que un individuo con el test diagnóstico negativo no tenga la enfermedad.

$$\text{VPN} = \frac{\text{NO ENFERMOS CON LA PRUEBA NEGATIVA}}{\text{TOTAL CON LA PRUEBA NEGATIVA}} = \frac{D}{c + d}$$

Interpretación
Los valores que se obtienen en los indicadores son probabilidades, o sea están entre cero y uno. Mientras más cerca de 1 se halle un valor indica que es más certero y mientras más se acerque a 0 indica que es menos exacto. Al interpretar los indicadores se debe hacer de conjunto, o sea considerando los 4 indicadores fundamentales. Se trata de obtener valores máximos de sensibilidad y especificidad. A mayor sensibilidad y especificidad mayor poder discriminatorio tiene la prueba evaluada.

Ejemplo
Investigación sobre el Factor de Crecimiento de Carcinoma Mamario (FCCM) donde un estudio piloto reveló que estaba elevado en los pacientes con carcinomas de mama confirmados. Se hizo un estudio clínico que incluyó a 1600 pacientes donde por biopsia se determinó carcinoma en 600 y 1000 estaban sanos. Se consideró como positivo al FCCM un resultado mayor o igual a 150 unidades por litro.

Se obtuvo la siguiente tabla:

	Con carcinoma de mama		**Sin carcinoma de mama**		**Totales**
FCCM +	**570**	**VP**	**FP**	**150**	**720**
FCCM -	**30**	**FN**	**VN**	**850**	**880**
Totales	**600**			**1000**	**1600**

Sensibilidad = a / a + c = **VP / VP + FN** = 570 / 600 = 0.95

Especificidad = d / b + d = **VN / VN + FP** = 850 / 1000 = 0.85

VPP = a / a + b = **VP / VP + FP** = 570 / 720 = 0.79

VPN = d / c + d = **VN / VN + FN** = 850 / 880 = 0.97

La interpretación depende de cada prueba y del criterio que adopte el investigador, así como de cada situación concreta particular. En este ejemplo, los valores obtenidos son muy altos indicando un buen diagnóstico de Carcinoma de Mama mediante el FCCM.

INDICE

Capítulo 15
LA CIENCIA Y LA ACTIVIDAD CIENTÍFICA

Autor: Dr. Emilio Morales Jiménez

SUMARIO

- *El método científico.*
- *Etapas del método científico. El problema científico. Formulación de objetivos. Formulación de hipótesis. Planificación de la verificación de las hipótesis.*
- *La observación científica. Importancia. Errores.*
- *Relación del método científico con los métodos de las ciencias particulares (método clínico y método epidemiológico)*
- *La investigación científica. Tipos de investigación científica.*
- *El proyecto de investigación científica. Importancia y partes que lo integran.*
- *El informe de investigación. Importancia y partes que lo integran.*

El método científico

Antes de hablar de método científico es imprescindible ponernos de acuerdo en que cosa es ***"ciencia"***. Haciendo un breve, pero necesario recuento histórico, tenemos que el eminente fisiólogo ruso Pavlov la definió hace dos siglos como ***"una unidad dialéctica entre el sistema de conocimientos, (compuesto por conceptos, categorías y leyes); el método del conocimiento; y la unión con la práctica como punto. inicial, fin supremo y criterio del conocimiento".*** En el siglo pasado el científico Mario Bunge la precisó como: ***un estilo de pensamiento y acción, el más reciente, el más universal y más provechoso de todos los estilos"***. La Academia Española de la Lengua la definió como: ***"el reconocimiento por sus principios y causas, así como cuerpo de doctrinas metódicamente formado y ordenado que constituye una rama particular del saber humano".***

En resumen, hoy día consideramos que **CIENCIA** es ***"una forma de la conciencia social y constituye un sistema de conocimientos acerca de la naturaleza, sociedad y el pensamiento y representa el reflejo de las leyes del mundo objetivo en forma de conceptos, símbolos, sistemas, juicios y leyes"***.

Toda ciencia vincula 3 elementos: el sistema de conocimientos, el método de conocimientos y la actividad práctica inherentes a la rama específica del saber a que corresponde. También en toda ciencia existen tres características fundamentales que son la sistematicidad, la objetividad y la racionalidad.

Los aspectos que caracterizan a una ciencia en específico son las siguientes:

1- Existencia de un objeto de estudio propio.
2- Existencia de un sistema de conceptos y categorías propios.
3- Existencia de un cuerpo de conocimientos sistemáticos sobre el objeto de estudio.
4- Existencia de una necesidad social para su desarrollo.
5- Existencia del personal científico e instituciones para su desarrollo.

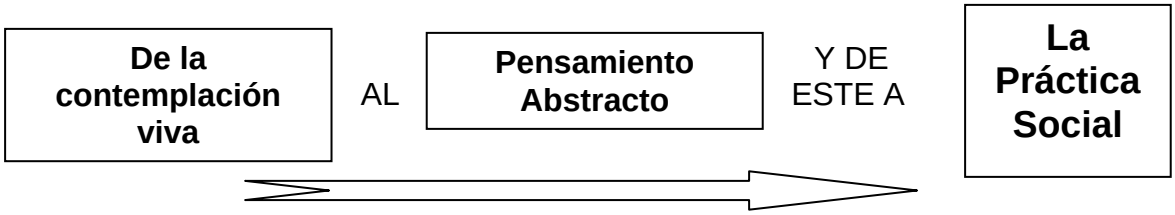
De esta forma cada ciencia en particular tiene su campo de acción e investigación muy bien definido, aunque hoy día dado el desarrollo tan acelerado de la investigación científica y la revolución científico-técnica; el desarrollo de las tecnologías de avanzada, la computación electrónica, informática, ingeniería genética y otras muchas nuevas especialidades han hecho que los campos de acción de ciencias particulares se imbriquen unos con otros facilitando un desarrollo exponencial del conocimiento científico. Testigo de ello hemos sido en los últimos 30 años.

Ahora bien, siempre en cualquier ciencia hay que diferenciar dos aspectos: el trabajo y el resultado. El trabajo se refiere a la actividad científica propiamente dicha, o sea a la investigación científica y el resultado se refiere al nuevo conocimiento científico que se deriva del mismo. Ambos aspectos se complementan siendo el último el fin del primero.

Cada ciencia en específico tiene un ***método*** de conocimiento. Este consiste en un sistema de principios y normas de razonamiento que permiten establecer conclusiones en forma objetiva. Así tenemos en medicina el método clínico y el método epidemiológico; en Estadística el método estadístico; en Pedagogía el método pedagógico y en fin, en cada ciencia un método particular para ese preciso quehacer. También en cada ciencia existen diferentes ***técnicas*** que no son más que principios y normas que auxilian a los investigadores en la aplicación del método. En Medicina podemos hablar de técnicas de enfermería, quirúrgicas y otras, en estadística se puede hablar de pruebas estadísticas y técnicas de recolección de información por citar algunas con las que estamos familiarizados.

Sin embargo, hay una ciencia general que es la filosofía que sienta las bases para estudiar de forma organizada y sistemática el mundo circundante para llegar a su conocimiento y comprensión. Esa ciencia tiene un método de conocimiento que se le ha dado en llamar ***"método científico"***. Este es la base de todos los demás métodos empleados por las ciencias específicas. La forma racional y objetiva en que plantea analizar el mundo circundante para conocer la realidad objetiva que nos rodea es utilizado por todos los métodos de las ciencias particulares. Este método no plantea reglas absolutas, ni fórmulas específicas para la obtención de conocimientos sino solamente organiza y sistematiza la marcha del proceso cognoscitivo pudiendo considerarse una brújula hacia la verdad. El método científico no suplanta los métodos particulares de las ciencias específicas sino que es la base filosófica de todos ellos. El método científico se basa en el camino dialéctico del conocimiento que va de la contemplación viva al pensamiento abstracto y de ahí a la práctica social.

CAMINO DIALÉCTICO DEL CONOCIMIENTO.

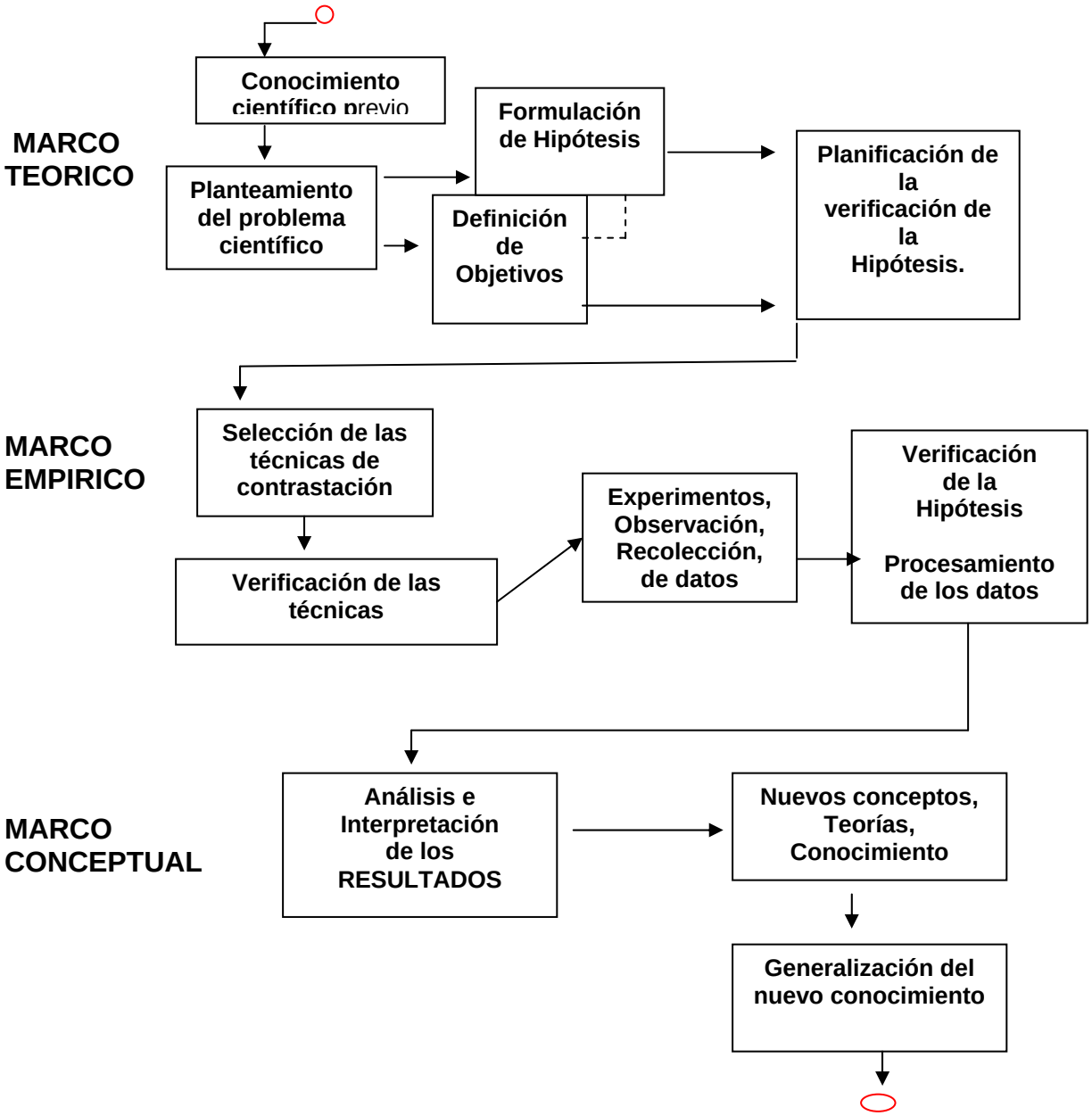


Para comprender como el método científico organiza y sistematiza la marcha del proceso cognoscitivo en las ciencias, se puede dividir didácticamente en etapas.

***Etapas del método científico. El problema científico. Formulación de objetivos. Formulación de hipótesis. Planificación de la verificación de las hipótesis*.**

Las etapas del método científico las podemos enmarcar en tres marcos o grandes fases del mismo: marco teórico, marco empírico y marco conceptual. El esquema que sigue resume los diferentes pasos que se requieren para organizar el proceso cognoscitivo de las ciencias en la obtención de nuevos conocimientos mediante la investigación científica.

ETAPAS DEL MÉTODO CIENTÍFICO



El método científico es el proceso mediante el cual se avanza en el conocimiento científico en forma de una espiral ascendente. Se parte siempre de un nivel de conocimientos determinado para mediante la investigación científica obtener nuevos conocimientos, punto de partida para la repetición del proceso.

El marco teórico agrupa los pasos del método científico que precisan, partiendo del conocimiento científico existente en un momento determinado, el problema científico que se desea investigar, los objetivos a investigar, la hipótesis científica, y la planificación de cómo se va a realizar la investigación. En esta fase todo el quehacer científico es teórico. En ella se determinan los aspectos imprescindibles y necesarios para mediante una investigación llegar a un nivel superior de conocimientos en un aspecto determinado del saber.

Todo nuevo conocimiento parte del conocimiento existente en un momento dado que es insuficiente para explicar un fenómeno o problema que se observa y para el cual no se tiene una explicación. Ello, en determinados individuos o grupos de individuos, despierta la necesidad de buscar una explicación que satisfaga el porque de ese problema, sus causas y sus consecuencias. Cualquier aspecto puede ser un problema científico si se tiene la suficiente intuición de reconocerlo como tal. Nos gusta citar un ejemplo sencillo. Cuantas personas, siglos atrás, veían como las manzanas, al igual que todas las demás frutas, cuando maduraban se desprendían de los árboles y caían al suelo? Para nadie eso representaba un problema. Sin embargo, para Isaac Newton si resultó un problema e indagando e investigando al respecto descubrió la Ley de Gravedad.

Basado en el conocimiento existente y una vez precisado cual es el fenómeno, dificultad, obstáculo o problema del cual no se tiene una explicación convincente, es necesario hacer el ***"Planteamiento del problema científico"*** o simplemente ***"Planteamiento del problema"***. Ese problema científico es la interrogante que surge en la frontera del conocimiento existente y que presupone la existencia de un marco real de conocimiento para cuya solución se genera un proceso investigativo. Para ello es necesario precisar cuan real es la falta de conocimiento sobre ese problema mediante una exhaustiva revisión bibliográfica al respecto. Hoy día, donde diariamente se publican miles de artículos científicos en el mundo entero, mediante los medios automatizados e INTERNET se realiza esa búsqueda bibliográfica en muchísimo menos tiempo que en épocas pretéritas. Ello permite precisar con relativa celeridad si lo que constituyó una dificultad u obstáculo para lo que no se tenía explicación es real o simplemente falta de conocimientos. Una vez constatada la existencia de un problema real para el cual no se tiene explicación se procede a realizar el planteamiento del problema científico.

El planteamiento del problema científico requiere determinar la esencia del problema y delimitar sus fronteras, o sea cual(es) es(son) el(los) aspecto(s) fundamental(es) que se requiere(n) conocer. Para ello se requiere creatividad, ideas nuevas y sabiduría. Es imprescindible definir la naturaleza del problema, -(a quien afecta, su importancia y sus consecuencias)-; que se desea investigar, -(que se conoce y que no se conoce del problema)-; y que se va a investigar, -(que factores o aspectos deben considerarse en la investigación del problema).

Los requisitos básicos del planteamiento del problema científico son: (1) tener objetividad y presentarlo con una buena fundamentación; (2) ser específico, o sea que sea un solo problema; y (3) que pueda ser contrastado empíricamente, esto es, que mediante la investigación científica pueda encontrarse una solución con los métodos, conocimientos y recursos disponibles. El adecuado planteamiento del problema científico, después de una exhaustiva revisión bibliográfica, permite entonces pasar a los pasos siguientes: ***"Definición de objetivos" y "Formulación de hipótesis"***.

La ***"Definición de objetivos"*** es un aspecto imprescindible en todo quehacer científico. Los objetivos van a delimitar aun más las fronteras del problema, indicar como se desarrollará la investigación, definir los resultados concretos que se desean obtener y servir de guía para elaborar el método que se requiere en la investigación. Existen diferentes clasificaciones de los objetivos, pero la más utilizada es aquella que los divide en Generales y Específicos. Por **Objetivos Generales** se entiende aquel o aquellos que precisan **"¿Qué?"** se va a investigar, lo que se espera alcanzar con el estudio en términos generales. Tienen un fin amplio y van a guiar y orientar toda la investigación. Por lo general se elabora un objetivo general aunque hay determinadas investigaciones que dado su alcance y envergadura pueden tener más de un objetivo general. Los **Objetivos Específicos** determinan el **"¿Cómo?"** se realizará la investigación. Ellos precisan y desglosan cada objetivo general. Cada uno tiene una finalidad concreta y van a determinar el método de la investigación especificando **"¿qué se hará?"** en el estudio, **"¿dónde?"** y **"¿conqué propósito?"**. Los objetivos específicos los podemos agrupar en tres categorías secuenciales para su formulación:

- **objetivos descriptivos** - caracterizan el problema y los factores contribuyentes o asociados
- **objetivos analíticos** - analizan posibles relaciones o asociaciones para esclarecer el problema
- **objetivos de utilidad** - muestran cómo utilizar los resultados

Los objetivos se expresan siempre con el verbo en infinitivo y se deben emplear verbos de acción lo suficientemente específicos para ser evaluados, tales como determinar, precisar, comparar, verificar, calcular, establecer, relacionar, entre muchos otros. No se deben emplear verbos que no sean de acción como apreciar, entender, entre otros.

La ***"Formulación de hipótesis"*** es un paso que, lamentablemente, no siempre se realiza. Acorde a la experiencia de los investigadores y a la naturaleza del estudio que se realizará, puede ser posible plantear **explicaciones o causas del problema que se van a probar**. Si eso es así, además de los objetivos se pueden plantear hipótesis. Las **HIPOTESIS** son la predicción de una relación entre uno o más factores y el problema que se está estudiando. Pueden surgir de los objetivos analíticos. Son proposiciones lógicas que, basadas en los conocimientos existentes, pretenden dar, a priori, respuesta a determinadas interrogantes de investigación que son susceptibles de contrastación, o sea de ser probada. Dicho de otra forma, es la explicación provisional al problema científico que se va a investigar acorde con los conocimientos que tiene el investigador, con el fin de constatar si es cierta o no.

Toda hipótesis científica requiere de: (1) **fundamentación teórica** basada en el conocimiento científico existente; (2) **fundamentación lógica** avalada por una formulación correcta, clara y precisa; y (3) **fundamentación empírica** que sea susceptible de contrastación. Ella debe reflejar la exhaustiva revisión bibliográfica realizada durante el planteamiento del problema científico.

Existen diferentes tipos de hipótesis. Las **hipótesis descriptivas** son aquellas que se limitan a una simple descripción de lo que se pretende responder. Las **hipótesis causales** son aquellas que tratan de explicar la causa de un fenómeno. Las **hipótesis estadísticas** son aquellas se formulan para tratar de decidir la existencia o no de diferencias o vínculos entre las variables en estudio.

Un aspecto muy valioso a tener en cuenta en todo quehacer científico es el hecho de que tan importante es al final de la investigación que la hipótesis haya sido probada, o sea ser considerada cierta, como haber sido negada, o sea haber sido rechazada. En ambos casos se obtuvo un nuevo conocimiento. En el primer caso el nuevo conocimiento coincidió con la posible explicación que se tenía del fenómeno estudiado; en el segundo indica que la explicación no era la adecuada y que se requiere buscar otra. Lo que no es admisible y va en contra de toda ética científica es forzar los resultados para que una hipótesis siempre resulte cierta.

El último paso del marco teórico del método científico es la ***"Planificación de la verificación (contrastación) de las hipótesis"***. En aquellas investigaciones que no se hayan planteado hipótesis lo que se planifica es la contrastación de los objetivos. Pero siempre es menester planificar como se va a desarrollar la investigación. La planificación de la investigación se plasma en un documento que ha recibido diferentes denominaciones. Se le llama indistintamente **protocolo o proyecto de investigación.** Más adelante en el capítulo se explicará con más detalle. En este documento hay que precisar pormenorizadamente como se realizará toda la investigación. Ello incluye desde la preparación del personal que intervendrá en la investigación, la recogida de los datos, las técnicas de contrastación que se utilizarán, el procesamiento y análisis de los datos, quienes y como realizarán las tareas durante la investigación para garantizar unos resultados válidos y confiables evitando errores.

El protocolo o proyecto de investigación se puede comparar con los planos de un arquitecto o ingeniero antes de proceder a la construcción de un edificio u obra civil. Sin esos planos es muy probable que el edificio o la obra construida no se pueda terminar o se derrumbe una vez concluida. En el quehacer científico es semejante. Hay que planificar hasta el último detalle de cómo se va a proceder y lo que se necesita para que la investigación llegue a feliz término. Sin planos no se puede hacer un edificio o un puente. Sin proyecto o protocolo no se puede hacer una investigación científica.

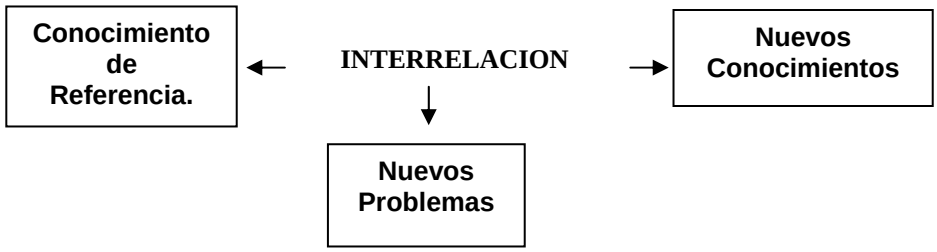
Una vez concluido el marco teórico se pasa al marco empírico. En éste la investigación es realizada. Todo lo planificado se debe cumplir al pie de la letra. Si fuera menester proceder a realizar algún cambio a lo planificado, cuestión que sucede con relativa frecuencia, por imponderables de la vida, se debe revisar el protocolo y plasmar los cambios requeridos. Ello precisará el nuevo derrotero de trabajo y realización. Los pasos

ejecutar en esta fase incluyen la selección y verificación de las técnicas de contrastación que se utilizarán, la realización de los experimentos, observación y recolección de datos en la muestra objeto de estudio y el procesamiento de los datos. Todo es actividad práctica, de ahí el nombre de marco empírico.

El método científico finaliza en el marco conceptual. Una vez con los resultados del trabajo práctico de la investigación, hay que proceder a su análisis e interpretación tomando en consideración toda la revisión bibliográfica realizada. La comparación de los resultados obtenidos con los criterios de otros autores e investigadores es un paso obligado para poder llegar a conclusiones y recomendaciones adecuadas; si se acepta o rechaza la hipótesis científica; que nuevos conceptos, leyes o conocimientos se lograron; y que nuevas interrogantes pueden haber surgido, constituyendo a su vez posibles problemas de investigación. Por último, la generalización del nuevo conocimiento alcanzado garantiza el desarrollo ulterior del quehacer científico.

Resumiendo esquemáticamente el método científico, como guía y motor del proceso cognoscitivo del quehacer científico, tenemos que:

I



Los diferentes marcos y pasos que se mostraron en el esquema del método científico no se encuentran tan nítidamente separados en el quehacer científico diario. Ello todo es un proceso dialéctico que tiene que irse ensamblando poco a poco. Los diferentes pasos dentro de cada marco se imbrican entre si y en la mayoría de los casos se van elaborando de conjunto. Por ejemplo, cuando se está planteando el problema y realizando la revisión bibliográfica inicial, se está pensando también en los objetivos y en las posibles hipótesis, así como en la forma de realizar la investigación. El esquema planteado solamente nos ayuda e comprender como es el proceso.

La observación científica. Importancia. Errores.

Como se señaló en el Capítulo 2, Recolección de la Información, ***la observación científica se define como el proceso perceptivo mediante el cual la atención es dirigida al elemento en estudio***. La observación científica hay que diferenciarla de la observación corriente que a diario realiza todo individuo. Efectivamente, en todo momento el ser humano está utilizando el sentido de la vista y **"observa"** lo que sucede en su derredor. Esa observación es, por lo general, superficial y sólo nos permite ver por donde vamos, que estamos haciendo, orientarnos, leer y otras muchas cosas. La observación científica es mucho más completa y profunda interviniendo en ella todos los sentidos. Ella comprende la vista, el olfato, el tacto, el oído y en contadas ocasiones hasta el gusto.

Toda la percepción que se puede obtener del elemento que se está estudiando forma parte de la observación científica. Se utiliza para estudiar al sujeto u objeto en la gran mayoría de las investigaciones científicas, siendo la forma más utilizada de recopilar la información necesaria para ello, no se puede soslayar.

La importancia de la observación científica radica en que es el método empírico que constituye la base de los demás métodos empíricos para la recogida de información. Ha sido el método clásico de la investigación científica siendo, a su vez, el más objetivo recomendándose su uso siempre que sea posible. Este proceder requiere de entrenamiento previo por parte de los observadores y su utilidad se restringe cuando se requiere información de manifestaciones subjetivas, hechos pasados e inquietudes futuras. Tiene el inconveniente de que es costoso no siendo muy recomendable para estudios masivos de población.

La observación científica tiene cinco factores componentes que la determinan. Ellos son:
- El observador
- El sujeto u objeto observado
- El medio de observación
- El entorno que rodea la observación
- El sistema de conocimientos

Cada uno de ellos influye en la observación y puede ser fuente de errores ***"sistemáticos"*** que de no ser detectados y eliminados son causa de ***"sesgo"*** en la recogida de información dando por consecuencia resultados y conclusiones sesgados (erróneos). Error sistemático es aquel que se produce involuntariamente durante la investigación producto de algunos de los cinco factores que intervienen en la misma y que se puede evitar. Se entiende por sesgo el recoger información (datos) que no se corresponde con la realidad, o sea, información desviada de su verdadero valor que da al traste con los resultados reales de la investigación. Detallemos las características de los cinco factores de la observación científica y como pueden resultar fuente de errores sistemáticos. Nos referiremos siempre al campo de las ciencias médicas.

El observador.
Es quien realiza la investigación. En el ámbito de la salud puede ser cualquier profesional y técnico que labore en el mismo. La observación científica debe realizarse siempre en condiciones idóneas y tratar que todas sean de la misma forma. Sin embargo, el estado físico y psíquico del observador puede influir en la misma. No de observa igual en una consulta al comienzo de la misma cuando se llega descansado, que después de haber visto a 30 pacientes o se está de post-guardia de 24 horas. Estas dos últimas situaciones son propensas a que ocurra sesgo en la recogida de información. Por ello cuando se está realizando una investigación científica hay que procurar recoger la información cuando en mejores condiciones se esté para ello y que sea igual con todos los sujetos que se estudien.

El sujeto u objeto observado.
Es el sujeto u objeto de nuestro estudio. Por lo general, en nuestro ámbito de trabajo, es el paciente salvo en un estudio con animales de experimentación. El observado también

puede ser causa de sesgo en una investigación. Por ejemplo, a un grupo de pacientes seleccionados para una investigación, se les indica que al otro día deben estar en ayunas para unos análisis de laboratorio. Si alguno no lo cumple, va a presentar resultados sesgados en sus análisis.

El medio de observación.

Por lo general la observación en salud requiere de medios que amplíen las posibilidades senso-perceptivas de nuestros sentidos. Ellos son los medios de observación que normalmente empleamos. Son de dos tipos: directos e indirectos.

- Medios directos de observación: Son aquellos que incrementan la capacidad de los sentidos del observador tales como el microscopio para incrementar la capacidad de la visión y el estetoscopio para magnificar las posibilidades de audición por mencionar dos ejemplos muy bien conocidos.
- Medios indirectos de observación: Son aquellos que transforman magnitudes fisiológicas imperceptibles para los sentidos del observador, en perceptibles. Así tenemos el electrocardiograma para precisar la conducción eléctrica en el miocardio, el esfigmomanómetro para precisar la tensión de la sangre por las arterias, el electroencefalograma para conocer la actividad eléctrica cerebral y muchos otros.

El medio de observación puede ser fuente de errores sistemáticos. Por ejemplo, si se está estudiando la tensión arterial en un grupo de pacientes y no se tiene el cuidado de utilizar el mismo esfigmomanómetro calibrado, las mediciones resultarán diferentes en dependencia del que se utilice. Ello introduce un sesgo en los valores de la tensión arterial en los pacientes estudiados. Si se requiere el peso del paciente sucede lo mismo si no se utiliza una pesa previamente calibrada en todos ellos.

El entorno que rodea la observación.

El medio ambiente circundante a toda observación científica es muy importante para lograr uniformidad y sistematicidad en la recogida de información. Pongamos dos ejemplos muy frecuentes y fácilmente entendibles. Si en una consulta médica, donde el facultativo está examinando a un paciente, no se logra la privacidad y aislamiento necesarios, la observación que se realiza está propensa a sesgo. No es lo mismo examinar al paciente acompañado solamente por el personal de enfermería que lo asiste dedicando toda su atención a la misma, que hacerlo con interrupciones frecuentes por parte de otros que irrumpen en el local e interfieren el proceso de observación. Si durante la realización de un electrocardiograma, que se debe realizar con el paciente en reposo y tranquilo en un local aislado, está entrando constantemente personal a preguntar cualquier cosa, el trazado puede resultar sesgado por alteraciones del paciente. Por ello hay que lograr que el entorno que rodea la observación en toda investigación científica sea el ideal y semejante en todos los sujetos u objetos obervados.

El sistema de conocimientos.

La observación depende también de los conocimientos que se tengan sobre lo que se está observando. No puede hacer una observación semejante un médico recién graduado que un especialista de II grado con 30 años de experiencia en su especialidad. Por ello cuando hay varios observadores en una investigación es imprescindible ponerse todos de acuerdo en como van a realizar la observación y como van a interpretar los hallazgos de

la misma. Si no se logra una uniformidad de criterios comprobada por una supervisión aleatoria y constante habrá alta inclinación a sesgo en la información recogida.

El sesgo durante la observación científica se puede evitar con una planificación cuidadosa del proceso de recolección de datos, la ejecución de una prueba previa de las herramientas que se utilizarán y una supervisión sistemática del proceso de recolección de datos

Es importante siempre tener en cuenta las consideraciones éticas durante la observación científica. Para ello se debe prestar atención a no violar la privacidad mediante preguntas muy sensibles, no hacer la observación sobre un sujeto sin su consentimiento y respetar determinados valores culturales, tradiciones o tabúes del sujeto. Las principales recomendaciones para cumplir con la ética durante la observación científica son no explorar aspectos sensibles con el sujeto antes de haber establecido una adecuada relación con él y asegurar la confidencialidad de los datos obtenidos.

La observación científica tiene como ventaja sobre otras técnicas de recolección de información el hecho de que el observador y el sujeto u objeto observado se encuentran en contacto directo permitiendo repetir la misma si el resultado es dudoso o se sospeche que puede tener sesgo.

Relación del método científico con los métodos de las ciencias particulares (método clínico y método epidemiológico).

Como vimos anteriormente el método científico es el instrumento del conocimiento de donde se nutren las ciencias particulares. En la salud se aplican dos métodos inherentes a las ciencias médicas, el Método Clínico y el Método Epidemiológico.

El ***Método Clínico*** es la aplicación del método científico a la clínica, su objeto de estudio es la recuperación de la salud del hombre enfermo. .

El ***Método Epidemiológico*** se utiliza para abordar científicamente las causas de las enfermedades y desviaciones de la salud que afectan a la población, está enfocado hacia los problemas de salud de grupos humanos.

El ***Método Estadístico*** explicado en el capítulo 1, Introducción a la Estadística, es muy utilizado también en las Ciencias Médicas, tanto en el quehacer cotidiano de profesionales y técnicos como en el quehacer científico.

A continuación exponemos en una tabla como el método científico se refleja tanto en el método clínico como en el epidemiológico y en una tabla siguiente como se refleja el método estadístico en los dos métodos particulares de las ciencias médicas.

REFLEJO DEL METODO CIENTIFICO CON EL METODO CLINICO Y EL METODO EPIDEMIOLOGICO

METODO CIENTIFICO	METODO CLINICO	METODO EPIDEMIOLOGICO
Marco Teórico Planteamiento del Problema Formulación de hipótesis	**Planteamiento del problema:** (Llega paciente y plantea su problema de salud) **Búsqueda de Información:** (Anamnesis y examen físico) **Formulación de hipótesis:** (Diagnóstico presuntivo)	***Observación*** (Signos de colectividad enferma) (Historia de la interacción enfermedad - población en área afectada) (Inspección del área en general y algunos servicios en particular) **Formulación de hipótesis:** (Hipótesis Epidemiológica)
Marco Empírico Verificación de hipótesis (recogida y procesamiento de datos)	**Verificación de la hipótesis:** (Exámenes complementarios) (Tratamiento provisional) (Evolución del paciente)	***Verificación*** (Medidas preventivas; exámenes de higiene y epidemiología; investigaciones en grupo control)
Marco Conceptual (Conclusiones, nuevas leyes, teorías, modelos, conocimientos)	**Diagnóstico de certeza:** (Diagnóstico definitivo) (Tratamiento definitivo)	***Conclusión*** (Diagnóstico epidemiológico definitivo) (Medidas de control y prevención)

REFLEJO DEL METODO ESTADISTICO CON EL METODO CLINICO Y EL METODO EPIDEMIOLOGICO

METODO ESTADISTICO	METODO CLINICO	METODO EPIDEMIOLOGICO
Planificación de la investigación	**Planteamiento del problema:** (Llega paciente y plantea su problema de salud) **Búsqueda de Información:** (Anamnesis y examen físico) **Formulación de hipótesis:** (Diagnóstico presuntivo)	***Observación*** (Signos de colectividad enferma) (Historia de la interacción enfermedad - población en área afectada) (Inspección del área en general y algunos servicios en particular) **Formulación de hipótesis:** (Hipótesis Epidemiológica)
Recolección de Información **Procesamiento de datos**	**Verificación de la hipótesis:** (Exámenes complementarios) (Tratamiento provisional) (Evolución del paciente)	***Verificación*** (Medidas preventivas; exámenes de higiene y epidemiología; investigaciones en grupo control)
Análisis e Interpretación	**Diagnóstico de certeza:** (Diagnóstico definitivo) (Tratamiento definitivo)	***Conclusión*** (Diagnóstico epidemiológico definitivo) (Medidas de control y prevención)

La investigación científica. Tipos de investigación científica.

En los primeros párrafos del presente capítulo planteamos que en toda ciencia había que diferenciar dos aspectos: el trabajo científico que se refiere a la investigación científica y el resultado que se refiere al nuevo conocimiento científico que se deriva de ella. Luego el fin de toda investigación científica, basada en el método científico, es alcanzar un peldaño superior del conocimiento.

La investigación científica, al igual que cualquier otra actividad humana, necesita ser organizada. Al analizar el método científico vimos que debe planificarse todo lo relacionado con la futura investigación (marco teórico) antes de realizarla. Una vez que se tiene el proyecto de investigación (protocolo) se procede a realizar la investigación (marco empírico) y por último a analizar los resultados obtenidos (marco conceptual) para llegar a nuevos conocimientos.

Aunque la investigación científica es un término genérico existen diferentes clasificaciones de la misma. Así cada clasificación define diferentes tipos de investigaciones. Veamos algunas clasificaciones.

- **El Ministerio de Ciencias, Tecnología y Medio Ambiente (CITMA) de nuestro país propugna la clasificación de las investigaciones según el alcance de los resultados.** Así se clasifican en:.
  - **Fundamental**: Investigación cuyo propósito está dirigido a la búsqueda de un nuevo conocimiento, generalmente teórico en las ciencias básicas, pero no puede precisarse su relación con la práctica social.

  - **Fundamental Orientada**: Investigación cuyo propósito está dirigido a la búsqueda de un nuevo conocimiento teórico pero vinculado con la práctica social de forma indirecta y mediata. Sus resultados no tienen una aplicación inmediata, pero pueden conducir a resultados que sí la tienen.

  - **Aplicada.** Investigación cuyo propósito es la solución de un problema que surge directamente de la práctica social, y genera resultados que son aplicables de manera inmediata.

  - **De desarrollo**. Investigación cuyo propósito es completar, desarrollar y perfeccionar nuevos materiales, productos o procedimientos. Se incluyen aquí los estudios realizados para evaluar el resultado de la implantación de nuevos procedimientos o técnicas, la evaluación de nuevos métodos de diagnóstico, evaluación de nuevos procederes terapéuticos y evaluación de tecnologías sanitarias entre otras.

- **Clasificación de las investigaciones según sus objetivos.** Se clasifican en:
  - **Exploratoria:** Investigación que se realiza para obtener información sobre un problema del cual no se tiene la suficiente información para realizar una investigación de mayor envergadura. Como su nombre lo indica su objetivo

fundamental es explorar una situación dada y poder delimitarlo con vistas a poder plantear una investigación más precisa después.

- **Descriptiva:** Investigación que se realiza fundamentalmente para describir un problema, conocer sus características, determinar a quienes afecta, cuales son sus principales manifestaciones, en fin precisar aspectos que enriquezcan los conocimientos sobre el problema investigado en un entorno específico. Esta investigación puede ser realizada hacia atrás en el tiempo **(retrospectiva)** o hacia adelante **(prospectiva).** Esta última puede ser **longitudinal** cuando los pacientes se siguen en el tiempo por un periodo más o menos prolongado y **transversal** cuando la información recogida viene a ser como una foto de un momento o proceso determinado.

- **Explicativa (Analítica):** Investigación cuyo objetivo fundamental es dar una explicación del problema que se está investigando, bien mediante el establecimiento de relación entre variables como mediante la determinación de relación causa-efecto. Esta investigación puede sub-clasificarse en **experimental** donde se efectúa un experimento donde el investigador manipula una variable determinada manteniendo constantes las demás y en **no-experimental** donde no se manipula una variable en especial precisándose posibles relaciones entre las variables estudiadas y en consecuencia con esos resultados precisar posibles relaciones causales.

.

– **La Organización Mundial de la Salud (OMS) propugna en su libro "DISEÑO Y CONDUCCION DE PROYECTOS DE INVESTIGACION EN EL SISTEMA DE SALUD" de Varkevisser CM, Pathmanathan I y Brownlee la siguiente clasificación:**

  ❖ **Estudios de no intervención:** Investigación donde el investigador solamente describe y analiza objetos o situaciones pero no interviene. Los mismos pueden ser:

    ➢ **Estudio exploratorio:** Estudio a pequeña escala de relativamente corta duración que se realiza cuando se conoce poco sobre la situación o el problema que se desea investigar. Al hacer un estudio exploratorio se DESCRIBEN las características sobre la situación o el problema que se desea investigar obteniéndose elementos sobre por ejemplo, diferentes categorías de pacientes y posibilidades para estudiar distintos factores. En ocasiones se desea ir mas allá y tratar de explicar diferencias que se observan o identificar posibles causas, requiriéndose entonces comparar grupos. La comparación es una estrategia fundamental en la investigación para identificar variables que nos ayuden a explicar porque un grupo de personas difiere de otro. Los estudios exploratorios ganan en valor explicativo si enfocamos el problema desde diferentes ángulos al mismo tiempo, o sea obteniendo información sobre el problema a partir de fuentes de informaciones diferentes e independientes para luego cruzarla. Por ejemplo, un estudio exploratorio sobre trauma ocular en el escolar, se puede enfocar a la vez mediante la entrevista al paciente, a la madre, a la maestra, al medico y enfermera de familia, al personal de salud del

policlínico, y al personal de salud del hospital especializado. **Si el problema y sus factores contribuyentes no están bien definidos es siempre aconsejable hacer un estudio exploratorio antes de involucrarse en un estudio descriptivo o analítico a gran escala.**

- **Estudios descriptivos:** Los estudios descriptivos se basan en la recolección sistemática de datos y su presentación de manera de brindar una idea clara de una situación particular. Pueden ser de dos tipos:
  - **Estudios descriptivos de casos:** Estos estudios describen a profundidad las características de uno o más "casos". Un caso puede ser, por ejemplo, un paciente, una institución de salud, una comunidad, un municipio, entre otras muchas. Pueden brindar información muy útil y son muy usados en las ciencias sociales, ciencias de la administración y en medicina clínica. Por ejemplo, en medicina, las características de una enfermedad no reconocida pueden ser documentadas como un estudio de caso, siendo frecuentemente el primer paso para desarrollar el cuadro clínico de esa enfermedad.
  - **Encuesta transversal (cross-sectional):** Está dirigida a cuantificar la distribución de determinadas variables en una población en estudio durante un mismo momento de tiempo. Por ejemplo,
    - Características físicas de personas, materiales, o el medio ambiente en: encuestas de prevalencia (cualquier enfermedad) y evaluación de cobertura (inmunización, agua potable, servicios de salud u otros).
    - Características socioeconómicas de personas, tales como edad, educación, estado civil, número de hijos, y otras.
    - El comportamiento de personas y el conocimiento, actitudes, creencias y opiniones que puedan ayudar a explicar el comportamiento ante determinados eventos que ocurren en la población (estudios CAP)..

  Las encuestas transversales pueden ser grandes, con cientos o miles de unidades de estudio (elementos muestrales). En estos casos se limita el número de variables a estudiar para evitar problemas en el análisis y en la redacción del informe final. Si las encuestas transversales son pequeñas se pueden hacer más complejas. En estos casos pueden incluso revelar asociaciones entre variables. En estos casos se combina la DESCRIPCION de la población objeto de estudio con la COMPARACION entre grupos de la población. Esas combinaciones son frecuentes y la distinción entre estudios descriptivos y analíticos es en ocasiones algo abigarrada.

- **Estudios analíticos (comparativos):**.Los estudios analíticos tratan de establecer causas o factores de riesgo para determinados problemas. Se realiza mediante la comparación de dos o más grupos donde algún grupo tiene o ha desarrollado el problema y otros no. En cualquier estudio analítico hay que tener mucho cuidado con variables de confusión. Los estudios analíticos más comunes son:

  - **Estudio de cohorte:** En este tipo de estudio un grupo de individuos expuesto a un factor de riesgo, (grupo de estudio), es comparado con otro

grupo de individuos no expuestos al factor de riesgo (grupo control). **Se siguen ambos grupos prospectivamente en el tiempo y se compara la ocurrencia del problema que se espera esté relacionado con el factor de riesgo en ambos grupos.** El grupo control y el grupo de estudio deben ser seleccionados al mismo tiempo y seguidos en el tiempo con la misma intensidad. Ambos grupos deben ser lo más similar posible excepto en el factor de riesgo objeto de estudio.

- **Estudio de cohorte histórica (retrospectivo):** Este tipo de estudio es muy semejante al de cohorte siendo la única diferencia que **se parte del pasado. Los grupos de individuos expuestos o no al riesgo se buscan en el pasado y en el momento presente se determina la ocurrencia del problema que se espera relacionado con el factor de riesgo y se compara**.

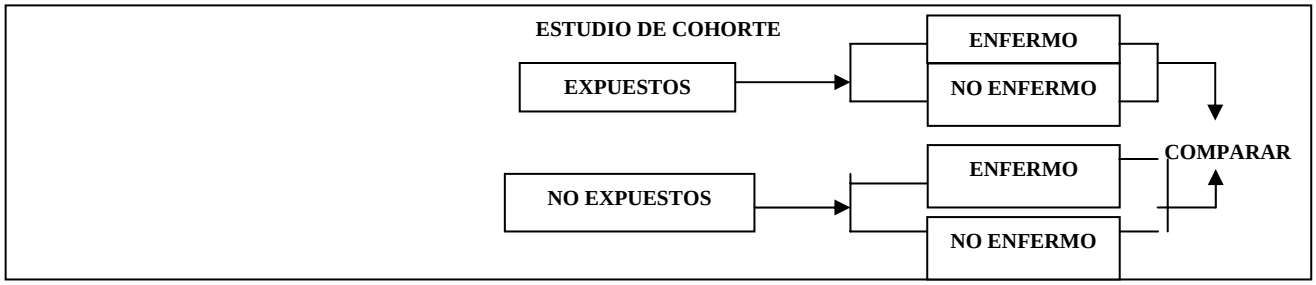
- **Estudio de casos y controles:** Se compara un grupo [casos], donde está evidente el problema (por ejemplo, desnutrición) con otro, que llamaremos grupo de control o de comparación, donde está ausente el problema. **Se averigua retrospectivamente los factores que han contribuido al problema. En ambos grupos, a saber casos y controles, se determina la presencia o no de los factores de riesgo que se están estudiando y luego se hacen las comparaciones.** Se puede minimizar la influencia de variables de confusión mediante la composición de los grupos, haciendo que ambos grupos sean similares en relación a la distribución por una o más variables de confusión. Idealmente, se quisiera equiparar el grupo de casos y el grupo control por todas las variables excepto aquellas que estamos estudiando como posibles factores de riesgo o causas del problema pero esto, en la práctica, es imposible y además no es aconsejable. (Se pueden eliminar así variables de interés.) Estos estudios utilizan la estratificación de las muestras además de la equiparación de grupos para controlar las variables de confusión.

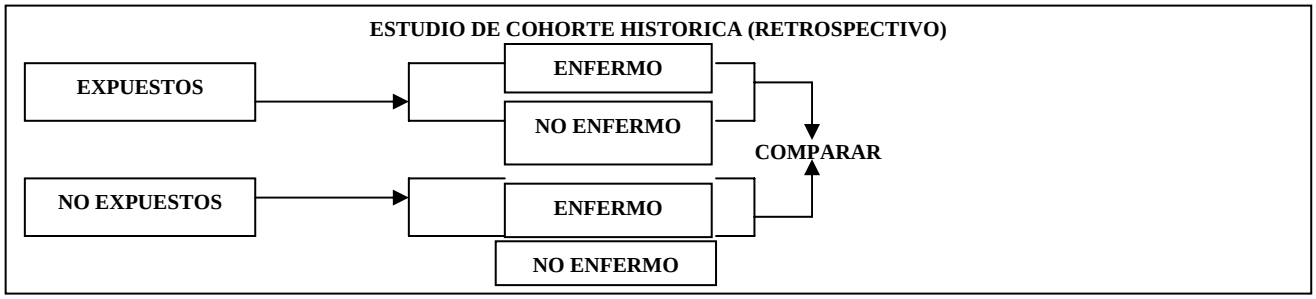
- **Estudio comparativo transversal:** Se hace comparaciones entre grupos así como se describen las características de cada grupo. **En un mismo momento de tiempo se tratan de determinar causas y riesgos y a la vez se clasifican los elementos muestrales en los grupos de estudio** estudiándose en ellos la presencia o no de las posibles causas o factores de riesgo.

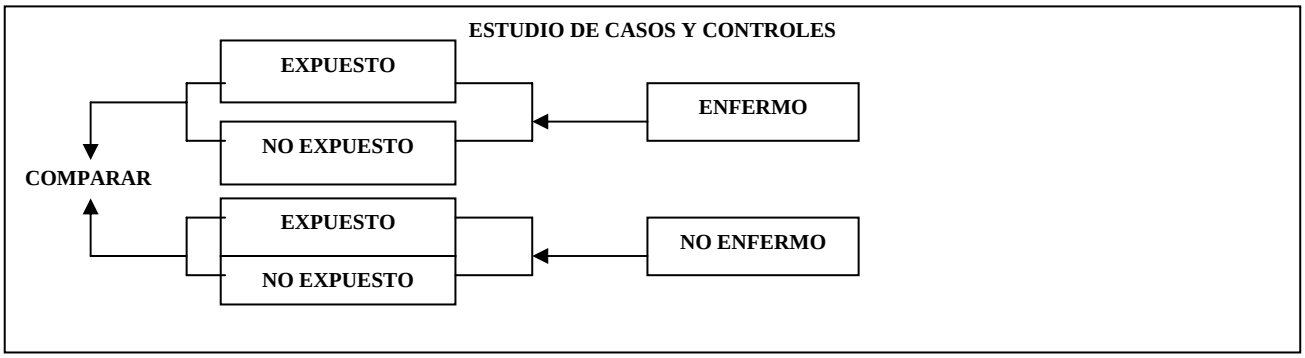
En la página siguiente se esquematizan estos cuatro tipos de estudios tan utilizados en las investigaciones en salud.

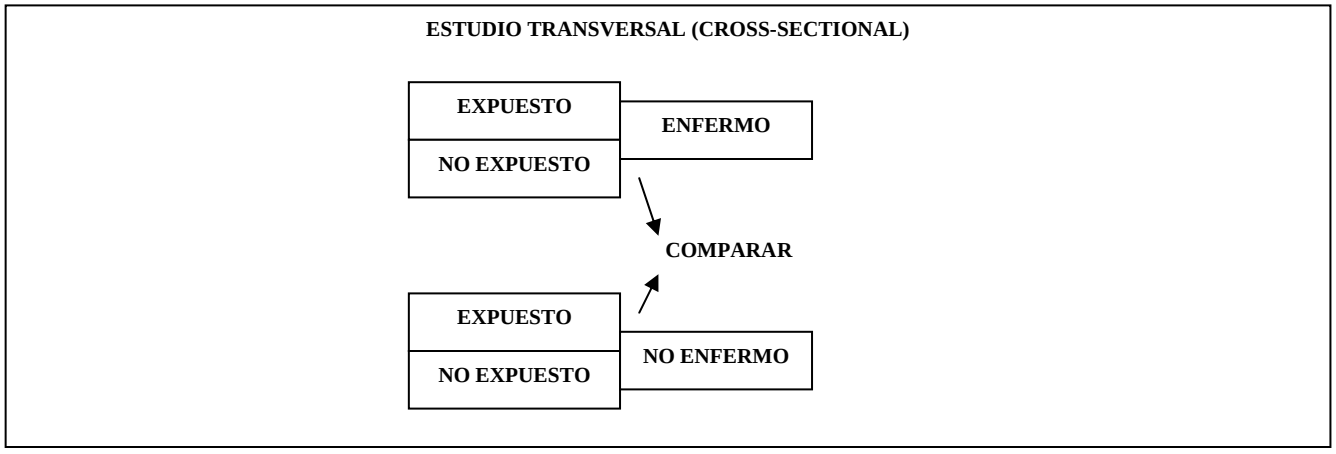
TIPOS DE ESTUDIO

PASADO **PRESENTE** **FUTURO**









- **Estudios de intervención:** Investigación donde el investigador manipula objetos o situaciones y mide los resultados de sus manipulaciones (por ejemplo, implementando campañas masivas de educación para la salud y midiendo el mejoramiento de las tasas de inmunización). Usualmente se comparan dos grupos:
  - uno donde se hace la intervención (por ejemplo, tratamiento con medicamento)
  - otro donde no se interviene (se le administra un placebo).

  Existen dos tipos de estos estudios: estudio experimental y estudio cuasi-experimental.
  - **Estudio experimental:** El diseño experimental es el único tipo de estudio que puede probar fielmente la causalidad. En el estudio experimental los individuos son aleatoriamente ubicados en al menos dos grupos: un grupo se somete a la intervención y el otro no. El efecto de la intervención se obtiene comparando los grupos. Este estudio tiene tres características:
    - MANIPULACION - se le hace algo a un grupo de sujetos
    - CONTROL - el investigador introduce uno o más grupos de control para comparar con el grupo experimental
    - ALEATORIEDAD - se tiene cuidado en la asignación aleatoria de los sujetos a uno u otro grupo.

    La fuerza de los estudios experimentales radica en que por la aleatoriedad se elimina el efecto de variables de confusión.

DIAGRAMA DEL ESTUDIO EXPERIMENTAL

Población de estudio
(muestreo)
⇓
Muestra de la población
(aleatoriedad)
⇓

⇓	⇓
Grupo de estudio (experimental)	Grupo control (grupo de comparación)
⇓	⇓
Primera recolección de datos (antes de la intervención)	Primera recolección de datos (al mismo tiempo que el grupo de estudio)
⇓	⇓
Intervención	No intervención
⇓	⇓
Ultima recolección de datos (después de la intervención)	Ultima recolección de datos (al mismo tiempo que el grupo de estudio)
⇓	⇓

⇓---------- comparación ----------⇓

- **Estudio cuasi-experimental:** En este tipo de estudio está ausente al menos una de las características del estudio experimental, bien la aleatoriedad o el uso de grupo de control. Siempre incluye la manipulación de una variable independiente que sirve de intervención. Se puede hacer con uno o dos grupos de estudio.

DIAGRAMA DEL ESTUDIO CUASI-EXPERIMENTAL CON DOS GRUPOS

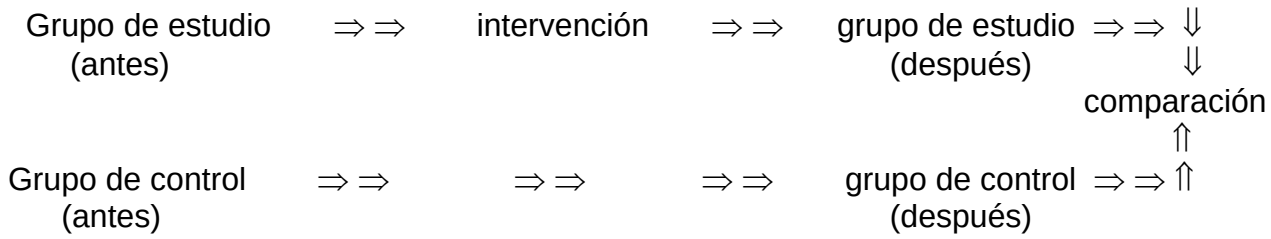
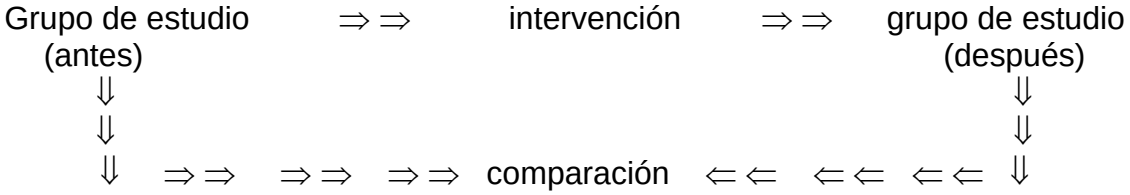


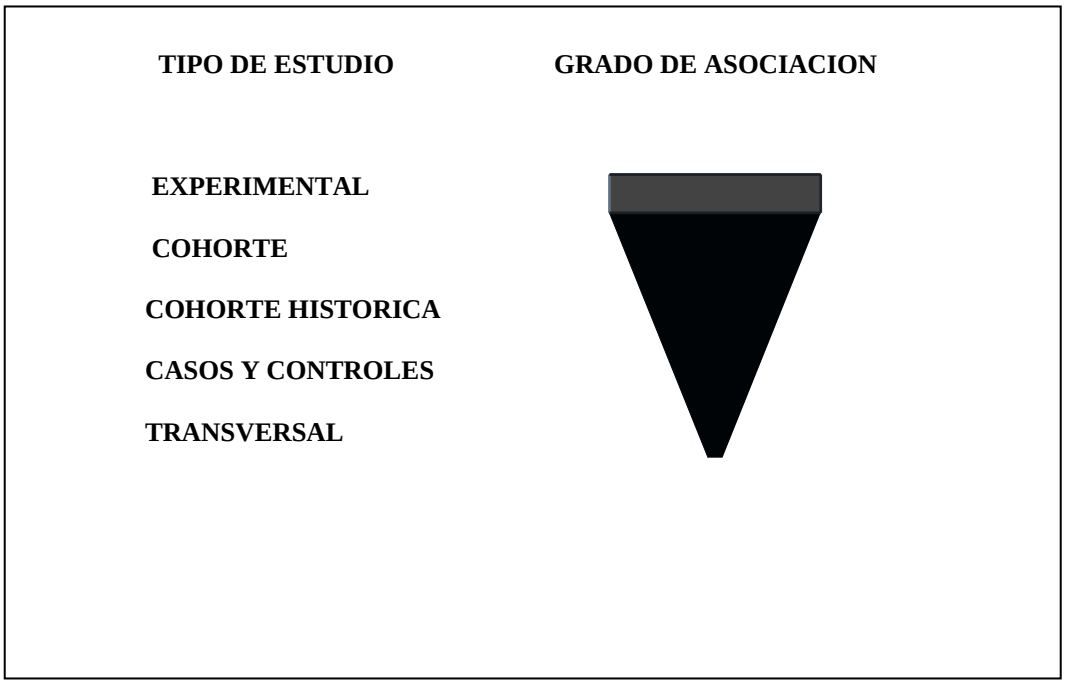
DIAGRAMA DEL ESTUDIO CUASI-EXPERIMENTAL CON UN GRUPO
(Estudio antes – después)



Tanto los estudios analíticos como los experimentales se emplean para determinar relaciones causales. La fortaleza relativa de esos estudios para demostrar causalidad se muestra en el siguiente esquema.

ESTABLECIENDO CAUSALIDAD

FORTALEZA RELATIVA DE LOS DISTINTOS ESTUDIOS



El triángulo invertido representa la fortaleza de la asociación causal que se puede demostrar según el tipo de estudio empleado. El estudio que más fortaleza brinda es el experimental seguido por el de cohorte, cohorte histórica, casos y controles y en último lugar el transversal, que paradójicamente es el que con mayor frecuencia se realiza.

La clasificación propugnada por la OMS es muy completa y satisface plenamente cualquier investigación que sea necesario realizar en el ámbito de la salud. Ahora bien cualquiera sea la clasificación que se utilice, en toda investigación científica se requieren dos documentos básicos: ***el proyecto de investigación o protocolo y el informe final de la investigación***. Ambos son imprescindibles y sus aspectos principales se reseñan en los próximos acápites.

El proyecto de investigación científica. Importancia y partes que lo integran.

En nuestro sistema de salud la investigación científica toma cada día más relevancia en el quehacer de profesores, especialistas, técnicos y personal en formación en las Ciencias Médicas. Ya hemos visto que el primer paso para ejecutar una investigación científica es precisar el problema a investigar, definir los objetivos de la investigación, establecer las hipótesis si procede y planificar la realización de la investigación que no es más que elaborar el ***proyecto de investigación*** **o protocolo**. Es de todos conocidos que la

investigación científica conlleva un gasto en recursos humanos y materiales que necesariamente hay que considerar durante la elaboración de un proyecto de investigación sobre todo si el mismo debe ser financiado por alguna institución nacional, internacional u organismos no gubernamentales. El proyecto de investigación (protocolo) debe ser por tanto un documento que permita no tan solo llevar a vías de hecho una investigación sino también brindar todos los elementos necesarios para que pueda ser aprobado y financiado para tener un feliz término.

Todavía en muchos profesionales y técnicos de la salud prevalece el criterio formalista de que el ***"proyecto de investigación" (protocolo)*** tiene como utilidad básica el inscribir una investigación en el Plan de Investigaciones de la Institución de Salud en que trabaja o que es la forma de solucionar fondos para acometer una investigación. No lo consideran como la herramienta de trabajo básica e indispensable para llevar a vías de hecho, sobre bases verdaderamente científicas, una investigación. En nuestro quehacer diario, nacional e internacionalmente, conocemos de muchos proyectos de investigación que están incompletos o que simplemente ***"se engavetan"*** una vez aprobados en lugar de utilizarse para realizar adecuadamente la investigación.

Importancia.

Las consideraciones más relevantes sobre la importancia del proyecto de investigación se listan a continuación.

- Lograr que la investigación sea aprobada inicialmente por la institución correspondiente.
- Servir como instrumento de guía y control al investigador y optimizar la utilización de los recursos y tiempo disponible.
- Permitir planificar detalladamente todo lo que se va a ejecutar durante la investigación definiendo claramente que hará quien y cuando.
- Permitir la reproducción de la investigación por otros autores.
- Servir como instrumento de control para las autoridades de salud competentes u organismos financieros en la supervisión de la investigación.

Partes que lo integran.

Existen disímiles formatos para presentar un proyecto de investigación. De hecho no existe un formato único y cada institución, organización internacional, organizaciones No gubernamentales, en fin cada cual puede tener un formato diferente. En nuestro país existe el formato CITMA, pero también existen formatos particulares de otras instituciones como el de la OMS, PNUD entre otras, si el proyecto hubiera que presentarlo a ellas. No obstante la diferencia del formato de presentación, el contenido del proyecto de investigación es semejante en todos pues refleja el marco teórico del método científico. Veremos a continuación cuales son los aspectos que, independientemente del formato de presentación, se plasman en cualquier proyecto de investigación. Ellos son:

- Título
- Introducción (Planteamiento del problema)
- Objetivos (Hipótesis)
- Método
- Cronograma (Plan de trabajo)
- Recursos necesarios (Presupuesto)

- Bibliografía
- Anexos
  - Cuestionario (Modelos de recogida de datos)
  - Plan de tabulación
  - Otros

Título

El título de la investigación científica debe dar una idea clara del problema científico a estudiar, su contenido y expresar realmente el trabajo que se realizará. Debe responder a las interrogantes de ¿Qué? ¿Dónde? y ¿Cuándo?.

No existe una norma internacional en cuanto a la longitud del título de una investigación, o sea cuantas palabras deben utilizarse para expresar el mismo. El autor considera que ello resultaría imposible ya que la disimilitud en contenido de las investigaciones así lo determina. Hay investigaciones que el título se puede expresar en pocas palabras y otras en que se requieren muchas más. Todo depende del problema científico y en la forma que se vaya a abordar. Lo fundamental del título es que satisfaga lo planteado en el párrafo anterior de manera clara, concisa y completa.

Introducción (Planteamiento del problema)

El planteamiento del problema científico es la base para el desarrollo ulterior de la propuesta de investigación, o sea, los objetivos, método, plan de trabajo y presupuesto. Facilita el encontrar información y reportes sobre estudios similares de los cuales se puede beneficiar el diseño del estudio, permite señalar porque debe acometerse la investigación propuesta, y que se espera alcanzar con los resultados. Esto es importante resaltarlo cuando se presente el proyecto al personal de la salud, al ministerio correspondiente, a dirigentes y a agencias donantes los cuales financiarán o aprobarán el estudio.

La información que debe ser incluida en el planteamiento del problema es la siguiente:

1. Una breve descripción de las características culturales y socio-económicas y una panorámica del sistema de salud y del estado de salud en el país o área siempre y cuando éstas sean relevantes al problema. Si están disponibles, incluya algunas estadísticas ilustrativas que permitan describir el contexto en que ocurre el problema.
2. Una descripción concisa de la naturaleza del problema (la discrepancia entre lo que es y lo que debiera ser) y de su magnitud, distribución y severidad (quién es afectado, dónde, desde cuándo y cuáles son las consecuencias para aquellos afectados y para los servicios de salud).
3. Un análisis de los principales factores que pueden influenciar en el problema y un argumento convincente de que la información disponible es insuficiente para solucionarlo.
4. Una breve descripción de cualquier solución que se haya empleado en el pasado, que tal ha funcionado la misma y porqué se requiere otra investigación.
5. Una descripción del tipo de información que se espera obtener como resultado del proyecto y cómo esta información será utilizada para resolver el problema.
6. Si es necesario, puede incluir una lista corta de definiciones de los conceptos fundamentales usadas en el planteamiento del problema.

Para desarrollar este punto se requiere una búsqueda bibliográfica actualizada que no requiere sobrepasar de 20 citas. Las mismas deben ser en su mayoría de los últimos cinco años. Las citas se deben acotar en el texto y listarse en el acápite de bibliografía según orden de aparición

En resumen, este aspecto va a introducir al que tiene que aprobar el proyecto en cual es el problema científico que se desea abordar, su importancia, la necesidad de su investigación, los principales resultados que se esperan lograr y su utilidad en la práctica social. En dependencia del formato que se utilice será la extensión de este aspecto. Hay formatos que lo restringen a dos páginas como máximo y otros que los dividen en varias partes siendo mucho más extenso.

Objetivos. (Hipótesis)

A comienzo de este capítulo se habló exhaustivamente de los objetivos y también de la hipótesis científica. Los objetivos resumen lo que se quiere lograr con el estudio y están estrechamente relacionados con el planteamiento del problema y con el título de la investigación, requiriéndose armonía y coherencia entre estos tres aspectos.

Se acostumbra a plantear objetivos generales y específicos. El objetivo general establece lo que se espera alcanzar con el estudio y casi siempre es único, aunque pueden plantearse más en dependencia de la investigación. Los objetivos específicos abordan los diferentes aspectos del problema y los factores que lo influyen desglosando el objetivo general en partes más pequeñas pero lógicamente conectadas que indican los diferentes pasos del trabajo investigativo que se desea realizar.

La cantidad de objetivos específicos que se establecen en un proyecto de investigación está en correspondencia con la magnitud de la misma. No hay regla fija, pero no deben ser muchos. Se tratará siempre de que no sobrepasen de 6 a 8 objetivos. Se deben colocar en una secuencia lógica (objetivos descriptivos, analíticos y de utilidad), siempre con un verbo de acción en infinitivo lo suficientemente específico para ser evaluado.

La hipótesis científica, de requerirse en el proyecto de investigación, se puede plantear al final de la Introducción (Planteamiento del problema) o junto a los Objetivos, sino especifica otra cosa el formato que se utilice.

Método.

También llamado Material y Métodos, Metodología o Metódica, es la parte más extensa de cualquier proyecto de investigación. En este aspecto hay que precisar como se va a hacer la investigación. Mientras más se detalle todo el trabajo que se requiere más fácil será la ejecución de la investigación. Aquí se deben pormenorizar las cuestiones siguientes:

- **Tipo de estudio:** Se precisa que tipo de estudio se realizará especificando donde, cuando y con quien.
- **Método de investigación:** Se especifica como se va a realizar la investigación. Si es un grupo de estudio o más grupos. Como se estudiarán los sujetos de la investigación. Los procederes diagnósticos que se seguirán. En resumen, todos los aspectos de la investigación en sí.

- **Universo y muestra, tipo de muestreo:** Se definirá el universo de estudio y se especifica cual será el tipo de muestreo a utilizar así como la muestra que se empleará y como se seleccionarán los elementos muestrales. Si se tienen dos o más grupos de estudio se explicará como se asignan los elementos maestrales a cada grupo. (Si el procedimiento de la selección de la muestra requiere una explicación extensa se puede colocar en un anexo haciendo la referencia correspondiente.)
- **Variables:** Se definen todas las variables que se utilizarán en la investigación, su unidad de medida y su método de determinación. Se precisarán cuales son las variables independientes y cuales las dependientes si es menester, así como cuales pueden resultar variables de confusión y como se minificará el efecto de las mismas.
- **Técnicas y procedimientos a utilizar:** Se describirán pormenorizadamente las técnicas y procedimientos que se utilizarán en la investigación.
- **Instrumentos de recolección de información (Cuestionario):** Se confeccionarán los instrumentos para la recogida de la información donde deben estar únicamente las variables que se van a utilizar en el estudio. No se debe recoger dato que no se vaya a utilizar en el estudio. El cuestionario tiene que estar debidamente elaborado e identificado y se tiene que precisar donde se almacenará una vez lleno y quien es responsable de su custodia. Si es menester realizar un instructivo de su llenado o una guía de entrevista se confecciona también. (Los instrumentos para la recolección de la información, cuestionarios y guías, se colocan en un anexo.)
- **Plan de recolección de datos:** Se confeccionará un plan para la recolección de la información determinando quien, donde y cuando hará la recogida de que datos y durante que periodo de tiempo. También se especificará como se hará la supervisión de la recolección de datos, como se detectarán posibles errores sistemáticos, las medidas para evitarlos y como se subsanarán los errores detectados durante esta fase del trabajo.
- **Plan de procesamiento de datos:** Se confeccionará un plan de procesamiento de datos especificándose como se realizará el procesamiento y que medios técnicos se emplearán. Si es menester utilizar medios computacionales se determinará que sistemas se requieren y tipo de computadora, la base de datos a utilizar con su diseño, cuando, quien y como se realizará el llenado de la base de datos. Se precisarán las medidas para evitar errores de trascripción y su corrección así como las medidas de seguridad y protección, cantidad de salvas, frecuencia de las salvas y donde se guardarán las mismas. Es necesario también especificar como se supervisará el llenado de las bases de datos. También se identificarán los procederes y métodos estadísticos a emplear para cumplir cada uno de los objetivos del trabajo, precisando el nivel de significación estadística a utilizar.
- **Plan de análisis de los datos:** Se precisará como se van a analizar los datos para dar respuesta a cada uno de los objetivos del estudio. Es menester para ello diseñar las tablas estadísticas que se utilizarán para presentar y analizar la información. Ello garantizará que todas las variables necesarias se hayan incluido en el estudio y haya correspondencia entre los objetivos, variables y resultados. (El Plan de tabulación, o sea, la relación de tablas diseñadas se colocan en un anexo.)
- **Plan de adiestramiento de personal**: Si se requiere adiestrar personal en alguno de las técnicas o procedimientos a emplear, en la realización de entrevista, o llenado de cuestionarios, en el llenado de la base de datos, en fin en cualquier aspecto de la

investigación es imprescindible elaborar un plan de adiestramiento. En el mismo se precisa el contenido del adiestramiento, donde se realizará, quien lo imparte. quien lo recibe, duración del mismo y forma de comprobar su asimilación.

- **Realización de prueba piloto si se requiere:** Se recomienda realizar una prueba piloto de los instrumentos fundamentales de la investigación antes de su real ejecución. Esto permitirá detectar posibles errores, fallas u omisiones involuntarias y proceder a su rectificación antes de la investigación a gran escala. La prueba piloto requiere un plan y especificar quienes participarán, su magnitud, alcance, periodo de tiempo de realización y cuales son los aspectos que se van a valorar en ella.
- **Consideraciones éticas:** Es imprescindible plantear las consideraciones éticas que se seguirán durante toda la investigación y en cada una de sus partes. Para los elementos muestrales, en caso de ser pacientes, se debe recoger consentimiento informado previo informe al paciente o familiar responsable en que consiste la investigación. (El modelo de consentimiento informado se coloca en un anexo.)

Para más detalles de estos aspectos consultar el libro **Metodología de la Investigación (Apuntes)** del mismo autor.

Plan de trabajo (Cronograma)

El plan de trabajo es un cronograma, gráfico o diagrama que resume de una forma clara las diferentes tareas del proyecto de investigación y su acoplamiento. Durante su preparación se deben de tener en cuenta algunos aspectos importantes. Debe ser sencillo, realista y de fácil comprensión por aquellos que están directamente involucrados. Debe cubrir las fases de preparación e implementación del proyecto, así como el análisis de los datos y el informe final y si es menester la diseminación y utilización de resultados.
Es conveniente considerar las costumbres locales (días festivos y feriados) así como el horario laboral y considerar las variaciones estacionales que pueden influir en el estudio.

El mismo se puede presentar en forma de tabla colocando en la primera columna las tareas a desarrollar, en la segunda columna la fecha de inicio, en la tercera la fecha de terminación y en la cuarta columna quien ejecuta la tarea. También se puede presentar como un diagrama de Gantt donde en la primera columna se colocan las tareas y ejecutor y las otras columnas representan la secuencia de tiempo en meses y años. En ambas formas de presentación, cada fila corresponde a una tarea colocándose en orden cronológico y lógico desde la primera que por lo general corresponde a la confección del proyecto de investigación hasta la última que es la presentación del informe final. Mientras más se detallen las tareas a realizar, mayor utilidad tendrá el plan de trabajo.

El plan de trabajo es una herramienta muy útil para planificar los detalles de las actividades del proyecto, para la elaboración del presupuesto, como visión esquemática de la secuencia de las tareas, pudiendo facilitar presentaciones y negociaciones con los que tienen que autorizar o financiar el proyecto. Además, es una herramienta de trabajo para el investigador principal y el equipo de investigación mostrando las tareas, cuándo deben ejecutarse y quién está involucrado. Es imprescindible para el monitoreo y evaluación, cuando el estado actual del proyecto es comparado con el plan al determinarse en él los plazos de cumplimiento de cada tarea, y sobre todo, cuando hay que brindar reportes parciales de su cumplimiento.

El primer borrador de plan de trabajo debe prepararse durante la elaboración de la propuesta para discutir con las autoridades pertinentes. Después se confecciona un plan de trabajo más elaborado y ajustado con los resultados de la prueba o estudio piloto. Si fuera menester, no se debe dudar en revisar el plan de trabajo y hacer uno nuevo durante la ejecución del proyecto, en base a la evaluación de lo que se está haciendo realmente y lo que falta por hacer.

Recursos necesarios (Presupuesto)

Todo proyecto de investigación requiere establecer los recursos humanos y materiales que se necesitan para acometerla. Ello se determina por la cantidad de profesionales, técnicos y otro personal que hiciera falta para su realización, cuanto tiempo van a trabajar y cuanto representa eso en salarios. También se requiere conocer detalladamente los recursos materiales que se van a necesitar, equipos, exámenes complementarios, insumos, y transportación. En fin, todo gasto material que se requiere y debidamente cuantificado en moneda nacional o libremente convertible según sea el caso. Los recursos materiales se presentan en detalle con la cantidad que se requiere y el costo de cada uno. Ello permite conocer el presupuesto de la investigación.

Aunque el presupuesto se prepara al final de la planificación del proyecto, los costos son un factor limitante y siempre deben tenerse en mente durante la elaboración de la propuesta de investigación. La elaboración del presupuesto ayuda a identificar los recursos que se encuentran localmente disponibles y cuales deben ser adquiridos; y considerar siempre que el uso de recursos locales incrementa la factibilidad desde el punto de vista financiero de un proyecto. También el presupuesto estimulará a considerar aspectos del plan de trabajo que no se habían pensado antes y servirá como un recordatorio de las actividades planificadas. Es importante recordar que los ministerios y las agencias otorgadoras de fondos generalmente ponen límite al presupuesto de los proyectos de investigación.

Bibliografia

La revisión de literatura (búsqueda bibliográfica) es un aspecto de suma Importancia cuando se prepara una propuesta de investigación. Ello previene duplicar esfuerzos y ayuda a encontrar lo que otros han aprendido y reportado referente al problema que se desea investigar. Esto ayuda a precisar el planteamiento del problema, familiarizarse con diferentes metodologías que pueden usarse en el estudio y proveer argumentos convincentes relacionados con la necesidad del proyecto de investigación. Aunque la revisión bibliográfica de una investigación puede comenzar incluso antes de iniciar la elaboración del proyecto, la misma no se termina hasta que se redacta el informe final. Como ya se señaló anteriormente, en el proyecto de investigación no se requieren muchas citas bibliográficas.

Las fuentes de información son disímiles e incluyen desde individuos, grupos y organizaciones, pasando por información publicada (libros, revistas, índices, revistas referativas y otros) hasta información no publicada (otras propuestas de investigación, informes de investigaciones, registros, bases de datos de computadoras). Hoy día se viabiliza mucho la búsqueda bibliográfica haciendo uso de INTERNET y accediendo a diferentes bibliotecas y revistas médicas así como utilizando los sistemas que existen

para la recuperación de información bibliográfica mediante palabras claves tales como Medline, Bireme y otros similares.
La bibliografía se debe acotar según el sistema VANCOUVER que establece las siguientes reglas:
Para un artículo: Autores (Apellido seguido de iniciales). Título del artículo. Nombre de la revista, año; número del volumen: páginas del artículo
Para un libro: Autores (Apellido seguido de iniciales). Título del libro. Edición. Lugar: Editora, año; número de páginas del libro
Para un capítulo de un libro: Autores del capítulo (Apellido seguido de iniciales). Título del capítulo. En: Editores del libro (Apellido seguido de iniciales). eds., Título del libro. Edición. Lugar: Editora, año; páginas del capítulo
Para un sitio de internet: Autores y resto de información según sea un artículo de revista, libro o capítulo del libro como se expresó anteriormente. A continuación se coloca: Dirección del sitio WEB. fecha de acceso

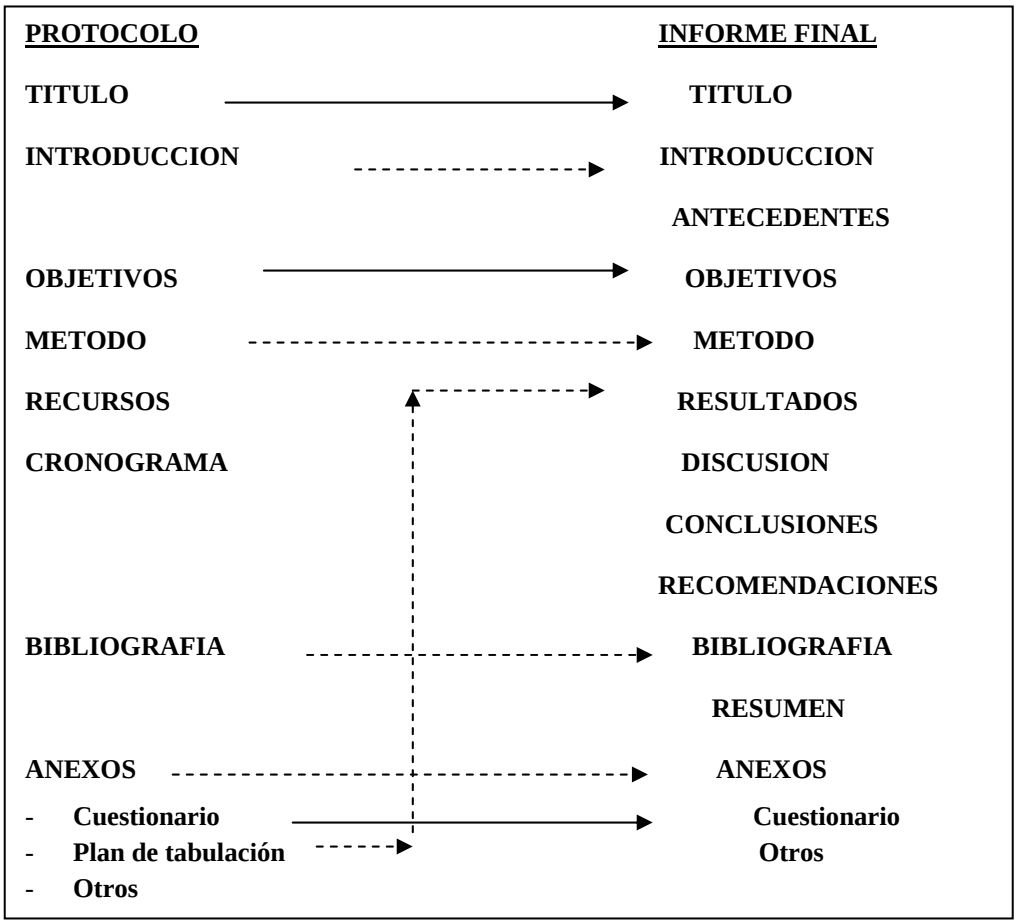
Por lo general, la bibliografía se numera consecutivamente comenzando por el número 1 en el mismo orden en que aparecen citadas o utilizadas en el texto. Otra forma de utilizarse es por orden alfabético del autor principal.

Anexos
Los anexos se numeran por lo general con números romanos comenzando por el uno. Hay dos anexos obligados, los instrumentos de recolección de información y el Plan de tabulación. El resto de los anexos está en dependencia de la investigación. Se exponen algunos ejemplos a continuación:

- **Instrumentos de recolección de información**
- **Plan de tabulación**
- **Otros**
  - **Planilla de consentimiento informado**
  - **Método de muestreo**
  - **Definiciones y conceptos**
  - **Fotos**

El informe de investigación. Importancia y partes que lo integran.
El informe de investigación también conocido como ***"informe final de la investigación"*** es el documento con el cual se concluye una investigación. El mismo resume el quehacer científico realizado y debe expresar en forma clara el nuevo conocimiento alcanzado sustentado por la base teórica utilizada, la forma de realización de la investigación, los resultados obtenidos y su discusión y las conclusiones y recomendaciones.

Este documento se vincula estrechamente con el ***"proyecto de investigación o protocolo"*** donde se planificó la investigación. Esa relación se presenta en el esquema siguiente:



La línea continua en el esquema significa que debe ser el mismo en ambos documentos y la discontinua que lo que se coloca en el informe final incluye parte de lo que se expresó en el proyecto de investigación.

Así, el titulo de la investigación es el mismo en ambos documentos, al igual que los objetivos, incluyendo el enunciado de la hipótesis si la hubiera, y los instrumentos de recolección de información que se exponen en los anexos. Algunos elementos de la introducción en el protocolo se pueden utilizar en la introducción del informe final; el método del informe final es un resumen del método expresado en el protocolo; los resultados del informe final se basan en el plan de tabulación del protocolo; la bibliografía utilizada en el protocolo puede ser utilizada en el informe final y, por último, algunos anexos incluidos en la categoría de otros en el protocolo pudieran repetirse en el informe final si fuera necesario.

Las fuentes de información para los aspectos esenciales del contenido del informe final de una investigación se resumen en el esquema que se presenta a continuación:

CONTENIDO	FUENTE DE INFORMACION
INTRODUCCION	**BIBLIOGRAFIA**
ANTECEDENTES	**BIBLIOGRAFIA**
OBJETIVOS E HIPOTESIS	**PROTOCOLO**
METODO	**PROTOCOLO**
RESULTADOS	**TRABAJO INVESTIGATIVO** **PROCESAMIENTO DE DATOS** **PROTOCOLO (Plan Tabulación)** **ANALISIS E INTERPRETACION DE DATOS**
DISCUSION	**BIBLIOGRAFIA** **ANALISIS E INTERPRETACION DE RESULTADOS**
CONCLUSIONES	**ANALISIS E INTERPRETACION DE RESULTADOS Y DISCUSION**
RECOMENDACIONES	**ANALISIS E INTERPRETACION DE RESULTADOS Y DISCUSION** **ANALISIS E INTERPRETACION DE CONCLUSIONES**
BIBLIOGRAFIA	**BUSQUEDA BIBLIOGRAFICA** **BIBLIOGRAFIA DEL PROTOCOLO**

Las principales fuentes de información están en la búsqueda bibliográfica que se realiza en toda investigación, en el protocolo o proyecto de investigación que se elaboró con anterioridad y en los resultados de la investigación. Como es fácil apreciar a medida que se avanza en la investigación un aspecto terminado es fuente de información para el que sigue. Así la discusión se obtiene de comentar los resultados encontrados sobre la base de compararlos con lo reportado en la bibliografía consultada. Las conclusiones se obtienen de los resultados y la discusión. Las recomendaciones surgen de las conclusiones, resultados y discusión. Lo que debe quedar bien claro es que todo aspecto que se reseñe en alguna parte del informe final de la investigación tiene que tener una fuente explícita. No puede haber aspecto alguno ***"sacado de la copa del sombrero de un mago"***.

Importancia.

La importancia fundamental del informe final de cualquier investigación es ***"mostrar los nuevos conocimientos científicos que la misma aportó"***. Los aspectos más relevantes sobre la importancia del informe final de investigación se listan a continuación.

- Presentar los resultados de la investigación.

- Mostrar fehacientemente los fundamentos teóricos sobre la base de los cuales se realizó la investigación.
- Exponer, mediante un análisis crítico comparativo de los resultados obtenidos con los de otros autores, su veracidad, relevancia así como cualquier aspecto novedoso.
- Concluir claramente los nuevos conocimientos hallados en la investigación incluido en ello el análisis de la hipótesis científica inicial.
- Indicar los mecanismos más idóneos para incorporar a la práctica social los resultados de la investigación.

Partes que lo integran.

No existe un formato único para presentar el informe final de una investigación y cada institución, organización internacional, organizaciones No gubernamentales, en fin cada cual puede tener un formato diferente. En nuestro país existe un formato específico, por ejemplo, para presentar el resultado del trabajo investigativo cuando se está optando por el grado de Doctor en Ciencias, otro para presentar el resultado de la investigación cuando se opta por una categoría científica menor como Master o Especialista en cualquier rama de la medicina entre otros. No obstante la diferencia del formato de presentación, el contenido del informe final de investigación es muy parecido en todos pues refleja el método científico en su conjunto. Veremos a continuación cuales son los aspectos que, independientemente del formato de presentación, se deben plasmar en todo informe final de investigación. Ellos son:

- Título
- Resumen
- Introducción
- Antecedentes
- Objetivos
- Método
- Resultados
- Discusión
- Conclusiones
- Recomendaciones
- Bibliografía
- Anexos
  - Cuestionario (Modelos de recogida de datos)
  - Otros

Título

Ya se señaló que el título debe ser el mismo que el planteado en el proyecto de investigación. En contadas ocasiones se hacen pequeñas modificaciones que no cambian la esencia del problema.

En la misma hoja del título se colocan el o los autores de la investigación.

Resumen

El resumen de la investigación puede presentar una extensión variable en dependencia del formato que se esté utilizando. Por lo general se solicita un resumen de una o dos

páginas como máximo donde se reflejen los aspectos fundamentales de la investigación. En ocasiones se solicita que sea un resumen estructurado donde se plasmen los aspectos esenciales de cada acápite o capítulo del informe final de la investigación. Lo que se persigue con el resumen es conocer ***"a vuelo de pájaro"*** el contenido del informe final de la investigación.

Introducción

La introducción del informe final de la investigación deviene el ***"sustento socio-técnico-político del trabajo científico realizado"***. En el mismo se incluye la importancia del problema de investigación, el impacto del problema en la sociedad en general, en los países desarrollados y desarrollados y el impacto del problema en el país donde se desarrolla el trabajo. Ello es llevar el problema de lo general a lo particular del entorno donde se hace la investigación, profundizando entonces en el impacto del problema en el ámbito de estudio y/o trabajo que debe haber sido la fuente inspiradora de la investigación.

Otro aspecto a tener en cuenta en la introducción es la utilidad que se espera obtener con los resultados de la investigación realizada y que se expondrá en los acápites siguientes. Esa utilidad se debe exponer en términos generales y, si es factible, particularizarla en aspectos específicos.

Hay formatos que establecen un límite para la introducción, otros no lo hacen. De no precisarse límite, el mismo estará dado por el contenido e importancia del problema de investigación así como de la búsqueda bibliográfica que se haya realizado.

Antecedentes

Es el fundamento teórico del trabajo científico realizado. Los antecedentes muestran todo el conocimiento previo y logros alcanzados relativos al problema que se está analizando, hasta el momento en que se inició la investigación. Por ello la bibliografía utilizada debe tener una fuerte proporción de citas (alrededor del 50%) de los últimos cinco años.

Se pueden estructurar los antecedentes en ***"antecedentes históricos"*** donde se refleje la historia del problema de investigación desde que se tuvo conciencia del mismo por primera vez. A continuación los ***"antecedentes generales"*** donde si la investigación trata sobre una enfermedad determinada se debe exponer la etiología, patogénesis y fisiopatología, síntomas y signos, tratamiento y evolución. Si la investigación versa sobre un proceder terapéutico o método diagnóstico se debe plantear quien, cuando y donde se aplicó primero, diferentes enfoques que puedan existir en cuanto a su aplicación, ventajas y desventajas de cada enfoque, aplicación en el país y aplicación en el ámbito de trabajo y/o estudio. Por último, tenemos los ***"antecedentes específicos"*** relacionados con el problema de investigación y al contexto en que está siendo investigado. Se deben especificar trabajos previos realizados en el país y enfatizar en aquellos hechos en el ámbito de trabajo y/o estudio de los investigadores.

Es aconsejable finalizar estos fundamentos teóricos de la investigación con el ***"enunciado del problema científico"*** que se abordó, las ***"interrogantes de investigación"*** que se pudieron tener y la ***"hipótesis científica"***.

En ocasiones se unen dentro de un mismo acápite del informe final la Introducción y los antecedentes.

Objetivos
Los objetivos deben ser los mismos planteados en el proyecto de investigación. En caso que durante la investigación haya surgido algún nuevo objetivo o haya sido necesario eliminar alguno de los señalados en el protocolo, se debe aclarar explícitamente en alguno de los acápites del informe final.

Método
En esta parte del informe final se resume el método planteado en el proyecto de investigación. Por lo general, en el cuerpo del informe final se describe el tipo de estudio, la población o universo de estudio, el método de muestreo utilizado y el tamaño de la muestra, los criterios para seleccionar los elementos muestrales, una breve descripción de los métodos empleados y una breve descripción de la recolección, procesamiento y el análisis de los datos. El análisis estadístico se puede reseñar por cada objetivo específico de la investigación

En los anexos se deben colocar la definición de términos y variables, el cuestionario (modelos y guías para la recogida de información), el método de muestreo (sí se requiere y es extenso), la descripción detallada de procederes quirúrgicos, terapéuticos o diagnósticos (sí se requiere), y otros aspectos que se considere importantes.

Resultados
Los resultados son los logros prácticos del trabajo investigativo realizado. Los mismos se presentan en forma de tablas y gráficos estadísticos que pueden estar incluidos en el acápite de resultados o en un anexo según sea la preferencia de los autores o lo establezca el formato utilizado. Aquí solamente se plasma lo que se encontró en la investigación una vez que se procesaron y analizaron los datos recogidos.

Esta parte del informe final se debe desarrollar en una secuencia lógica siguiendo la secuencia de los objetivos. Se pueden agrupar las diferentes tablas estadísticas con diferentes subtítulos o por objetivos

Discusión
La discusión es el resultado teórico de la investigación realizada. Es el sustento de los nuevos conocimientos científicos logrados. La discusión puede ser un acápite por sí sola o hacerse conjuntamente con los resultados, titulándose entonces Resultados y Discusión.

Esta parte, de hacerse de forma independiente, debe seguir el mismo orden que los resultados, que fueron escritos en una secuencia lógica. Para cada resultado (tabla estadística) se debe tener una discusión comparando los resultados con los de otros autores. Ello requiere de una bibliografía actualizada siendo necesario que la mayoría de las citas corresponda, preferiblemente, a los últimos cinco años. A veces no es posible discutir todas las tablas estadísticas pues no se halla bibliografía al respecto. Se deben discutir la mayoría de los resultados, al menos el 70% de las tablas estadísticas, con lo

planteado por otros autores. En la comparación con otros autores se debe seguir una secuencia lógica yendo de lo general a lo particular (condiciones similares a las de nuestro entorno). En ocasiones, para resaltar aquellos resultados mas relevantes es conveniente mostrarlos mediante un gráfico estadístico, como complemento de la tabla estadística.

Es aconsejable finalizar la discusión con un análisis crítico de la solución del problema científico, respuestas a las interrogantes de investigación y de la hipótesis científica así como de la novedad científica del nuevo conocimiento alcanzado.

Conclusiones
Las conclusiones del informe final de una investigación devienen el resumen de los nuevos conocimientos científicos obtenidos por el trabajo realizado. Por ello tienen que ser muy claras, responder a los objetivos del trabajo y estar explícitas en los resultados o en la discusión. Es recomendable que en aquellas investigaciones donde exista hipótesis científica, la primera conclusión se refiera a ella y se establezca claramente si la misma se cumple o no con los resultados encontrados.

Las conclusiones no son la repetición de los principales resultados, ni la repetición de los datos estadísticos relevantes mostrados en las tablas y gráficos estadísticos. Las mismas son un enunciado del nuevo conocimiento alcanzado producto del análisis de los resultados obtenidos al contrastarse, en la discusión, con los criterios y resultados de otros autores.

Recomendaciones
Las recomendaciones del informe final de la investigación pueden plasmar la forma en que los investigadores consideran que se debe generalizar el nuevo conocimiento científico alcanzado. También es el lugar para colocar sugerencias de nuevas investigaciones a realizar para darle continuidad a los resultados encontrados así como para aclarar incógnitas surgidas en la investigación terminada.

Las recomendaciones tienen que ser muy claras y estar en correspondencia con las conclusiones. Ellas surgen del análisis de los resultados, la discusión y de las conclusiones.

Bibliografía
La bibliografía o referencias bibliográficas es la enumeración de todas las citas bibliográficas que se emplearon en el informe final de la investigación. Los aspectos señalados en párrafos anteriores al respecto cuando se habló del proyecto de investigación son válidos aquí por lo que no se repetirán. No existe un límite máximo de bibliografías a utilizar pero usualmente el mínimo de referencias bibliográficas requeridas en una investigación científica es alrededor de setenta u ochenta.

Anexos
Los anexos es la parte final del informe de investigación. Lo aspectos generales señalados en párrafos anteriores cuando se describieron los anexos del proyecto de investigación son válidos aquí. Es obligado colocar los instrumentos de recolección de

información utilizados, así como las variables, conceptos y definición de términos utilizados en el informe final. El resto de los anexos está en dependencia de la investigación. Se exponen algunos ejemplos a continuación:

- **Instrumentos de recolección de información**
- **Listado de variables**
- **Conceptos y definición de términos empleados**
- **Otros**
  - **Tablas estadísticas si no se muestran en los resultados**
  - **Gráficos estadísticos si no se muestran en los resultados**
  - **Esquemas o diagramas explicativos**
  - **Planilla de consentimiento informado**
  - **Método de muestreo si procede**
  - **Técnicas o procederes empleados en la investigación**
  - **Fotos**

Por último, consideramos oportuno reseñar algunas sugerencias para la redacción del informe final de la investigación:

- Señale opiniones y criterios de otros autores
- Siempre siga un orden lógico, de los aspectos generales a los particulares
- Puede colocar citas textuales si lo desea utilizando “comillas”
- Puede hacer resúmenes bibliográficos si así lo desea
- Todos los anexos deben estar referenciados en el cuerpo de la investigación
- Toda la bibliografía utilizada debe ser acotada

Después de haber escrito el texto, déjelo y revíselo después de varios días. !No se sorprenda si lo cambia por completo!

INDICE

BIBLIOGRAFIA CONSULTADA

Knapp RG, Clinton Miller M. Clinical Epidemiology and Biostatistics. First Edition. Harwal Publishing Company, Malvern, Pennsylvania., 1992: 435

Beaumont GP. Elementary Mathematical Statistics. First Edition. Londom. McGraw Hill Inc., 1973: 312

Bliss CI. Statistics in Biology. Vol. 1. First Edition. London: McGraw-Hill Book Company, 1967: 557

Bradford Hill A. A short Textbook of medical Statistics. Tenth Edition. London: The English Language Book Society and Hodder and Stoughton, 1981: 325

Morales Jiménez EL, Hattab AS, Reinoso Medrano T. Health Statistics and Research Methodology. First Edition. University of Aden Printing & Publishing House, 1997: 237

Brito JA, Cathcart F, Colunga C, Domínguez GI, Durán I, García MN, González N, Lamartche L, O'Farril E, Oliva L, Paez LA, Pozo R, Rios E, Rodríguez L, Rodríguez E. Torres JA. Bioestadística y Computación. Primera Edición. Habana. Editorial Ciencia y Educación, 1987: 228

Cochram WG. Sampling Techniques. Third Edition. New York; John Wiley and Sons, Inc., 1966: 611

Daly LE, Bourke GI McGilvray J. Interpretation and Uses of medical Statistics. Fourth Edition. London: Blackwell Scientific Publications, 1981: 441

Daniel WW. Biostatistics: A Foundation for Analysis in the Health Sciences. Third Edition. new York: John Wiley and Sons, 1983: 534

Edwards AL. Statistical Methods. Third Edition. London: Holt, Rinehart and Winston Inc., 1973: 312

Hoel PG. Estadística Elemental. Segunda Edición. Habana: Instituto del Libro, 1969: 351

WHO. International Stattistical Classification of Diseases and Related health problems. Tenth revision. Vol. 2. Geneva: World Health organization, 1992: 160

Colectivo de autores. Informática Médica Tomo 2 Bioestadística. Primera Edición. Editorial Ciencias Médicas. Habana, 2004:183-632

Morales Jiménez, EL. Metodología de la Investigación (Apuntes). Primera Edición. ELACM, La Habana, 2001:102