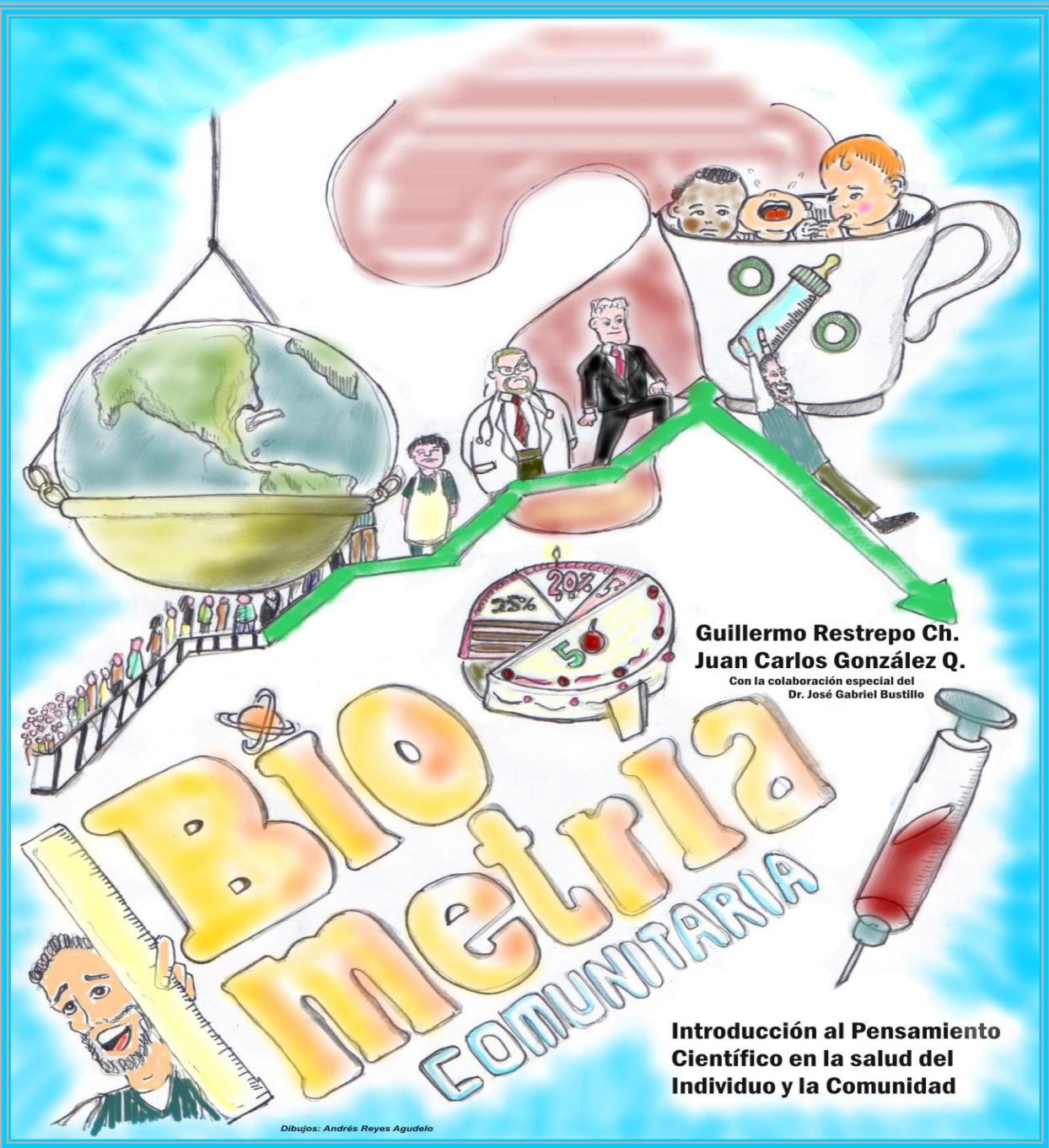




TEXTO BÁSICO DE BIOMETRÍA





FUNDACIÓN UNIVERSITARIA
JUAN N. CORPAS

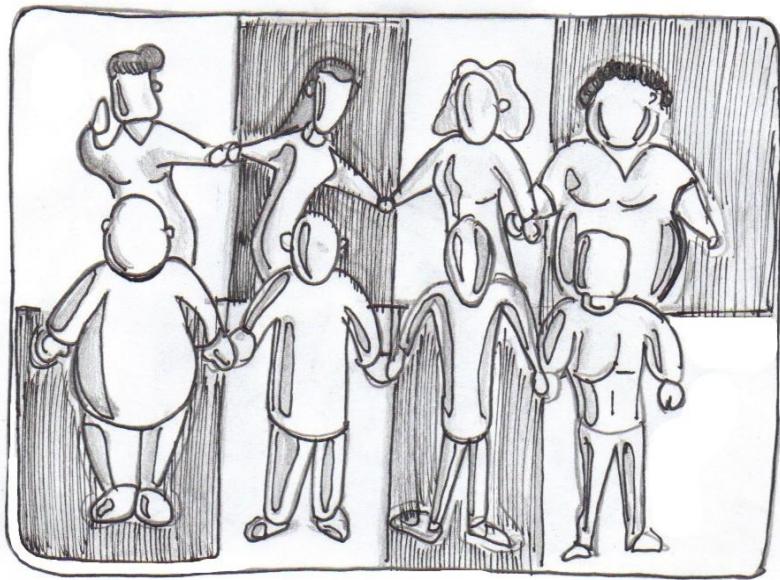
ISSN 978-958-44-6894-9
Libro de Biometría Comunitaria
Fundación Universitaria Juan N Corpas
Facultad de Medicina
Departamento de Medicina Comunitaria
Bogotá - Colombia
2010

CONTENIDO

INTRODUCCIÓN AL TEXTO	6
CAPÍTULO 1: INTRODUCCIÓN A LA EPIDEMIOLOGÍA	8
1.1. Palabras para una definición.....	9
1.2. La utilidad de la epidemiología.....	11
1.2.1. Buscando las causas	11
1.2.2. La ayuda para la salud pública	14
1.3. Las herramientas de la epidemiología.....	16
1.3.1. La observación	16
1.3.1.1. La descripción	18
1.3.1.2. La comparación.....	20
1.3.1.3. La intervención	21
1.4. Las ayudas en la epidemiología	22
1.4.1. El raciocinio científico	22
1.4.2. El Papel de la estadística.....	23
1.5. La estrategia de la epidemiología	25
1.5.1. La investigación	25
1.5.2. Las variables	26
1.5.3. Los datos y las medidas	29
1.5.4. La búsqueda de la validez para la salud.....	30
CAPÍTULO 2: LA UTILIDAD DE LA EPIDEMIOLOGÍA	33
2.1. La lucha contra la enfermedad.....	33
2.1.1. El arte de describir	33
2.1.1.1. La historia natural de las enfermedades	37
2.1.1.1.1. Ejemplo: Historia de una epidemia (SIDA)	45
2.1.1.1.2. La cadena epidemiológica	48
2.1.2. La tabla de frecuencias	51
2.1.3. La forma de aprender: comparar	53
2.1.4. La tabla de comparación: los factores de riesgo y la evaluación de lo que sirve.....	54
2.1.4.1. Usos de la tabla de comparación	58
2.1.4.1.1. Identificación de riesgos	59
2.1.4.1.1.1. Presentación de una tabla de contingencia.....	64
2.1.4.1.2. Otras comparaciones: sensibilidad y especificidad.....	65
2.2. Buscando las causas	69
2.2.1 La maravilla de las relaciones y las asociaciones	70
2.2.2. La red causal.....	78
2.2.2.1. Las causas necesarias y suficientes	79
2.3. Buscando la salud pública: La vigilancia epidemiológica:.....	83
2.3.1. La importancia de los registros	84
2.3.2. El arte de la predicción.....	85
2.3.3. Aporte de la epidemiología a la salud comunitaria.....	87
2.3.3.1. La estrategia de la identificación.....	90
2.3.3.2. Para decidir	91
2.3.3.3. Para organizar.....	91
2.4 Hacia dónde va la epidemiología. Mirando la evidencia de la utilidad ...	93

2.4.1. Una mención de éxitos y limitaciones	96
Usos de la epidemiología	103
CAPÍTULO 3. LAS AYUDAS EN LA EPIDEMIOLOGÍA: LA INVESTIGACIÓN	106
3.1. <i>El raciocinio científico. La base.</i>	107
3.2. <i>Los paradigmas de la investigación.</i>	111
3.2.1. Concepto de sesgos	114
3.3. <i>Las fases de la investigación...</i>	118
FASE UNO: PLANTEAMIENTO DEL PROBLEMA	121
1.1. <i>Paso 1: Selección del tema</i>	121
1.2. <i>Paso 2: Revisión bibliográfica</i>	122
1.2.1. Cómo se revisa un artículo.....	124
1.3. <i>Paso 3: Especificación del problema o de la pregunta de investigación.</i>	128
1.4. <i>Paso 4: Identificar las probables respuestas (las hipótesis)</i>	130
1.4.1. Algo más sobre las hipótesis de investigación	134
1.4.2. Las hipótesis estadísticas	137
1.5. <i>Paso 5: Pertinencia de contestar la pregunta de investigación (o importancia de la respuesta de la pregunta)</i>	139
1.6. <i>Paso 6. Construcción de un marco teórico</i>	141
1.6.1. Definición de los componentes del problema:.....	142
1.6.2. Qué respuestas se conocen a la pregunta.....	143
1.6.3. Cómo la han contestado en otros estudios (diseños empleados, variables y manejo estadístico)	144
1.6.4. Recopilación del marco teórico (qué debe decir una tesis)	148
FASE DOS: DISEÑO Y PLANEACIÓN (PLAN PARA CONTESTAR LA PREGUNTA)	154
2. Paso 7. Selección del diseño que se empleará	154
2.1. Diseños cuantitativos	156
2.1.1. Intervencionales o experimentales	158
2.1.2. Observacionales	162
2.2. Diseños cualitativos.....	170
2.3. Investigación multi-método.....	171
2.4. Paso 8. Selección de la población	172
2.4.1. Acceso a la población	173
2.4.1.1. Criterios de inclusión	173
2.4.2. Cumplimiento de principios éticos.....	176
2.5. Muestras (de población)	178
2.5.1. Muestras representativas	178
2.5.2. Muestras por conveniencia:	180
2.6. Paso 9. Selección de las variables (instrumentos a emplear)	182
2.7. Sesgos.....	187
2.7.1. Atributos de una medida.....	189
2.8. Paso 10. Manejo estadístico de las variables	190
2.8.1. Base de datos	190
2.8.2. Paso 11. Construcción del formulario.....	192
2.9. Paso 12. Plan para el trabajo de campo	194
2.9.1. Descripción de las actividades.....	195
Consecución de la población	195
Trabajo de campo.....	197
Análisis de la información	199
Presentación de la investigación	200
2.9.2. Recursos que se necesitan para cumplir con las actividades	202

2.10. Paso 12. El protocolo de investigación	204
FASE TRES: TRABAJO DE CAMPO.....	208
3. Paso 14. Recolección de la información.....	208
3.1. Cómo se accedió a la población.....	208
3.2. Cómo se recogió la información	208
FASE CUATRO: ANÁLISIS DE DATOS.....	209
4. Paso 15. Registro de información recogida en la base de datos.....	209
4.1. Paso 16. Procesamiento la información.....	210
FASE CINCO: DIVULGACIÓN DE LA INVESTIGACIÓN	212
5. Paso 17. Publicación de la investigación.....	212
5.1. Cómo escribir un artículo	212
5.2. Paso 18. Presentación de la investigación	215
CAPÍTULO 4. EL PAPEL DE LA ESTADÍSTICA: CONSTRUYENDO UN IDIOMA ENTENDIBLE	218
4.1. <i>Los datos</i>	219
4.2. <i>Las medidas</i>	223
4.2.1. Medidas de frecuencia	224
4.2.1.1. Concepto de indicadores: Palabras para una definición	226
4.2.2. Medidas de asociación	231
4.2.3. Medidas de impacto	232
4.3. <i>La estadística descriptiva</i>	232
4.3.1. La tabla de frecuencia	233
4.3.2. Las gráficas	236
4.3.3. Las medidas de tendencia central	241
4.3.3.1.La media o promedio	243
4.3.3.2. La moda	246
4.3.3.3.La mediana	247
4.3.4. Las medidas de dispersión	248
4.3.4.1. El recorrido	249
4.3.4.2. Los recorridos intercuartiles.....	251
4.3.4.3. La desviación estándar (con los pasos para obtenerla)	252
4.3.5. La curva de distribución	256
4.3.5.1. Explicación de la curva de distribución y su utilidad a través de las tablas de peso y talla	263
4.3.5.2. La curva de distribución y la similitud y diferencias estadísticas	268
4.3.5.3. La curva de distribución y las hipótesis estadísticas	271
4.4. La estadística analítica	278
1.4.1.La herramienta esencial de la comparación (Tabla de contingencia o de 2X2) y su aplicación con IC del 95% y prueba de la p	281
1.4.1.1.Ejemplo a través de un brote del uso de la tabla	282
1.4.1.1.1.El concepto de significancia a través del ejemplo	285
4.4.1.1.1.2. Ampliando la prueba p.....	289
4.1.1.1.3. La prueba Ji (Chi cuadrado)	292
4.1.1.1.4.Entendiendo la prueba T	294
4.4.2. El error tipo I y tipo II	297
4.4.3.Las correlaciones.....	299
4.4.4.La inferencia estadística	301
A manera de conclusión....	304
Lecturas recomendadas	305



INTRODUCCIÓN AL TEXTO

Este texto es un intento de integrar la epidemiología, la investigación y la estadística en un compendio elemental y sencillo para todos los que trabajan en el área de salud y se hacen preguntas en torno a sus trabajos. Surge por la convicción de que la realidad integra estas tres áreas del conocimiento y por ello pretende presentarlas de esa manera ante los ojos de los lectores.

Está escrito de manera sencilla y ha usado un sistema de clasificación en donde se parte de una visión general (que introduce a los otros capítulos) en el capítulo uno, para adentrarse en la utilidad práctica de la epidemiología haciendo énfasis en la comprensión de la tabla de contingencia y el mundo de las relaciones, pasando en el capítulo tres a una visión de la investigación como principal herramienta de la epidemiología. Allí se mencionan algunos aspectos claves de interpretación en las investigaciones. El último capítulo es dedicado a entender el lenguaje de la investigación y, por ende, de buena parte de la epidemiología. Se hace un esfuerzo a través de ejemplos con el fin de mostrar el uso de los principales términos estadísticos.

La razón de este libro es eminentemente pedagógica. A lo mejor, aquel que lo lea con deseos encuentre en sus explicaciones una ganancia que le servirá para hacer proyecciones en la población objeto de su trabajo. Ese es el anhelo: ¡Que sirva!

Los autores.



Capítulo 1

Introducción a la Epidemiología

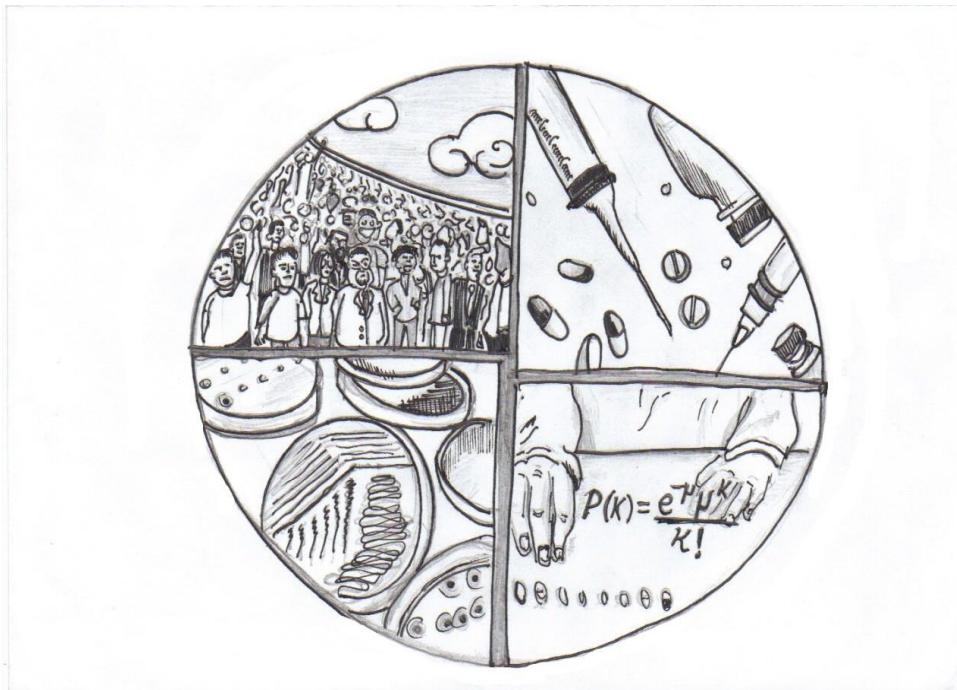


<u>1.1. Palabras para una definición...</u>	9
<u>1.2. La utilidad de la epidemiología...</u>	11
<u>1.2.1. Buscando las causas</u>	11
<u>1.2.2. La ayuda para la salud pública.....</u>	14
<u>1.3. Las herramientas de la epidemiología.....</u>	16
<u>1.3.1. La observación</u>	16
<u>1.3.1.1. La descripción</u>	18
<u>1.3.1.2. La comparación.....</u>	20
<u>1.3.1.3. La intervención</u>	21
<u>1.4. Las ayudas en la epidemiología</u>	22
<u>1.4.1. El raciocinio científico</u>	22
<u>1.4.2. El Papel de la estadística.....</u>	23
<u>1.5. La estrategia de la epidemiología.....</u>	25
<u>1.5.1. La investigación</u>	25
<u>1.5.2. Las variables</u>	26
<u>1.5.3. Los datos y las medidas</u>	29
<u>1.5.4. La búsqueda de la validez para la salud.....</u>	30



CAPÍTULO 1: INTRODUCCIÓN A LA EPIDEMIOLOGÍA

Entender la epidemiología es de un inmenso valor para todos aquellos que están dedicados a ayudar al ser humano desde el área de la salud. Como tal, el conocimiento agrupado bajo la epidemiología ha abierto una extraordinaria senda que ha permitido enormes avances para la humanidad. Su gran objeto ha sido y será combatir las enfermedades y hoy se incorpora en la búsqueda de la salud. A continuación se esbozará a manera de introducción algunos componentes del maravilloso mundo de la epidemiología.



1.1. Palabras para una definición

Unas palabras que necesariamente entran en cualquier definición de la epidemiología han de ser **enfermedad** y **población**. Sin duda, la ocupación de la epidemiología será estudiar la enfermedad en la población. Las epidemias (enfermedades casi siempre infecciosas que afectaban grupos poblacionales) como ejemplo clásico, fueron combatidas con las técnicas epidemiológicas (descripción, comparación e intervención). Ello establece un objetivo: preservar al ser humano. Las **frecuencias** (como expresión de ocurrencia), sin duda, han de acompañar cualquier definición. Y qué decir de la **etiología**, probablemente sin esta palabra no existirían los grandes éxitos que la epidemiología se ha anotado. Pero hoy... hoy no se habla solo de la enfermedad... se habla de la **salud**, por ende, la epidemiología tiene que ver también con la búsqueda de la etiología para mantener la salud. Por ello, la definición de epidemiología deberá contemplar, población, enfermedad, salud, frecuencias (ocurrencias), etiología... claro, falta la palabra que va integrar todo estos aspectos... ¿será ciencia? o, ¿técnica?... la verdad es que al momento de estas líneas, no se sabe. Una ciencia tiene una cantidad de componentes que tal vez la epidemiología aún no tiene (teorías e hipótesis) y una técnica parece ser muy simple para la epidemiología. A lo mejor, el **método científico** puede ser ese enlace que "activa" la epidemiología...



Definiciones¹:

Enfermedad: alteración más o menos grave de la salud.

Frecuencia: número de veces que se repite algo.

Etiología: estudio de las causas de las cosas.

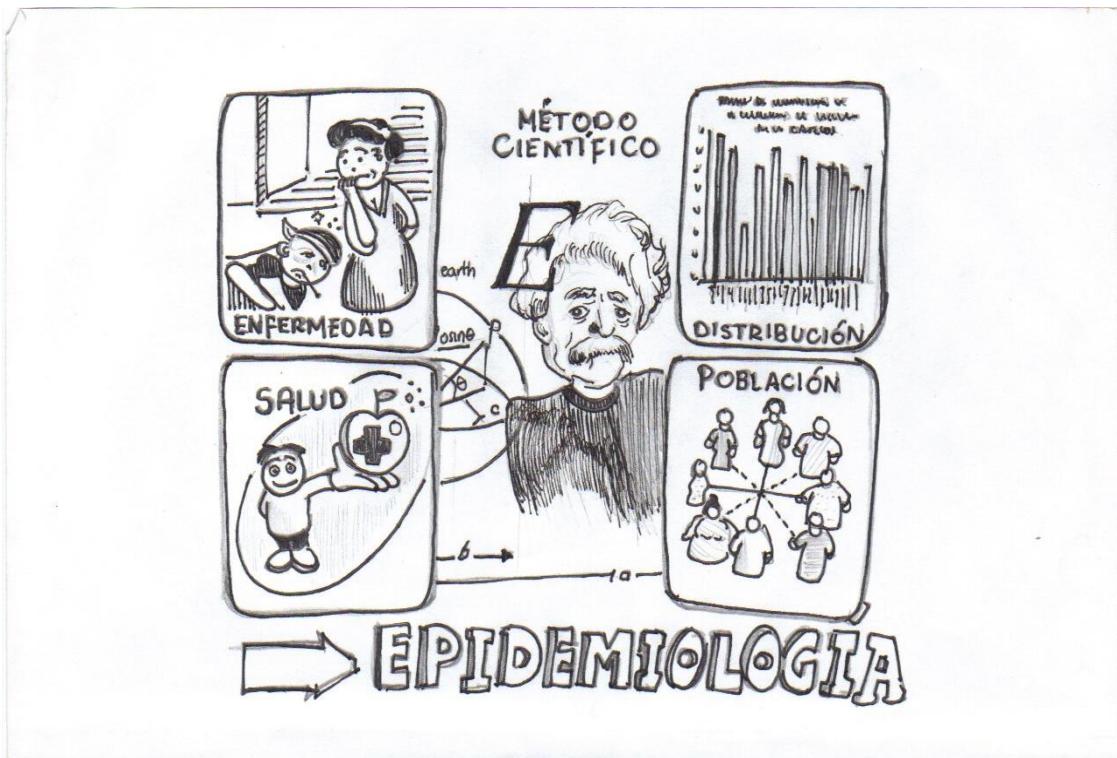
Salud: estado en el que el ser humano ejerce normalmente todas sus funciones.

Método científico²: método de estudio sistemático de la naturaleza que incluye las técnicas de observación, reglas para el razonamiento y la predicción, ideas sobre la experimentación planificada y los modos de comunicar los resultados experimentales y teóricos.

Población: Conjunto de personas que habitan la Tierra o cualquier división geográfica de ella.

¹ Real Diccionario de la Lengua.

² Enciclopedia Encarta 2005.



Como una propuesta de definición se propone: Es el uso del método científico en el estudio de la enfermedad y la salud de la población, buscando la causa para, así entender, qué se puede hacer en beneficio del hombre.

De todo lo anterior ha de quedar en claro que la epidemiología tiene una gran utilidad en la salud de los seres humanos.



Epidemiología: Según Last J M en su diccionario de epidemiología: el estudio de la distribución y de los determinantes de los estados o acontecimientos relacionados con la salud en las poblaciones específicas y las aplicaciones de este estudio al control de los problemas sanitarios.

Lilienfeld: el estudio de la distribución de una enfermedad o una condición en una población y de aquellos factores que influyen en su distribución.

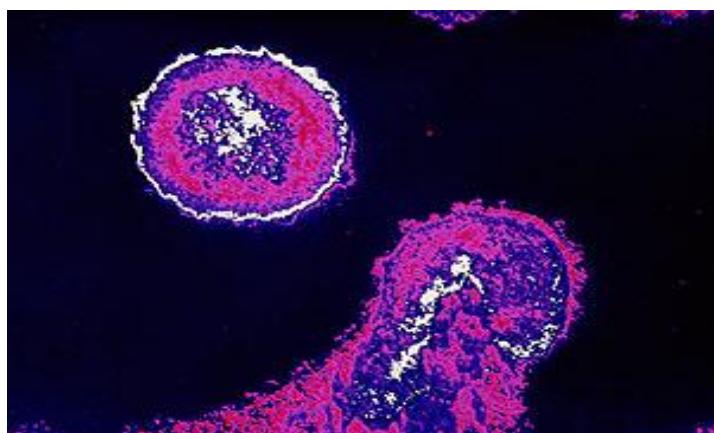
Gordon: el estudio de la enfermedad y como ella ocurre en la naturaleza.

Mac Mahon y Pugh: el estudio de la distribución y de los determinantes de la frecuencia de la enfermedad en el hombre.

1.2. La utilidad de la epidemiología

1.2.1. Buscando las causas

Para combatir la enfermedad, el saber la causa es un elemento importante, que no indispensable. La historia ha demostrado cómo a pesar de no saber la causa, la epidemiología ha contribuido a desarrollar una serie de medidas protectoras que protegen al ser humano. Por ejemplo, el médico James Lind en 1740 a través de un sencillo experimento, descubrió cómo combatir el escorbuto mucho antes de saber cuál era la causa. Pero al saber la causa, esta tarea se facilita enormemente. Es más... bajo esta condición, la enfermedad puede ser derrotada. Hoy surge con la misma fuerza la búsqueda de la "causa" que puede mantener sano al ser humano. Pero no cabe duda que la humanidad esté más entrenada en torno el estudio de la etiología de la enfermedad.



En 1980 apareció una temible enfermedad que empezó a "matar" a gente joven por infecciones que en situaciones normales no producían mayor mortalidad. El Dr. Gottlieb descubrió que esas personas tenían un "problema" en sus glóbulos blancos. Cuando revisó casos similares y dispuso que colegas de distintas partes indagaran, se dio la alarma en el mundo pues se estaba ante una PANDEMIA. Pronto se detectó que en la comunidad de San Francisco estaban apareciendo una gran cantidad de casos. Un conector común fue que en un comienzo hubo un alto porcentaje de casos en pacientes homosexuales, situación que luego pudo especificarse en torno a que se trataba de personas con una vida sexual promiscua. El mundo se alarmó pues la multiplicación de casos se hacia de una forma exagerada y los investigadores de los países con recursos se lanzaron a una búsqueda de la causa de esta enfermedad. Cuatro años después, el profesor Luc Montagnier y el Dr. Robert Gallo le decían al mundo que habían descubierto el virus responsable de esta temida enfermedad: el virus del SIDA. Gracias a ello y en solo tres años, se descubrieron los medicamentos antirretrovirales que tienen la capacidad de cambiar el pronóstico.

Hoy se sabe que nada es el producto de un solo componente. La multicausalidad parece reafirmar la certeza de la "complejidad". Se habla de "redes" y de "pasteles" causales. Se habla de causas necesarias y suficientes; intentos todos de ordenar y clasificar el sorprendente mundo etiológico (ver capítulo 2, numeral 2.2). Se habla de factores de riesgo y de factores protectores, de componentes "primordiales", en fin... según y como pretenden ordenar los grandes "pensadores" del mundo de la medicina, el estudio de la etiología o de la causa emblematiza la utilidad inmensa que tiene la epidemiología en torno al ser humano.



La gráfica expresa un ejemplo a través de la génesis de la TBC: cómo distintos componentes "causales" permiten que aparezca la enfermedad. Algunos pueden tener más "peso" que otros, pero, al fin y al cabo, es la suma de todos los que completan el "pastel" causal (círculo).



Un interesante ejemplo de la utilidad del pensamiento etiológico, o lo que es lo mismo, causal de la epidemiología, es la investigación de Semmenweis (5) en torno a la fiebre puerperal en 1840... A través de un juicioso análisis Semmenweis logró establecer lo que se podía hacer para combatir la "fiebre puerperal" aún antes de saber la existencia de los microbios responsables de las infecciones. Estableció que un lavado adecuado de las manos antes de valorar a las mujeres en trabajo de parto podía disminuir la incidencia (aparición de casos nuevos) de la temible enfermedad.

Si bien Semmenweis no descubrió al estreptococo, responsable etiológico de la fiebre puerperal, sí descubrió que en la "causa" de la fiebre puerperal, el no lavado de las manos es un factor esencial. Interviniendo sobre este componente causal, logró modificar la distribución de la enfermedad.



1.2.2. La ayuda para la salud pública

Así como contribuye al estudio de la causalidad, la epidemiología es una de las bases de la salud pública. ¿Por qué? Para ello se hace necesario “entender” el significado de la salud pública. Aceptando lo difícil que es encontrar una definición, se puede afirmar que la que sea, ha de tener como componentes esenciales: **comunidad** referida como hacia dónde se dirige; **bienestar** como el objetivo que se busca; **Estado** por ser el responsable. Tal vez con otros componentes, pero seguros de que estos términos han de participar en cualquier definición, se deja en la búsqueda de los lectores interesados la definición de la salud pública.



Comunidad: conjunto de personas de un pueblo o nación, organizados y con objetivos comunes.

Bienestar: condición de la persona en que funciona psíquica y somáticamente.

Estado: conjunto de los órganos de gobierno de un país soberano.

La salud pública está íntimamente relacionada con el desarrollo social. Condiciones económicas (revolución industrial), políticas (revolución francesa), sociales (surgimiento del proletariado) junto con las científicas (aportadas por la epidemiología) se integran en un ente ejecutor y responsable que es el Estado.



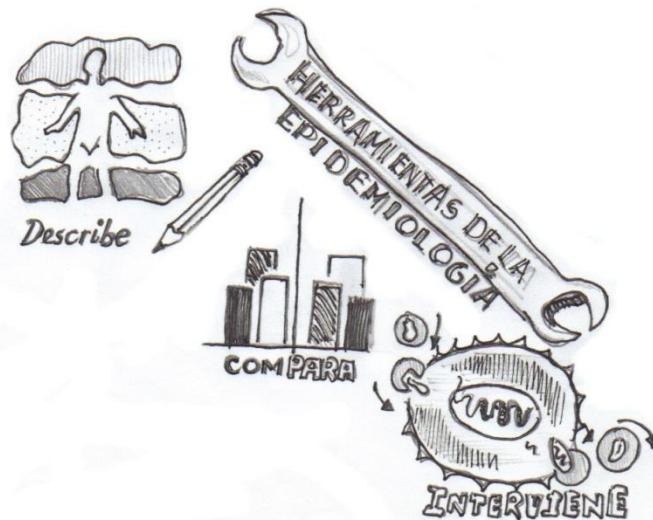
Johann Peter Frank en 1779: fue un extraordinario exponente del papel rector del Estado en la salud de la colectividad. Su obra SISTEMA DE UNA POLICÍA MÉDICA COMPLETA promovía una intervención (con sanciones) del Estado en la salud de los súbditos del imperio Austro-Húngaro. Promovía una policía sanitaria.

Dado que en la epidemiología el elemento esencial de estudio es la población y el ver cómo se comporta en la enfermedad, ofrece con sus análisis importantes aportes para la toma de decisiones, lo que sin duda, hace parte de la salud pública. Si bien la epidemiología entra como un componente básico de la salud pública, la política con sus desarrollos (cualidades y defectos) forman parte del otro componente fundamental de la salud pública.



Hoy ha surgido como disciplina la epidemiología clínica, la cual utiliza el método epidemiológico para el estudio de la enfermedad individual desarrollando conceptos tan interesantes como la medicina basada en la evidencia.

1.3. Las herramientas de la epidemiología



La epidemiología usa la “duda” que surge de observar y/o comparar. Ahí plantea sus problemas y usa (como cualquier otro conocimiento) la descripción como un primer paso para entender, para pasar ulteriormente a la comparación en donde plantea predicciones hipotéticas, pasando finalmente a las intervenciones (o experimentos) de donde obtiene el máximo de sus aplicaciones.

De lo anterior se entiende que las herramientas de la epidemiología son todas aquellas que se derivan de la observación objetiva (de descripciones, comparaciones e intervenciones).

1.3.1. La observación

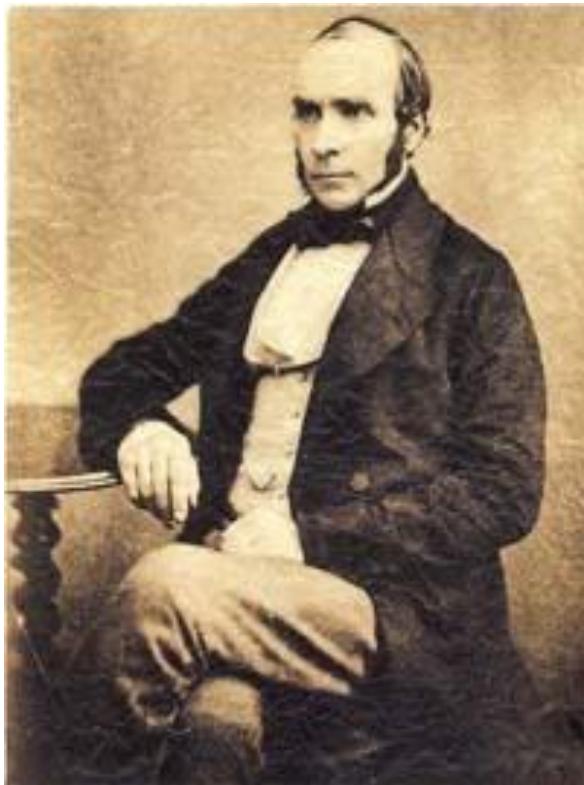
La observación tiene una connotación especial en epidemiología. Sin duda, corresponde al primer paso del método científico y concluye con el **planteamiento del problema**, expresando de esta manera la “concreción” de la observación. En general y, hasta hace muy poco, las observaciones correspondían a manifestaciones que dieron pie a la construcción de enfermedades. Hoy, además, lo hace a los procesos que preservan la salud. Lo cierto es que la epidemiología ha provocado extraordinarios planteamientos de problemas a través de los cuales ha logrado beneficiar a la humanidad cuando los contesta.



Doll y Hill en el British Medical Journal de 1950, son un ejemplo extraordinario de cómo la observación en el crecimiento de la incidencia de cáncer del pulmón dio pie al planteamiento de la relación entre el cigarrillo (cuyo consumo definitivamente se había incrementado) con la enfermedad. Los estudios los plantearon sospechando o que, bien era la contaminación del aire (automóviles, humo de fábricas etc.) o, el incremento del hábito de fumar tabaco. En la fotografía está el Dr. Doll.

1.3.1.1. La descripción

Una vez **planteado el problema**, la epidemiología **describe**, y lo hace para entender. Describir es mencionar lo que se ve usando referentes. Para ello, primero ordena y luego clasifica. Un ejemplo de la extraordinaria capacidad descriptiva que usa la epidemiología son los aportes que ha hecho a las enfermedades a través de lo que se conoce como la **historia natural**. Ahí la epidemiología, usando una serie de criterios, describe a las enfermedades permitiendo entenderlas.



John Snow en su texto "Sobre el modo de transmisión del cólera" en los alrededores de 1850 hace una extraordinaria descripción a partir de la cual podrá en su momento plantear las hipótesis de lo que se puede hacer para controlar la epidemia de cólera que azotó Londres. En primer lugar, Snow identificó el cólera asiático. Observó que el sitio donde primero aparecía era en los puertos. Estaba seguro que el tránsito de las personas de un sitio donde existe a otro donde no, es indispensable. Observó que una vez se presenta un caso, la expansión es inmediata y empieza con los allegados directos de la víctima. De ahí que es válida la sospecha de "transmisión", decía Snow en su escrito de 1789. ¿Cómo se trasmite? Plantea que dada la manifestación gastrointestinal tan "fuerte", el contacto con esa diarrea puede ser la forma. Y menciona con base a unos

"casos" dónde apareció el cólera en personas en contacto con enfermos (familiares), que el período de incubación suele ser muy corto. La posibilidad de que las heces "contaminen" expresa el peligro de convertirse en una epidemia. Enunció que las manos de los que cuidan el enfermo al ponerse en contacto con la ropa de cama "ensuciada" es muy seguramente un factor de contaminación. Además, el hacinamiento se convierte en un factor de propagación, pues cuando un enfermo duerme con los demás miembros de la familia y tiene una profusa evacuación, puede, y de seguro, contamina a muchos miembros de la familia que duermen con él. Esta suposición explica por qué el cólera es tan devastador entre los mineros. A ello le agrega las condiciones sanitarias dentro de la mina que hacen que "un minero" enfermo, propague con facilidad la enfermedad, pues la mezcla de material de letrina con la comida de los demás es muy probable dentro de la mina.

La estadística se ha convertido en una excelente "ayuda" para la epidemiología. Dado que la ciencia busca entender los datos (respuestas de las variables) la estadística ofrece "referentes" que permiten comprender dichas respuestas (Estadística Descriptiva) y ofrece alternativas para comparar (Estadística Analítica). Por ejemplo, el uso de referentes estadísticos (medidas de tendencia central o de dispersión...) contribuye enormemente en el proceso de entender.



La **estadística** es el conocimiento que permite recoger, resumir y analizar datos sujetos a variaciones aleatorias. El término **aleatorio** proviene del azar o que depende de algún evento fortuito. Por ende, la estadística a través de los números pretende hacer entender por qué suceden las cosas.

El ser humano aprende gracias a la habilidad de describir. Al hacerlo busca unos referentes (puntos de comparación) que le permiten incrementar su entendimiento acerca de lo descrito. Por ejemplo, en el estudio de las enfermedades, se usa la mortalidad (virulencia) de una enfermedad para poder saber qué tan grave es. Para ello, y cuando se está describiendo la enfermedad, se "mira" (por ejemplo) el dato de mortalidad: cuántos pacientes han muerto en determinado tiempo. Este dato debe ser "manejado" para que se pueda "entender". Lo que hace la estadística descriptiva es ofrecer unos referentes descriptivos, como la "media" de mortalidad de pacientes en determinado tiempo. Esta "media" va a permitir comparar este comportamiento con el de otras enfermedades y así se puede clasificar (ordenar) esta enfermedad. De hecho se usan dos grupos de medidas de referentes: las de tendencia central y las de dispersión, las que se ahondarán en el capítulo 4.

1.3.1.2. La comparación

		VARIABLE DEPENDIENTE		
		SI	NO	
		PRESENTES	PRESENTE	
VARIABLE INDEPENDIENTE	Si presente	a	b	a+b
	No presente	c	d	c+d
		a+c	b+d	

Tabla de comparación o de contingencia o del 2X2o del SI y el NO

El ser humano entiende mejor cuando compara. Por ello la epidemiología, como pocas ramas del conocimiento, usa una extraordinaria herramienta para la comparación. Se trata de la tabla de contingencia (o **tabla del 2X2**) que permite consolidar a los ojos lo que se compara. Una vez más la estadística contribuye con sus técnicas, sus pruebas, a confirmar si lo que se compara definitivamente es diferente o igual.

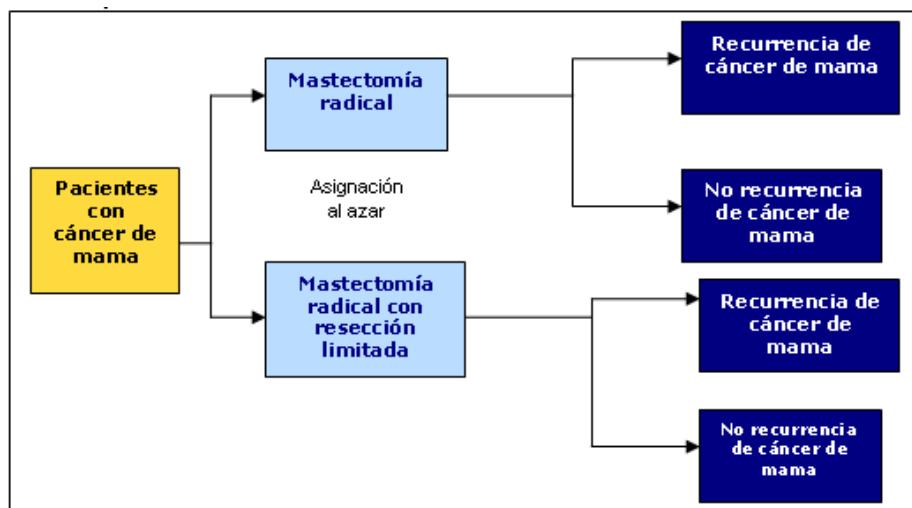


La comparación es una de las formas más usadas para que -el ser humano aprenda. Por ejemplo, al analizar (igual a comparar) las manifestaciones presentes en una enfermedad, se hace indispensable comparar enfermos con sanos y ver quiénes tienen o no la situación o característica a comparar. Esta selección, de hacerse sistemáticamente con "todos" los componentes, permitirá "ver" qué tienen y qué no tienen los enfermos, diferencia esencial para plantear interpretaciones. Para poder confirmar si existen o no diferencias, se usa la estadística. De hecho un área de la estadística está destinada a analizar si las diferencias halladas al comparar son producto del azar o si ciertamente son diferencias que "siempre" están. Lo cierto es que buena parte de este texto va estar dedicado a enseñar a usar e interpretar esta tabla (la de contingencia).

1.3.1.3. La intervención

El objeto final de la epidemiología es entender y con ello predecir qué hacer para evitar que lo nocivo ocurra. Esa predicción se construye a través de las intervenciones. **Intervenir** es hacer que la población sea sometida a una “acción” puesta de manera consciente, esperando que a través de dicha acción no ocurra (prevención) o se corrija una adversidad (disminuya la morbilidad). En esencia, las intervenciones se convierten en la máxima aplicación de la epidemiología.

Las intervenciones son la razón esencial de la medicina: se busca que con lo que el médico formula (variable interviniente) modifique el pronóstico (mueran menos) o las manifestaciones (se curen más rápido). Implica conocer la historia natural de las enfermedades en torno al pronóstico (intervenciones cuando está presente la enfermedad) o pretenden disminuir su incidencia (intervenciones preventivas).



Por ejemplo, en el esquema previo se muestra cómo se debe comparar: se está comparando una modalidad de mastectomía vs. otra y así se puede “aprender” cuál recomendar (comparación de intervenciones). Claro... para poder evaluar una intervención siempre debe haber un grupo de población que es intervenido y otro que no (o que es intervenido de otra manera). Así se compara. Las intervenciones pueden ser efectuadas con medicamentos, cirugías, procesos educativos, vacunas, psicoterapias. Lo que se busca es que al “darlas” se modifique la evolución de la enfermedad. Un ejemplo es la apendicitis. Durante miles de años el pronóstico de esta enfermedad fue malo. A quienes les daba, se morían en un altísimo porcentaje. Se la conocía como la “enfermedad del costado derecho”. Las intervenciones que se hacían no modificaban el fatídico desenlace. Cuando se decide intervenir operando y extirmando el apéndice, se cambia radicalmente el pronóstico: la mortalidad (dato) se disminuye al ser comparados los casos operados con los no operados.

1.4. Las ayudas en la epidemiología



La epidemiología usa el paradigma científico. De él construye su método. Maneja las medidas buscando que finalmente ellas puedan generar datos que permitan ser tratados por la estadística, ya para describir o para comparar.

1.4.1. El raciocinio científico

Sin petulancia, la epidemiología hace uso del raciocinio científico como su modelo de pensamiento. Y se menciona sin petulancia porque existen otros modelos o formas de pensar. Sin embargo, es válido afirmar que la epidemiología no existiría sin el raciocinio científico con su expresión, que es el método científico.

Partiendo de una observación, se plantea un problema. Luego se describe, recurriendo a la indagación de lo referido acerca de cada componente hasta el momento, para así proponer una serie de **hipótesis**. Con base en ellas se selecciona aquella que parece poseer más probabilidades y se aplica. Luego se correlaciona con los resultados y bien se aprende cómo se “hace” o cómo “no se hace”, caso en el cual se continúa con la siguiente probable respuesta.



Hipótesis: suposición de algo posible o imposible para sacar de ello una consecuencia, la que se establece provisionalmente como base de una investigación que puede confirmar o negar la validez de aquella.

1.4.2. El Papel de la estadística

El ser humano ha construido un idioma que pretende ser universal y que utiliza símbolos en donde el significado de los mismos sea siempre el mismo. Se trata del lenguaje de los números. Los investigadores buscan transformar las palabras de sus respuestas investigadas a números. La estadística da referentes descriptivos a esos números que contribuyen a "entenderlos" y permite comparar sus resultados. Dado que usa referentes para la descripción y pruebas para comparar lo encontrado, una forma de dividirla es precisamente como estadística descriptiva y analítica (o inferencial). La epidemiología siempre describe y de ahí la inmensa utilidad que tiene la estadística descriptiva. Así mismo, muchos de los planteamientos de la investigación surgen de comparar. La certeza de la similitud o de la diferencia se convierte en un requisito esencial para la aplicación de los resultados. Por ello, la estadística comparativa, también denominada analítica o inferencial se convierte en esencial.



Ver en qué nos parecemos y en qué nos diferenciamos de otra persona no parece ser una tarea difícil. Se busca describir unas características presentes en uno y ver si el otro las presenta. Pero hacer esas comparaciones entre grupos poblacionales, si bien es posible, requiere usar "criterios descriptivos" que al final permitan asegurarnos en qué se parecen y en qué se diferencian estas poblaciones. Eso es lo que hace la estadística: permite usar criterios descriptibles que podrán ser comparados. Por ejemplo, al describir EDAD. La edad se reconoce como **CATEGORÍA**. Pero se hace indispensable (según el criterio del investigador) subdividirla en unas **VARIABLES**, por ejemplo, de 0 a 5 años de 6 a 10 años, de 11 a 19 años, etc. El número de casos encontrado en cada variable se conoce como **DATOS** y el porcentaje derivado de ellos cuando se usa para comparaciones con otros trabajos u otras investigaciones, se reconoce como **INDICADOR**.

Datos

Categoría

Variables

Indicadores

Tabla # 1 CARACTERISTICAS DE LA POBLACIÓN Y LAS FAMILIAS DE LOS ADOLESCENTES DE 12 COLEGIOS PÚBLICOS DE LA LOCALIDAD DE SUBA EN EL AÑO 2006.		
EDADES	Nº	%
De 10 A 12 años	3965	32
De 13 A 16 años	7200	59
De 17 A 19 años	597	8
GENERO	Nº	%
Mujeres	6274	51
ESTRATO	Nº	%
Estrato 1	347	3
Estrato 2	8958	73
Estrato 3	2475	20
ESTRUCTURA FAMILIAR	Nº	%
Nuclear completa	6862	56
Nuclear incompleta	3774	31
Recompuestas	1112	9
Hogares donde faltan ambos padres	227	2
FUNCIONALIDAD FAMILIAR	Nº	%
Buena función familiar	7722	63
Disfunción moderada	2731	22
Disfunción severa	920	7



¡qué vaina!

Uno de los problemas que existe en la humanidad es que no nos hemos puesto de acuerdo en torno a la manera de ordenar, de clasificar... es más, ni siquiera en muchas definiciones. Por ello, los términos que acá empleamos, pueden tener una interpretación distinta en otros textos...

1.5. La estrategia de la epidemiología

La epidemiología se desarrolla mediante el uso de la investigación. Usa una variedad de métodos, en los cuales pone condiciones específicas.

1.5.1. La investigación



Investigar es contestar preguntas usando un modelo. Modelo es una forma de pensar o de hacer las cosas. La epidemiología usa el pensamiento científico. Cuando surge la observación y se expresa en la pregunta, la epidemiología ayuda a escoger la “mejor” forma de responderla para el interesado. Existen muchas maneras de contestar la pregunta (muchos tipos de investigación) unos más **válidos** que otros pero todos útiles. El que se escoja depende de las posibilidades de los investigadores. La manera como se contesta la pregunta de investigación en epidemiología se conoce como DISEÑOS. Sea este el momento de decir que la epidemiología usa unos diseños específicos denominados diseños CUANTITATIVOS.



Por ejemplo, se quiere saber si un medicamento "cura" una enfermedad. Una forma de saberlo sería ver qué enfermos han recibido el medicamento y cuántos se han curado. Otra, sería dividir a los enfermos en dos grupos: a unos darle el medicamento y a otros no y ver en cuál de los dos grupos se curan más. Son dos diseños que pretenden averiguar lo mismo. Pero uno de ellos resulta **más válido**, es decir, más probable que de seguirlo, la respuesta sea más coincidente con la realidad... ¿cierto? Más adelante se volverá a tratar el asunto (Capítulo3, paso6). *Válido* es entonces la aproximación con la realidad... *más valido* es más cerca de la realidad.

Sea cual sea el diseño, la investigación epidemiológica descompone la pregunta en otras, denominadas variables. Una **variable** puede definirse como una característica a través de la cual se puede contestar la pregunta de investigación (o parte de ella).



Validez: Concordancia entre un resultado y la realidad. Válido es aquel resultado que refleja la realidad.

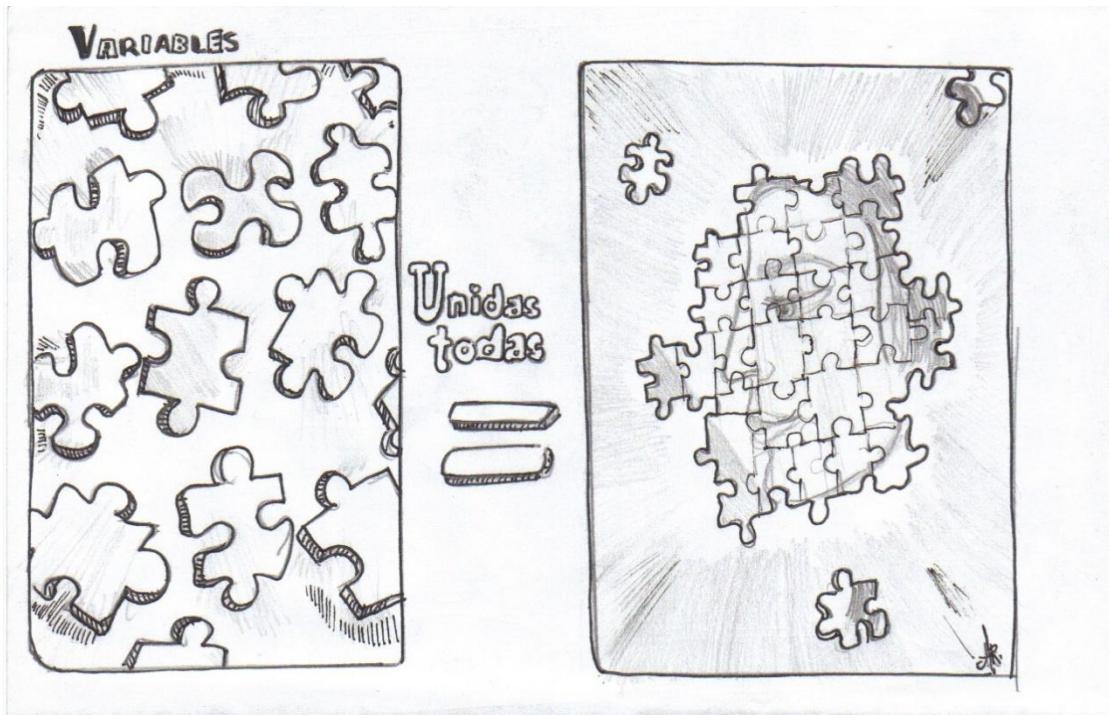
1.5.2. Las variables

La pregunta de investigación se "transforma" en variables que no son otra cosa que "otras" preguntas a través de cuyas respuestas (**datos**) se contesta la pregunta. ¡Lástima que las preguntas de investigación no se contestaran de manera simple (una pregunta una respuesta)! Por ejemplo, si se desea saber si un medicamento "cura" una enfermedad mortal, la variable mortalidad ha de ser **medida** (de seguro es la variable principal). Pero así mismo, la **morbilidad**, pues "cura" seguramente implica que no queda enfermo (claro... esto se aclara previamente en la definición de términos). ¡Ah...! la pregunta de morbilidad (entendida como enfermedad) muy seguramente deberá ser traducida con otras variables donde se expresa la presencia o ausencia de manifestaciones (que son la expresión de la enfermedad). Presencia de fiebre o de dolor o de leucocitosis, por ejemplo, son variables de morbilidad. Además, se deberá medir características del grupo intervenido versus el no intervenido (otras variables). En fin... la respuesta a "esa" pregunta, implica buscar variables (características) para contestarla. Como se observa en el ejemplo, y lo que constituye el mensaje que se quiere trasmitir, es que una pregunta de investigación se contesta a través de muchas otras denominadas variables. Lo cierto es que la epidemiología ofrece todo su potencial si el que quiere contestar la pregunta "selecciona" unas variables adecuadas.



Medir: comparar una cantidad con su unidad.

Morbilidad: proporción de personas que enferman.



 Una variable sola rara vez permite describir a través del dato lo que se quiere "ver". Habitualmente, es la suma de varias variables con sus respectivos datos los que, unidos permiten "ver" lo que se busca. Por ejemplo, si se quiere saber si un servicio médico es "bueno" se deberán medir muchas variables, tales como grado de recuperación de la salud de la población enferma, percepción de la población, costos en que se incurre, y claro, en este caso, se deberán construir "sub" variables de cada aspecto, que permitan calificar a través de los datos. Tal vez un ejemplo más sencillo (pero dejando el anterior para discusión) es cómo saber si un enfermo se está recuperando con unas medidas terapéuticas. Las variables que se deberán medir son aquellas que la enfermedad había afectado. Por ejemplo: los signos vitales y las otras manifestaciones. Sumando todos los datos que se obtienen de cada variable se podrá decir que sí o no se recuperó.

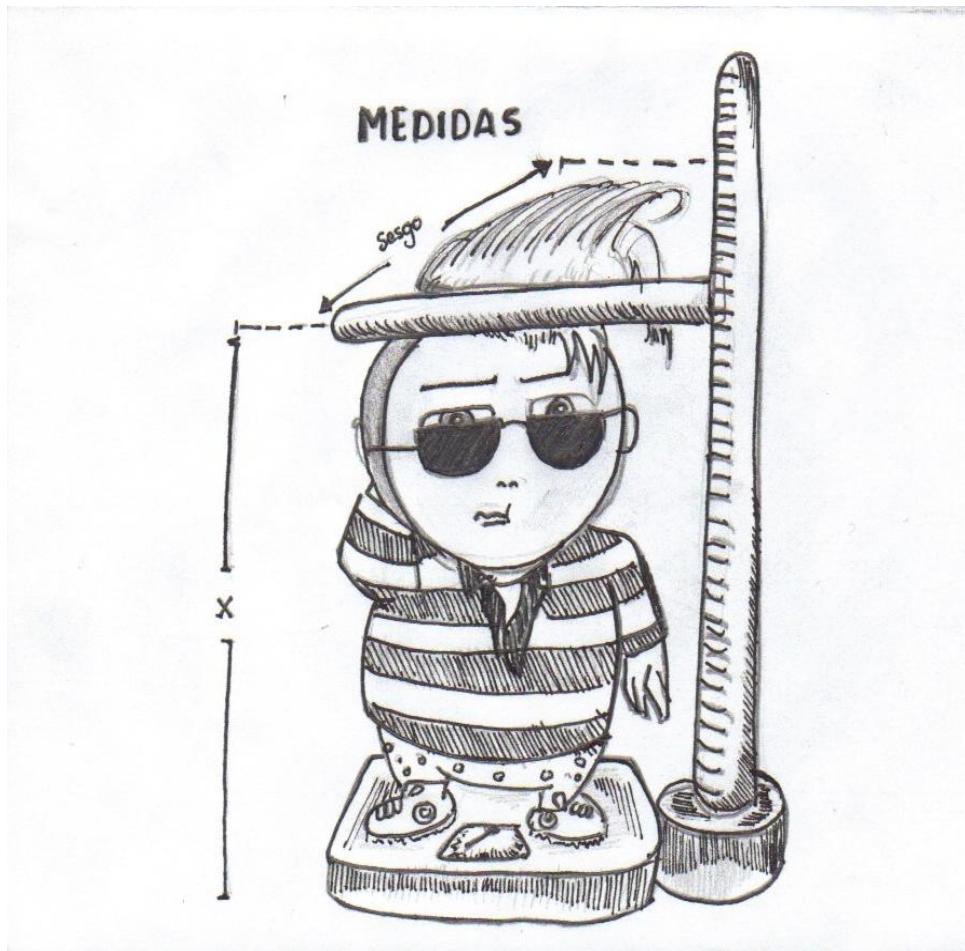
Repetiendo: la respuesta de las variables se denomina datos y estos se traducen en últimas, en las "medidas" que sirven para comparar y deducir la respuesta.

En torno a las variables existen muchas clasificaciones y definiciones, todas ellas efectuadas bajo distintos contextos. Ejemplos de clasificaciones son las variables dependiente e independiente, usadas particularmente cuando se comparan características. En algunos textos, clasifican a las variables con los atributos de los datos que producen (nominales, ordinales, continuas y discretas). En este texto esa clasificación se le otorga a los datos.



El ser humano crea clasificaciones para entender. El asunto grave se produce cuando ante unos mismos eventos o criterios, se construyen distintas clasificaciones. Ahí ponerse de acuerdo es muy difícil. Desafortunadamente esto ocurre con más frecuencia de la que se debería. **Las VARIABLES** son una buena muestra de múltiples clasificaciones. Por ejemplo, para el uso de la tabla de contingencia que se revisará a lo largo del texto, la clasificación como variable **DEPENDIENTE** (que es la que "siempre" se tiene en cuenta) e **INDEPENDIENTE** (que son las que se van "cruzando" con la dependiente) resulta clave. Pero al hacer análisis de investigaciones surge otra clasificación muy interesante: variable **INTERVINIENTE**, referida a aquellas características que "intervienen" de alguna forma en "algo" que se está estudiando. Lo interesante de todo esto es que se debe tener una mente amplia y dispuesta a no "encasillarse". Por ejemplo, si se desea estudiar factores que se relacionen con la aparición de una enfermedad, esta última es la variable dependiente y los factores que se estudien son la variable independiente. Un ejemplo de variable interviniante puede ser la relación entre cáncer de pulmón y café en una investigación: pero resulta que los que toman café presentan una condición: es que fuman más. Por ende, esa relación entre café y cáncer es el resultado de una variable interviniante de los que toman café, y es que fuman. ¡Ah...! pero en algunos textos mencionan que la variable interviniante es aquella que "pone" el investigador en los diseños experimentales. Por ello, es tan importante definir términos antes de discutir.

1.5.3. Los datos y las medidas



Toda pregunta de investigación se contesta a través de variables (otras preguntas). La respuesta de cada variable se denomina **dato**. Un buen dato debe tener referentes que permitan compararlo y, por ende, "medir" su resultado. Las variables pueden producir datos cuantitativos (continuos y discretos) y cualitativos (nominales, ordinales, interválicos y proporcionales). Los primeros usan números como referentes. Los segundos, características. El desarrollo estadístico se ha encargado de describir y comparar unos y otros.

Una variable no es otro elemento que una característica que será medida a través de los datos. Entonces, y bajo el contexto expresado en los párrafos previos, las preguntas de investigación se contestan a

través de los datos obtenidos al “medir” determinadas características (o variables)”. Cuando la pregunta se puede contestar con variables que producen datos “fáciles de entender”, la posibilidad del resultado se potencializa. Por ejemplo, si la pregunta es ¿cuál es el estado de nutrición de una población de niños?, la variable puede ser el peso o la talla de esos niños y los datos serán o centímetros o gramos, datos que ciertamente se aproximan a reflejar el estado nutricional. Sea este el momento de resaltar que un buen dato ha de cumplir con las características de: ser veraz (reflejar la verdad), pertinente (que refleje la importancia de lo que expresa), entendible (que el que lo “vea” lo entienda) y que sea fácil de obtener.



Las variables que usan datos cuantitativos son “más tratables” que aquellas que usan datos cualitativos. Más tratables se refiere a que la estadística descriptiva se “acomoda” mejor y, por ende, la obtención de referentes que permitan entenderla más. Así, por ejemplo, ante la variable talla de una persona, el dato en metros o centímetros es más fácil de entender (y por tanto de comparar) que el dato de disfunción severa o moderada en la aplicación del test de APGAR FAMILIAR (variable). Cuando se habla del tamaño en centímetros todos lo entendemos. Cuando se habla de disfunción familiar cada uno puede imaginar algo “un poco” diferente... ¡que vaina ser tan positivista, pero así son las cosas!

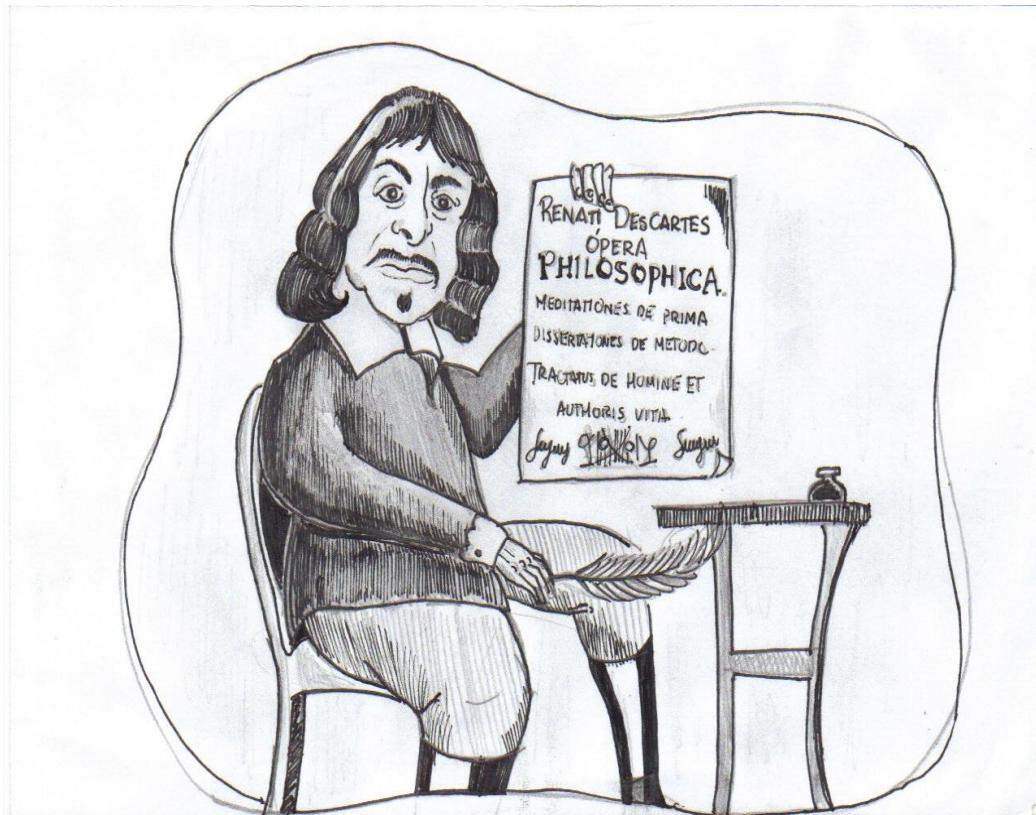
1.5.4. La búsqueda de la validez para la salud

Una “buena” variable debe producir datos que se entiendan y ello se consigue a través de la cercanía con la realidad de su respuesta (validez). Esto se conoce como el proceso de **validación** que no es otra cosa que la correspondencia entre lo que se pregunta y lo que responde (dato), con la realidad. Entre mayor sea la correspondencia, más **válida** será la variable y, por consiguiente, más útil.



La selección de variables implica un gran esfuerzo racional, sistemático y coherente. De hecho, lo que se busca es que las variables sean un fiel reflejo de la realidad. Por ello, se ha colocado a Descartes como ejemplo de la validez, ya que su “racionalismo” es un llamado a la búsqueda de la validez.

Proceso de validación: acciones a través de las cuales se pretende demostrar cómo “algo” refleja la realidad.



Ante una pregunta de investigación, será esencial que se seleccionen variables válidas. Para ello, nada mejor que revisar cómo han hecho otros investigadores. De dicha revisión pueden surgir variables ya "demostradas" o, lo que es lo mismo, VALIDADAS, lo que sin duda facilita la investigación.

La validez es la cercanía del método investigativo con la realidad y, si ello se cumple, hace probable que los resultados puedan **inferirse** y con ello predecir, que al fin de cuentas es una de las búsquedas más importantes del ser humano.



Inferir: generalizar una conclusión resultado de una investigación particular. Por ejemplo, y en torno a las investigaciones acerca de medicamentos en una enfermedad, inferir sería concluir que si en la población objeto de la investigación fue útil, también lo será en "un" enfermo (de lo general se particulariza). En el capítulo 3 subtítulo 3.2 se ampliarán estos conceptos.



Capítulo 2

La Utilidad de la Epidemiología



<u>2.1. La lucha contra la enfermedad</u>	33
2.1.1. El arte de describir	33
2.1.1.1. La historia natural de las enfermedades.....	37
2.1.1.1.1. Ejemplo: Historia de una epidemia (SIDA).....	45
2.1.1.2. La cadena epidemiológica	48
2.1.2. La tabla de frecuencias.....	51
2.1.3. La forma de aprender: comparar	53
2.1.4. La tabla de comparación: los factores de riesgo y la evaluación de lo que sirve	54
2.1.4.1. Usos de la tabla de comparación.....	58
2.1.4.1.1. Identificación de riesgos	59
2.1.4.1.1.1. Presentación de una tabla de contingencia	64
2.1.4.1.1.2. Otras comparaciones: sensibilidad y especificidad	65
2.2. Buscando las causas	69
2.2.1 La maravilla de las relaciones y las asociaciones.....	70
2.2.2. La red causal	78
2.2.2.1. Las causas necesarias y suficientes	79
2.3. Buscando la salud pública: La vigilancia epidemiológica:	83
2.3.1. La importancia de los registros.....	84
2.3.2. El arte de la predicción	85
2.3.3. Aporte de la epidemiología a la salud comunitaria.....	87
2.3.3.1. La estrategia de la identificación	90
2.3.3.2. Para decidir.....	91
2.3.3.3. Para organizar.....	91
2.4 Hacia dónde va la epidemiología. Mirando la evidencia de la utilidad.....	93
2.4.1. Una mención de éxitos y limitaciones	96
Usos de la epidemiología	103



CAPÍTULO 2: LA UTILIDAD DE LA EPIDEMIOLOGÍA

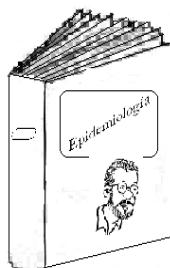
La epidemiología en esencia son conocimientos que se emplean para combatir la enfermedad (ayer) y buscar la salud (hoy) de la población. Para ello, descubrir las causas y evaluar intervenciones se convierte en esencial. De la misma manera, el monitoreo de los comportamientos de las enfermedades se convierte en una de las principales herramientas para vigilar y proteger a la población. A continuación se expondrá algunos de los usos de la epidemiología.

(Este capítulo es una extensión del subtítulo 1.2.)

2.1. La lucha contra la enfermedad

2.1.1. El arte de describir

¿Cómo combatir lo que no se conoce?... Es muy difícil; a veces y, por el azar, tal vez parezca (o incluso suceda) que lo que se hace, da resultados, pero lo cierto es que lo más frecuente es que todo lo que se haga no influye de manera real sobre el curso de los hechos. Esa fue por mucho tiempo la historia de la humanidad en torno a sus combates contra la enfermedad. Dado que no la reconocían, sus acciones para combatirla eran inútiles. Pero surge en algunas mentes la capacidad de observar. De esas observaciones empiezan a describir las manifestaciones y al hacerlo, son capaces de colocarles nombre a las entidades. Ahí ocurrió el primer paso. Al tener nombre, las manifestaciones pudieron ser reconocidas y luego comparadas. Así las cosas, es factible analizar si el hacer o no hacer en una entidad provoca diferencia. Se había conseguido a través del nombre (de la enfermedad) un **referente** y ese referente (basado sin duda en lo más frecuente) permitió comparar.



Referente: punto o aspecto de reparo que sirve para comparar otro. La gráfica muestra el mapa que diseñó el Dr. Snow en la epidemia de cólera que le valió reconocimiento mundial. Cada punto señala dónde vivían las personas que tenían cólera. A él se le ocurrió relacionar los casos con la zona geográfica donde vivían. Gracias a la descripción pudo notar que existía una concentración de casos en una zona. Con ello pudo formular una hipótesis que le permitió tomar una medida de tratamiento. Sea este el momento de mencionar que ese mapa se conoce como CROQUIS y es una importante forma de presentar información.

Describir es decir lo que se ve en contraste con un referente; ello es una de las formas como el ser humano puede entender. La primera gran descripción que se hizo fue la de las manifestaciones agrupadas como enfermedades (construidas por el hombre), claro, en los enfermos (quienes padecen las manifestaciones). Luego de esto la tarea ya era más fácil... Se podían clasificar a los enfermos (y por tanto, la enfermedad) por las manifestaciones de acuerdo al órgano correspondiente (enfermedades del corazón o del pulmón, por ejemplo)

o por la causa (o lo que se creía antes y que hoy se sabe) que provoca las manifestaciones (y, así mismo, la enfermedad), Por ejemplo, enfermedades infecciosas o entidades tumorales. Sin duda un elemento cardinal en la descripción fue el ordenamiento. **Ordenar** es un componente esencial de la descripción. Se ordena siguiendo criterios, que deben, en lo posible, ser simples y aplicables. Solo así se logran conseguir que otros describan y al hacerlo la capacidad de entender a la enfermedad se potencializa. La descripción permite reconocer.



Ordenar: colocar de acuerdo a un plan o un modo conveniente. Resulta evidente que cuando se ordena se incrementa la capacidad de ver...

Uno de los grandes aportes de la epidemiología ha sido utilizar como criterios de descripción, la **historia natural de las enfermedades**. Gracias a ello se puede entender a la enfermedad y cómo intervenir sobre ella.



Son muchos los médicos a lo largo de la historia de la humanidad que con la publicación de sus observaciones han contribuido a construir la historia natural de las enfermedades. Resulta práctico que al describir se sigan unos criterios. La historia natural usa de hecho unos, así se puede comparar mejor cuando se publica. Publicar es una de las actividades más importantes que se debe cultivar en los médicos. A través de ello es mucho lo que se puede aprender.

Para describir, la epidemiología propone usar unos criterios a través de los cuales expone el concepto de historia natural de las enfermedades; como herramienta descriptora propone a la **tabla de frecuencias** y cuando se requiera, propone el uso de la **tabla de contingencia** como herramienta esencial de la comparación de lo descrito.

La tabla de frecuencia es la principal forma de presentar los datos en las investigaciones. Está formada por tres columnas básicas: la primera donde están los criterios a describir, la segunda donde están las frecuencias y la tercera en donde está el porcentaje de cada frecuencia. En el capítulo 4 numeral 4.2 se puede leer ampliada la manera cómo se construye una tabla de frecuencia con todo.

TIPO DE FAMILIA	Nº	%
Familia Nuclear		
Completa	7 619	62
Incompleta	5 386	39
Extensa	2 854	21
Familia reconstituida		
Con padrastro	1 065	8
Con madrastra	171	1
Bajo el mismo techo vive con		
Hermanos	10 188	85
Abuelos	1 595	12
Tíos y primos	786	6
Primos	207	2
Cuñada	60	0.4
Ciclo Vital Familiar		
Familia con hijo mayor adolescente	5 161	38
Función Familiar		
Buena función	8 590	68
Disfunción moderada	3 029	24
Disfunción severa	998	8

Tabla de frecuencia que expresa cómo son las familias de 13 699 adolescentes de 12 colegios públicos de la localidad de Suba en el año 2007. Tiene tres columnas: la primera muestra los criterios de familia que se están presentando, la segunda el número y la tercera el porcentaje. Estas son las tres columnas básicas.

La tabla de contingencia (o tabla del si y el no) se basa en las cuatro casillas de comparación que surgen de poner los que SÍ tienen la variable dependiente con los que NO la tienen, contrastados con los que SÍ tienen la variable independiente junto con los que NO la tienen. Más adelante se aclaran estas dos tablas a través de ejemplos (subtítulos 2.1.3. y 2.1.4.).

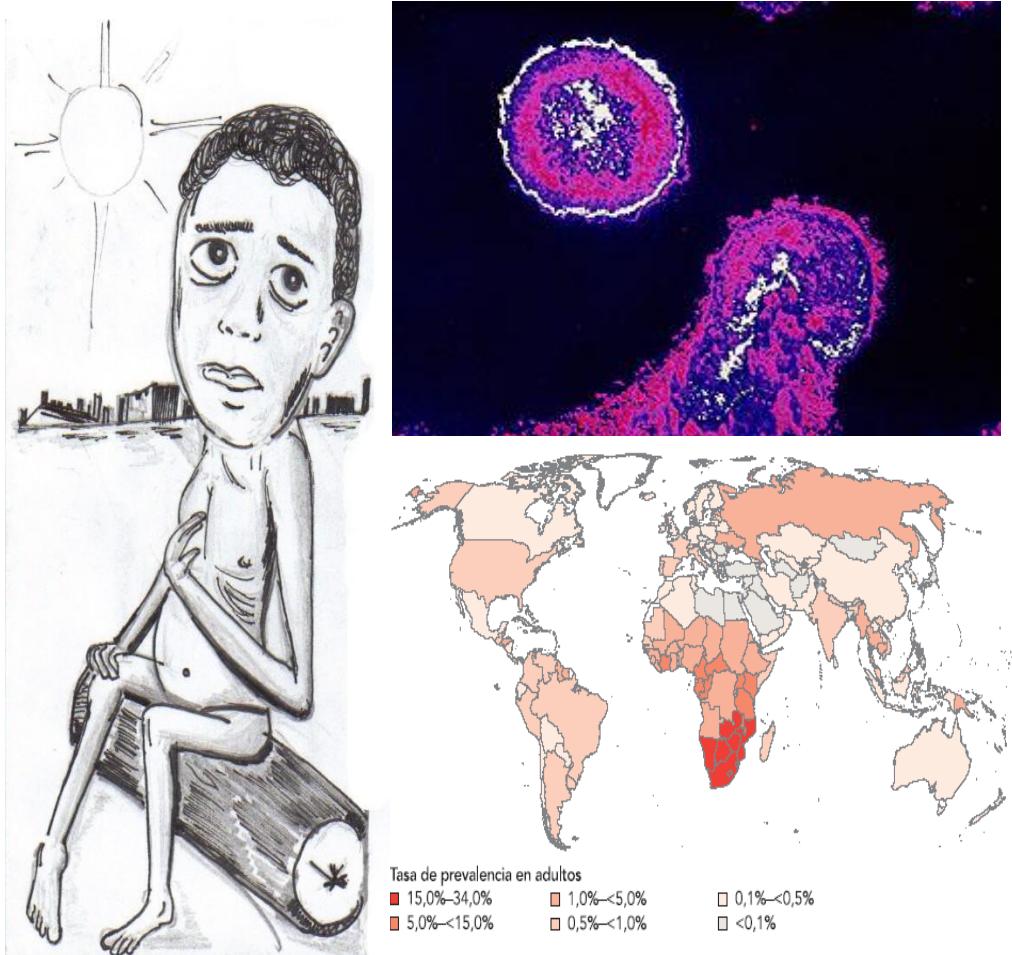
**VARIABLE
DEPENDIENTE =
CANCER DE PULMON**

VARIABLE INDEPENDIENTE = FUMAR		SI	NO	1269
		PRESENTES	PRESENTES	
	Si FUMA	647	622	
	No FUMA	2	27	29
		649	649	1298

Tabla de contingencia

2.1.1.1. La historia natural de las enfermedades

Como criterios de descripción en torno a las enfermedades, la epidemiología busca considerar los tres componentes de seguro más influyentes. Ellos son: El agente(s), el huésped y el medio. El agente busca estudiar las características causales que originan la enfermedad. El huésped, las condiciones del paciente que intervienen en la enfermedad y, el medio, la influencia que el entorno puede tener. A manera de ejemplo, no es lo mismo una infección por el virus del VIH en un joven que en un viejo o que se presente en una persona de estrato 1 o de estrato 6. Cada uno de estos aspectos influye incluso de manera determinante en el desarrollo de la enfermedad.

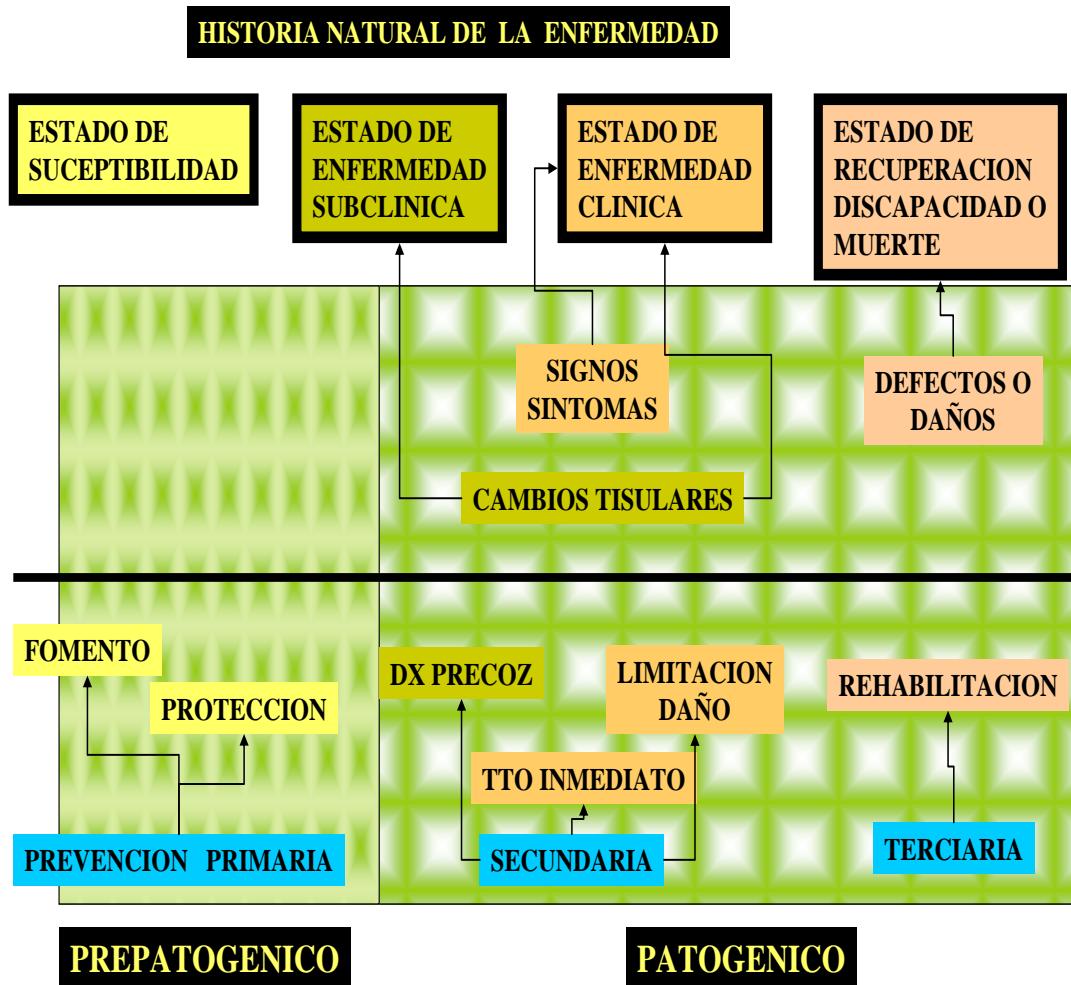


La tríada ecológica (agente, huésped y ambiente) recibe su nombre para hacer entender el concepto que la ruptura del equilibrio entre estos tres aspectos hace que aparezca la enfermedad. El surgimiento de agentes nuevos y, por consiguiente, desconocidos para el sistema inmune (huésped) y bajo condiciones ambientales especiales (pobreza) crea las características que permiten el surgimiento de enfermedades que se expanden en la comunidad.

Blum menciona que al estudiar la historia natural de las enfermedades se han de tener en cuenta las condiciones biológicas (donde la genética, por ejemplo, tiene una enorme importancia), las sociales (distribución de recursos, etc.), las culturales (hábitos y maneras de comportarse) y la organización sanitaria (acceso a los servicios). Por ejemplo, al estudiar la hipertensión, la raza tiene una influencia (más susceptible la población de color) o la situación de pobreza como propagadora de enfermedades infecciosas o el daño que está provocando el sedentarismo a nivel mundial o la diferencia que puede existir en torno al acceso a vacunación de un sistema de salud de un país con otro.

Como se nota, es difícil que algo no pueda ser colocado dentro de algunas de estas categorías. Teniendo en cuenta el fin de la epidemiología (estudiar la enfermedad en la población) se entiende cómo esta triada busca ser lo más abarcadora posible. Con esos tres aspectos de referente, la epidemiología propone para estudiar a las enfermedades estas fases:

1. Fase prepatogénica: acá se estudia a la enfermedad con sus referentes antes de que se enferme el huésped. En este momento se estudia las características que pueden influir en la enfermedad en la interacción de esta triada. Supone esta fase, el que exista un equilibrio que evita que aparezca la enfermedad y cuando éste se rompe, entonces se crean las condiciones para que el ser humano se enferme.
2. Fase patogénica. Acá se estudia luego de que el agente de la enfermedad ya comprometió al huésped. Puede tener:
 - 2.1. Una fase subclínica donde se está enfermo y no se sabe (solo pruebas de laboratorio podrían detectarlo).
 - 2.2. Una fase clínica, donde ya a través de las manifestaciones, la enfermedad hace su aparición.
 - 2.3. Fase de desenlace. Donde aparecen la recuperación con o sin secuelas y/o la muerte, siendo por ende, el momento donde se construye el pronóstico de las enfermedades.



¿Por qué es interesante esta clasificación? Entre otras, sencillamente por qué a través de ella se puede identificar el tipo de intervenciones que puede desarrollar el ser humano para enfrentar a la enfermedad. En la fase pre-patogénica se pueden hacer **intervenciones preventivas** (antes de que aparezca la enfermedad). En la fase patogénica, las intervenciones ya implican combatir directamente la enfermedad. Y en la fase de desenlace se hacen intervenciones en torno a la rehabilitación. El concepto de intervención tiene bajo este contexto a las acciones que se hacen (vacunas, medicamentos o fisioterapia, por ejemplo) para modificar el desenlace de las enfermedades.



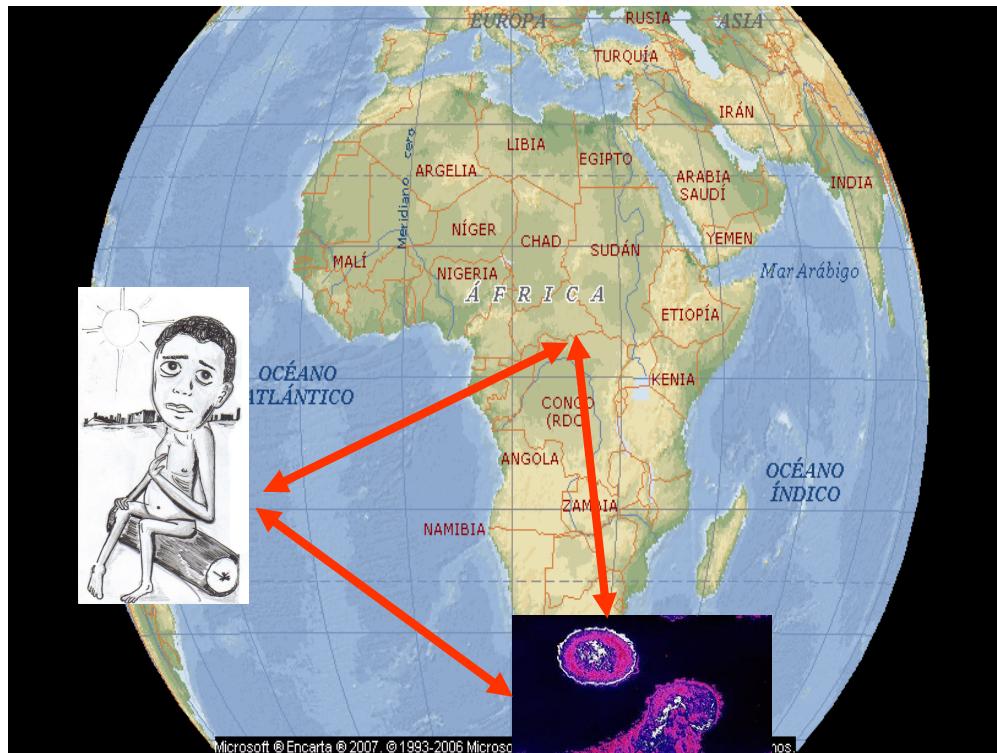
Bajo el contexto de la epidemiología, las intervenciones son todas las acciones en donde se introduce una variable que busca modificar un desenlace. Corresponde también al término de experimentación. El dibujo muestra la percepción de un pintor acerca de las epidemias.

Numerosos autores se han referido al interesante concepto de la historia natural de las enfermedades. En comunidad, se puede hablar de la historia de la vida cuyo objetivo esencial (según Darwin) es la lucha por la existencia... Los seres vivos tienen un ciclo de cumplimiento: nacen, se reproducen y mueren. ¿Cómo nace la vida? Existen teorías que se orientan hacia el creacionismo o hacia el evolucionismo.

La primera, pone la mano de Dios como la gran gestora. El evolucionismo a su vez plantea aspectos tan interesantes como el movimiento cósmico en donde elementos básicos tal como el hidrógeno se mezcla por fuerzas físicas que le impelen al movimiento y va formando nuevos elementos o moléculas. En algún momento se formó el carbono y cuando ello sucede surge la base química de la vida como la conocemos. De ahí a las proteínas y a las célula solo de requiere de más movimiento dentro de esa sopa cósmica. La otra teoría es la del BIG BANG. En ella una enorme y grandiosa explosión cósmica impulsa el polvo cósmico a través del universo y en ese viaje se mezclan partículas de un sitio con otro y en esa mezcla se origina la vida.



¿Qué hacen los seres humanos? En algunos casos pelean por los alimentos contra los agentes... eso se denomina ANTIBIOSIS, en otras se adaptan contra esos enemigos dando lugar a la SIMBIOSIS (pasando por las fases del parasitismo y comensalismo), etapas todas que solo demuestran el fenómeno de la adaptación de la vida. La enfermedad surge como un reflejo de esa lucha...

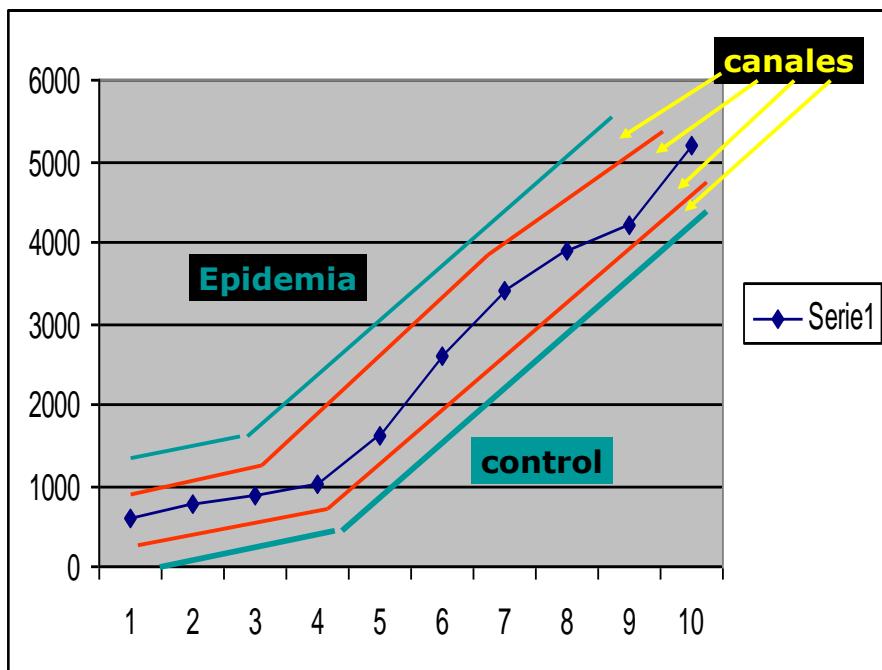


👉 Un buen ejemplo del rompimiento de la triada ecológica es el caso del SIDA. Surge en un ambiente propicio por la pobreza (con todas sus consecuencias), a través de un organismo nuevo (dado que no se conoce, no existe respuesta inmune) y por ende, muy agresivo (virulento) y afecta a personas en condiciones desventajosas.

¿Qué sucede en una comunidad cuando se rompe la triada ecológica? En torno a las frecuencias, se incrementa el número de casos de la enfermedad. Esto se conoce con el nombre de **EPIDEMIA**. Si ese incremento de casos se hace sin sobrepasar unos límites antes establecidos se denomina **BROTE**, pero si el incremento es más de lo esperado en una región o país, se denomina **ENDEMIA** y cuando sobrepasa las fronteras o se esparce entonces el término es **PANDEMIA**. En el caso del SIDA, cuando la humanidad lo reconoció ya estaba en fase de PANDEMIA. Para asegurar estos conceptos, se puede mencionar que la MALARIA es ENDEMICA de unas regiones de Colombia. Que cuando ocurre una intoxicación en una cafetería, ocurre un **BROTE**.

La forma técnica de identificación de una epidemia sigue los siguientes pasos: en primer lugar se debe identificar una frecuencia esperada. Para ello se identifica una medida de frecuencia Por ejemplo, número de casos u otra tal como promedio de la prevalencia en determinado

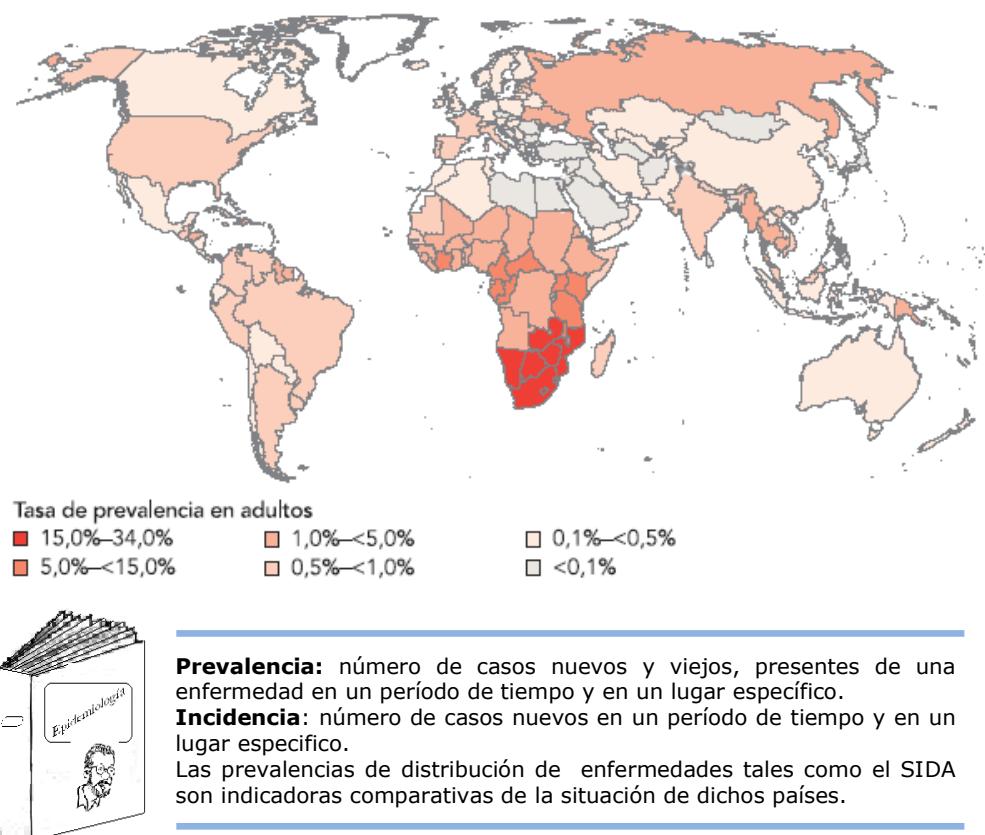
tiempo (número de casos sobre el total de población en cada año). Claro... entre más tiempo se tenga de referente, mejor, pues la información a obtener será más proyectada. Entonces, y con base al número de casos o a los promedios en los períodos de tiempo, se obtiene una **TENDENCIA**. La tendencia puede quedar limitada por los datos que se obtienen de aplicar la medida de la desviación estándar (una DS y dos DS). Ello dará pie a la aparición de unos canales... pero obsérvese la figura:



En la gráfica se observa en el eje X numerado con 1, 2 etc.... el tiempo de observación (por ejemplo 1 = 1984. 2 = 1985 y así sucesivamente). En el eje Y se observa el número de casos que se detectaron en cada uno de los años. Desde ya (y al graficar) se puede identificar cómo se está ante una enfermedad que tiene una tendencia al incremento (claro... medida a través de número de casos) pues es evidente su crecimiento en cuanto a número de casos. La línea azul con puntos es el promedio de número de casos. Con la línea roja se establece unos límites establecidos al aplicar una desviación estándar (concepto que se ampliará adelante). Con la línea verde se establece los límites con dos desviaciones estándar. Entre la línea azul y la roja se habla del canal NORMOENDÉMICO indicando que las frecuencias que se ubiquen entre esos límites son las esperadas. Las frecuencias que se ubiquen entre la línea roja y la verde describen al canal hiper (si es hacia arriba, situación que se podría presentar en un BROTE por ejemplo) o hipo endémico (hacia abajo), y si se presentan frecuencias por arriba de del límite verde superior se habla de EPIDEMIA y por debajo del mismo, de enfermedad controlada. La utilidad de estos canales está dada porque gracias a ello se puede entender la forma de comportarse de las enfermedades. Por ejemplo, y según esta gráfica, se está ante una enfermedad que tiene una tendencia al incremento. Por supuesto que esta descripción ha de servir para comparar.

2.1.1.1.1. Ejemplo: Historia de una epidemia (SIDA)

En 1980 el mundo detectó una enfermedad que (fase pre-patogénica) podía afectar a cualquier persona; sin embargo, se encontraba con mayor frecuencia en población promiscua sexualmente o población que requiriese transfusiones, enfermedad que daba como consecuencia de una infección viral por un germen de características especiales (retrovirus) que permite ser detectado, que puede provocar, (fase patogénica) una estado de portador asintomático (fase subclínica) el que evoluciona ulteriormente con la aparición de un cuadro cuya característica es la inmunosupresión (fase clínica), provocando por ende, diversos cuadros de infecciones oportunistas, desencadenando un cuadro mortal (fase de desenlace) casi siempre de rápida evolución. Se calcula que para este momento unos 40 millones de seres humanos padecen de esta enfermedad (**prevalencia**) Hace 15 años se calculaban en 21 millones. Ello implica que no se ha logrado controlar la **incidencia**.



Hoy, el mundo elabora numerosas campañas para evitar que la población susceptible y de mayor riesgo adquiera la enfermedad (intervenciones primarias) y la combate a través de la detección de la infección en población asintomática o con el uso de medicamentos antirretrovirales en la población sintomática (intervenciones secundarias). El desenlace ha cambiado de manera considerable por el uso oportuno de los antirretrovirales. Claro, esto último siempre y cuando se tenga acceso a los mismos, pues la gran mayoría de población que hoy padece del SIDA no tiene acceso a los medicamentos... ipor lo cual mueren igual que hace 26 años! Corolario: **El medio** es esencial para el pronóstico de las enfermedades.

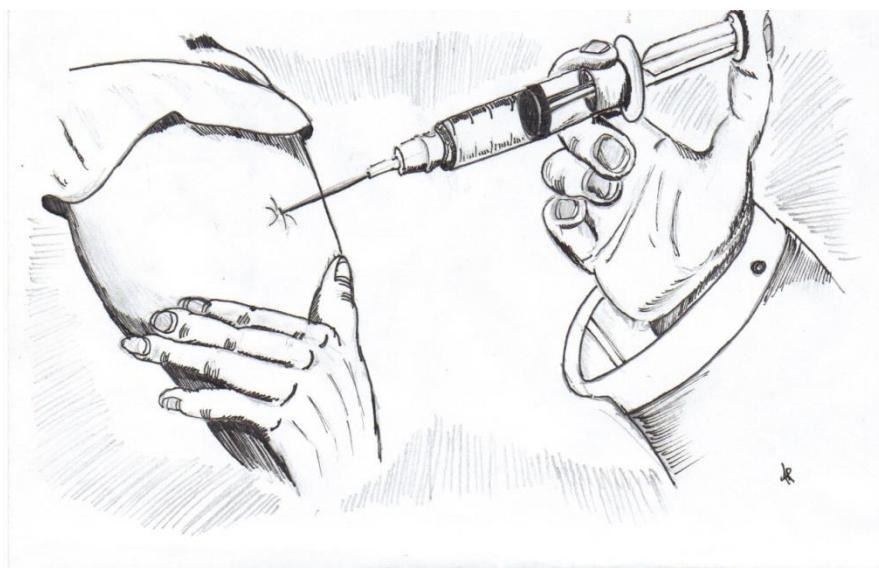


Uno de los problemas vigentes de la humanidad es la INEQUIDAD. Se define equidad como todos tener la misma oportunidad. Es muy evidente que en el mundo, **no todos tienen las mismas oportunidades**.

Como se observa, la forma de clasificar las enfermedades a través de la historia natural es distinta a la que emplea, por ejemplo, la nosología. En tanto que la visión de la modalidad epidemiológica está enfocada hacia la población, la nosología lo hace enfocado hacia el individuo

(fíjese cómo la medicina interna clasifica a las enfermedades). Por supuesto que tienen numerosos puntos de confluencia. A manera de ejemplo de uso, se puede mencionar que las actividades efectuadas en la fase pre-patogénica se denominan prevención primaria y están dirigidas a evitar la aparición de la enfermedad. En la fase patogénica, las actividades de intervención pueden buscar, o bien detectar lo antes posible, a los enfermos y así iniciar lo más rápido posible el tratamiento (si es que lo hay) y se denominan acciones de prevención secundarias. Y todas aquellas actividades que se hacen sobre las secuelas, se denominan de prevención terciarias y buscan esencialmente rehabilitar al enfermo de sus secuelas.

Finalmente, y como para redondear la idea, esta clasificación puede ser útil para entender los planes de salud pública. Las grandes y masivas acciones (intervenciones) de la salud pública están dirigidas a la prevención primaria. Las vacunas y la educación dirigidas a toda la población son un buen ejemplo. La citología (como intervención) a toda mujer en edad fértil es un ejemplo de prevención secundaria. Y los programas de rehabilitación son el ejemplo de las acciones de prevención terciaria, acciones que en una alta frecuencia se incrementan en la medida que los seres humanos viven más.



Eduardo Jenner en 1780 describió magistralmente cómo los granjeros que curaban a los caballos de un mal denominado Ajuagas y que después ordeñaban, trasmítían a las vacas, y desde estas a las personas, lo que denominaban el cowpox o viruela de las vacas. Pero lo realmente extraordinario de su descripción fue hacer notar que aquellos a los que les daba cowpox no les daba viruela. De ello dedujo que el contacto con la viruela de las vacas protegía contra la mortal viruela humana. Concluyó que si inoculaba una pequeña cantidad en las personas podría evitar la viruela.

2.1.1.2. La cadena epidemiológica

Dado que el estudio de las infecciones ha sido una de las áreas donde la epidemiología ha demostrado su desarrollo, la identificación de la ruta que sigue un agente infeccioso permite diseñar conceptualmente una cadena con los siguientes eslabones:

1. Fuente de infección
2. Puerta de salida
3. Vía de diseminación
4. Puerta de entrada
5. Susceptible

El concepto de la cadena infecciosa ha sido un esfuerzo por poder ordenar y así entender las infecciones. Siendo su utilidad inmensa, se debe reconocer que cuando se ha querido trasladar a las enfermedades crónicas, los criterios utilizados no han sido suficientes. Es decir, nos falta aún mucho que entender en las enfermedades crónicas.

La fuente de infección es todo ser vivo que alberga y disemina la infección. De la fuente de infección sale el agente infeccioso que pueden ser bacterias, virus, hongos y parásitos. Lo más frecuente es que sea el hombre la principal fuente de infección de los hombres.

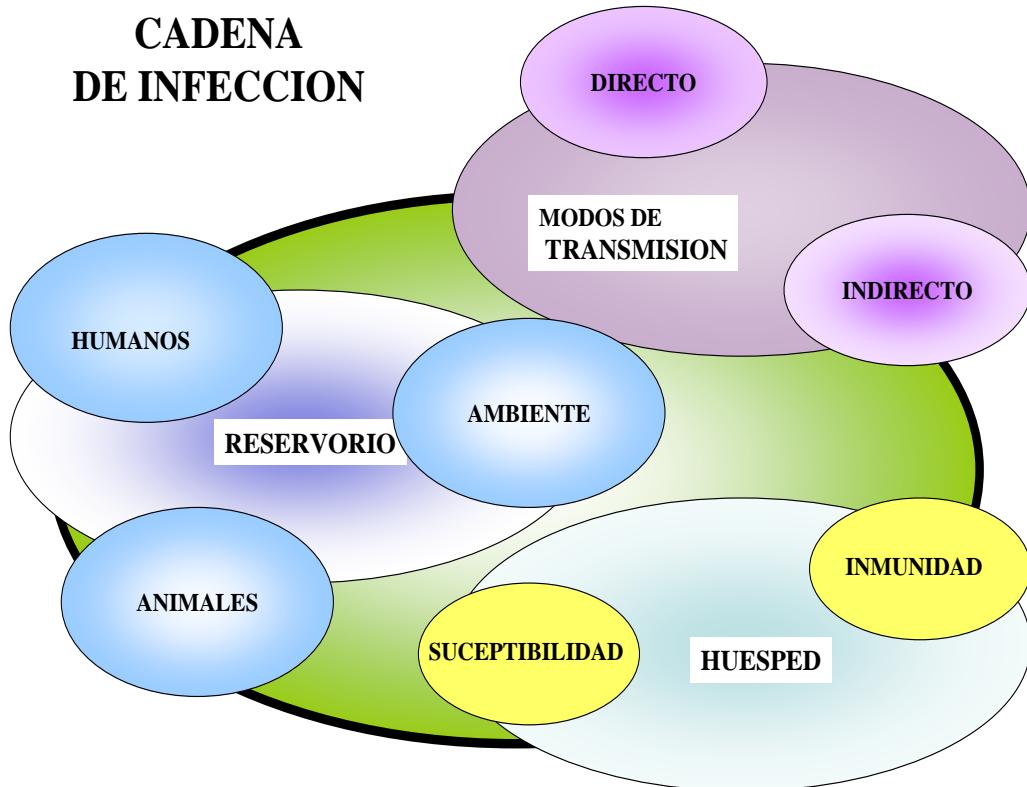
Ciertamente los microorganismos tienen un hábitat natural en las distintas especies. En general, los agresivos (virulentos) contra una especie no afectan otras. Sin embargo, existen microorganismos que afectan a especies distintas. ¿Cuándo se respeta lo primero o cuándo sucede lo segundo? Ello es uno de los aspectos que se busca entender y estudiar a través de la historia natural.

Cuando el hombre se pone en contacto con el agente infeccioso pueden sucederle varias opciones:

- Que se convierta en un **caso** que puede ser clínico (se identifica fácilmente por la presencia de las manifestaciones) o subclínico (se requiere de **pruebas** para identificarlo y tiende a pasar desapercibido). Acá surgen los **portadores** pues estas personas portan al agente infeccioso, que a su vez pueden ser sanos, ya por **inmunidad particular** o por estar en el período de incubación o ser un portador convaleciente (ya tuvo la enfermedad y que aún tiene el agente infeccioso, por lo que está en fase de convalecencia).



Los portadores pueden ser un enorme problema para la contención de las enfermedades. Por ejemplo, los portadores del virus del VIH o de la hepatitis B contagian a otros sin darse cuenta pues ellos mismos ignoran su situación. Tal vez el acceso a pruebas de detección masivas podría enfrentar este problema.



El sitio por donde sale el microorganismo se denomina **puerta de salida**. Puede ser una puerta permanente, por ejemplo, vía respiratoria (al toser o hablar) o intermitente como vía intestinal, urinaria, mamaria, piel y mucosas. Claro... la importancia de entender esto está en que acá existe un sólido criterio para entender la propagación de una enfermedad infecciosa y, por supuesto, contagiosa.

De ahí el susto que se tiene de una epidemia de gripe aviar. Tal parece que las aves pueden contagiar a los hombres y éstos, a través de la vía respiratoria, propagarla en grandes proporciones (**infectividad**).

La vía de diseminación busca hacer entender qué usa el microorganismo para transportarse desde la fuente de la infección hasta el sujeto que se puede enfermar (**susceptible**). Se han clasificado estas vías como:

1. Directa: referida a aquella en la cual se usa el menor número de intermediarios, así que la enfermedad se disemina persona a persona. Un buen ejemplo son las enfermedades de transmisión sexual donde el infectado contagia al contacto de manera directa.

Sin embargo, se sabe que existen enfermedades de transmisión sexual que el hombre traspasa a animales y éstos pueden devolvérsela al hombre en aquellas situaciones en donde se promueve el contacto sexual hombre-animal.

2. Indirecta: referida al uso por parte de los agentes infecciosos de transportes, condición que permite suponer que los agentes infecciosos son más resistentes (duran más fuera de donde se originan). La forma de trasportarse puede ser a través de vehículos inanimados (agua, leche, alimentos, suelo, aire, polvo u objetos contaminados). La otra forma es por vehículos animados, denominados también vectores y se subdividen en aquellos que sólo lo transportan (moscas en sus patas) o aquellos en que el microorganismo utiliza el metabolismo del vector en su desarrollo, Por ejemplo, el paludismo.

La puerta de entrada se refiere a la forma como entra el microorganismo en el susceptible. Existe una relación entre la fuente de salida y la de entrada (por donde sale entra).

Una identificación clave es la del **susceptible**. Se define como los individuos que por no tener inmunidad o algún tipo de resistencia al agente infeccioso les puede dar la infección.

El concepto y los términos de la cadena infecciosa resultan particularmente importantes para poder enfrentar a las infecciones y en particular las epidemias. Imagínense que se presenta un brote de una enfermedad que se manifiesta como diarrea y vómito en personas que almorzaron en un restaurante. Se identifica que se trata de una propagación de infección por estafilococo. Los pacientes que padecen de las manifestaciones son sospechosos de ser **casos clínicos**. Se dice sospechosos porque no se ha confirmado el microorganismo en sus cuerpos. Si se hacen pruebas y salen **positivas** entonces se describen como casos confirmados. Si la prueba sale positiva en uno de los comensales que no tiene manifestaciones se habla de un **portador**. Al buscar la fuente del contagio, se detecta que uno de los cocineros

presenta una infección en los dedos y se confirma la presencia del estafilococo. Ahí está la **puerta de salida**. La contaminación de los alimentos y el consumo de los mismos es la **puerta de entrada** y son los alimentos la **vía de diseminación**. Al armar todo esto, sin duda estaremos en la capacidad de intervenir y modificar en torno a este presente y proyectar acciones para evitar que suceda en un futuro. Sabiendo todo esto, ¿qué se hace?:

- Eliminar el agente. Uso de antibióticos. Se le dará a los casos confirmados y a los portadores asintomáticos (y así evitar que ellos propaguen; claro, si ellos se pueden convertir en propagadores, cosa menos frecuente en las intoxicaciones con estafilococo pero muy frecuente en las salmonelosis).
- Modificar la fuente. Tratar al cocinero. Aislarlo (cuarentena) de sus labores en tanto se cura.
- Tratados los casos se puede insistir a los cocineros del sitio (y de otros para que aprendan por la experiencia de los demás) en las normas y cuidados que han de tener con la manipulación de los alimentos.

2.1.2. La tabla de frecuencias

Las descripciones se presentan, en lo posible, de manera integral y para ello el mejor diseño es el la tabla de frecuencias.

La tabla de frecuencias está constituida por columnas y filas. Por ejemplo, el programa Epi Info construye las tablas de frecuencia de la siguiente manera: La primera fila contiene los ítems de lo que muestra. La primera columna expone el criterio que se va a medir. La segunda columna expone las frecuencias. La tercera el porcentaje de dichas frecuencia con denominador del total de frecuencias. La cuarta columna muestra el porcentaje acumulado de cada nuevo criterio (suma el porcentaje que venía más el nuevo. El Epi Info produce una columna donde aparece la representación en barras de las frecuencias en porcentaje.

CRITERIOS	Frecuencia	Porcentaje	Porcentaje acumulado	BARRAS
CRITERIO 1	2	6,70%	6,70%	
CRITERIO 2	2	6,70%	13,30%	
CRITERIO 3	2	6,70%	20,00%	
CRITERIO 4	2	6,70%	26,70%	
CRITERIO 5	2	6,70%	33,30%	
CRITERIO 6	4	13,30%	46,70%	
CRITERIO 7	8	26,70%	73,30%	
CRITERIO 8	2	6,70%	80,00%	
CRITERIO 9	1	3,30%	83,30%	
CRITERIO 10	1	3,30%	86,70%	
CRITERIO 11	1	3,30%	90,00%	
CRITERIO 12	1	3,30%	93,30%	
CRITERIO 13	1	3,30%	96,70%	
CRITERIO 14	1	3,30%	100,00%	
Total	30	100,00%	100,00%	

Una buena tabla de frecuencias debe hablar por sí misma: ello quiere decir que cualquiera que la vea sepa de qué se trata y qué describe. Así mismo, es muy importante colocarle un título a la tabla. Debe tener: las características que describe, la población que describe, y cuándo y dónde se realizó. La primera columna se denomina COLUMNA MATRIZ y ahí se ubica lo que se va a describir y la primera fila se denomina así mismo fila matriz y describe las características que se van a exponer.

Tabla # 1 CARACTERISTICAS DE LA POBLACIÓN Y LAS FAMILIAS DE LOS ADOLESCENTES DE 12 COLEGIOS PUBLICOS DE LA LOCALIDAD DE SUBA EN EL AÑO 2006.

EDADES	Nº	%
De 10 A 12 años	3965	32
De 13 A 16 años	7200	59
De 17 A 19 años	997	8
GENERO	Nº	%
Mujeres	6274	51
ESTRATO	Nº	%
Estrato 1	347	3
Estrato 2	8958	73
Estrato 3	2475	20
ESTRUCTURA FAMILIAR	Nº	%
Nuclear completa	6862	56
Nuclear incompleta	3774	31
Recompuestas	1112	9
Hogares donde faltan ambos padres	227	2
FUNCIONALIDAD FAMILIAR	Nº	%
Buena función familiar	7722	63
Disfunción moderada	2731	22
Disfunción severa	920	7

Por ejemplo, la tabla está numerada, dice que muestra (CARACTERISTICAS) de quién, de dónde y cuándo. ¡Ah...! es parecida a otra presentada antes (subtítulo 2.1.1.), pero aquella tiene más población que esta.

La tabla de frecuencia forma parte obligada de toda investigación. Por ejemplo, en el caso del SIDA se puede construir una tabla de frecuencia de manifestaciones. Para ello se revisan los casos de una población que padezca de SIDA. Como criterios se ponen las manifestaciones

presentes en dicha población. Como frecuencia el número de veces presente de cada característica en dicha población. El porcentaje representa la proporción de dicha característica con el total de la población. Si bien se podría poner el resto de componentes de la tabla (frecuencia acumulada y porcentaje acumulado), lo cierto es que lo habitual es solo usar frecuencia y porcentaje.



A manera de ejemplo busque, si desea, un artículo sobre SIDA y observe la tabla de manifestaciones.

2.1.3. La forma de aprender: comparar

Dos estrategias empleadas por el hombre para aprender son la descripción y la comparación. Comparar implica conocer un referente que se usa para comparar. La epidemiología ha desarrollado una herramienta formidable para la comparación, que se denomina la tabla de contingencia.

		VARIABLE DEPENDIENTE		
		SI PRESENTE	NO PRESENTE	
VARIABLE INDEPENDIENTE	Si presente	a	b	a+b
	No presente	c	d	c+d
		a+c	b+d	

El principio básico de la comparación es no tomar más de dos variables. Una variable es una característica (otra definición a las ya descritas). El elemento crucial es la presencia o la ausencia de la característica y relacionarlas entre ellas.



No debemos confundirnos con las denominaciones de las variables. Recordar que las casillas de variables dependientes reciben según el contexto otros nombres, por ejemplo, en el programa Epi Info se denominan variables de resultado y las independientes se buscan como variables de exposición.

2.1.4. La tabla de comparación: los factores de riesgo y la evaluación de lo que sirve

Para poder entender a cabalidad, se va a presentar la tabla de resultados encontrada por los doctores Doll y Hill en 1950 cuando efectuaron sus estudios sobre el cáncer de pulmón y su relación con fumar.

		VARIABLE DEPENDIENTE = CANCER DE PULMON		SI PRESENTE	NO PRESENTE	1298
VARIABLE INDEPENDIENTE = FUMAR		Si FUMA	No FUMA			
	Si FUMA	647	622			1269
	No FUMA	2	27			29
		649	649			

Ellos querían verificar si existía alguna relación entre el hábito de fumar y el cáncer de pulmón, el que estaba presentándose con una incidencia incrementada (**epidemia**). Para ello compararon personas con cáncer y sin cáncer (variable dependiente) y averiguaron si fumaban o no (variable independiente). Antes de interpretar la tabla, es muy conveniente familiarizarse con ella. Para ello se emplearán las letras descritas previamente en cada casilla.



Epidemia: aumento no esperado de casos en un lapso de tiempo. Enfermedad que se propaga acometiendo simultáneamente a gran número de personas. Endemia es la frecuencia elevada de una enfermedad en una región. Pandemia es una epidemia que no respeta fronteras.

La letra a expresa quiénes tienen cáncer y fuman (647).

La letra b quiénes fuman y no tienen cáncer (622).

La suma de a + b expresa el total de fumadores del estudio (1269).

La letra c expresa quiénes no fuman y tienen cáncer (2).

La letra d muestra quiénes no fuman y no tienen cáncer (27).

La suma de a + d muestra quiénes no fuman (29).

La suma de a+ c muestra el total de personas con cáncer del estudio (649).

La suma de b+d el total de personas que no tienen cáncer (649).

El total de personas de este estudio fueron 1298.

¿Qué es lo que se quiere comparar?... sencillamente del total de fumadores cuál es la proporción que presentan cáncer comparado con la proporción del total de no fumadores que presentan cáncer. Expresado en letras sería lo siguiente:

$$\frac{a}{a+b}$$

$$\frac{c}{c+d}$$



Dado que lo que se compara son dos eventos distintos, en este caso dos números (obtenidos de manera diferente o lo que simbólicamente es a/b) entonces la medida con que se conoce esto es la de **RAZÓN**.

Como se observa se está obteniendo una **razón** de comparación y el resultado de la operación matemática fue de 7,3. (Hágalo para que confirme). Este resultado (y la fórmula) se denomina el **RIESGO RELATIVO**. Se interpreta de la siguiente manera:

**VARIABLE
DEPENDIENTE =
CANCER DE PULMON**

SI PRESENTE

VARIABLE INDEPENDIENTE = FUMAR	Si FUMA	7,3
	No FUMA	1

 que en el capítulo 4 se ampliará el concepto, sea este el momento de decir que con el fin de hacer entender se usan números. Por ejemplo, el número de personas con cáncer del pulmón. Sin embargo, a veces el número no permite entender todo lo que se quiere, entonces la **proporción** se convierte en una importante alternativa. Ahí se comparan el número de casos dividido por el total de población referenciada y se multiplica por 100. Por ejemplo, número de fumadores dividido por el total de cánceres del pulmón multiplicado por 100. Se obtiene la proporción de cáncer en los fumadores. Pero existe otra forma más interesante de comparación: y son las **razones**: acá se compara un dato con otro dato diferente (a/b). Por ejemplo, el resultado de fumadores con cáncer dividido con el resultado de cánceres en no fumadores. Todas estas expresiones numéricas solo pretenden hacer entender y el uso de una u otra manera dependen de la capacidad de cada una de contribuir con el entendimiento.

Por cada persona que no fuma y le dio cáncer de pulmón (1 EN LA RELACIÓN), existen 7,3 que fuman y les dio cáncer. Como se observa, la medida a comparar **siempre es 1** que representa a los NO EXPUESTOS y el resultado de la fórmula de RR (Riesgo Relativo), en este caso 7,3, expresan los expuestos. Claramente se ve cómo **los que fuman** (si fuman) se exponen a un riesgo mayor de padecer de cáncer (7 veces más en este estudio) que los que **no fuman**.

 Ampliando el concepto: Las razones son una manera de presentar información, usando como referente el denominador de la división de dos datos. Se compara el numerador (A) con el denominador (B) y se usa siempre como 1 el denominador el que se compara con el resultado de la división A/B. Por ejemplo, en un salón existen 30 personas y de ellas 10 son hombres y 20 mujeres. Ante la pregunta cuál es la razón mujeres-hombres: A= 20 mujeres. B= 10 hombres. La razón mujeres/hombres será: $20/10 = 2$. Entonces se presenta así: Por cada hombre (siempre el denominador es 1) existen 2 mujeres (el resultado de la división). Si la pregunta fuera cuál es la razón hombres-mujeres, entonces sería $10/20 = 0.5$. Se expresaría que por cada mujer existe 0.5 hombre.

Como conclusión se puede afirmar que el riesgo relativo es una de las medidas usadas por la tabla de comparación o de contingencia. De hecho, y en epidemiología, la tabla de contingencia originan las MEDIDAS DE COMPARACION que, por supuesto que son razones (A/B) obtenidas de comparar el numerador con el denominador. El RR es la medida más representativa junto con el OR (Odds Ratio). Este último tiene la misma interpretación que el RR, pero dado lo que compara, sus resultados se consideran relaciones potencializadas. La fórmula es:

$$\frac{a \times d}{b \times c}$$

El resultado en este estudio del OR fue de 14. Expresado a manera de riesgo es así:

VARIABLE
DEPENDIENTE =
CANCER DE PULMON

SI PRESENTE

VARIABLE INDEPENDIENTE = FUMAR	Si FUMA	14
	No FUMA	1

Por cada persona que no fuma y le dio Cáncer de pulmón, existen 14 que fuman y presentaron cáncer. Es evidente la diferencia entre el RR y el OR, pero para entenderlo, se ha de pensar, por ejemplo, que en un caso se habla de centímetros en tanto en el otro de metros. En ambos casos se habla de lo mismo, (que alguien es más alto que otro) pero se usan referentes distintos.

La tabla comparativa o de contingencia es ampliamente usada en el mundo de la salud. A través de ella se buscan factores de riesgo, se comparan resultados... en fin, es una de las herramientas más valiosas para entender a través de la comparación.



Resulta interesante analizar que los doctores Doll y Hill detectaron esta poderosa relación en 1950. Quiere decir que desde entonces existían fuertes indicios de la relación entre fumar y el cáncer del pulmón. Ello no evitó que surgiera por parte de las tabacaleras una fuerte promoción al consumo de cigarrillos, la que fue tan efectiva que hoy se considera que es la principal adicción del mundo. Y ello sabiendo que induce no sólo al cáncer sino a otras muchas enfermedades. De ahí el dicho: TENER LA RAZÓN NO ES RAZÓN SUFICIENTE PARA SEGUIRLA.



Una explicación simple y sencilla del por qué no se ha podido prohibir la producción de cigarrillos es porque su venta produce dinero y con ese dinero se compra la influencia política que permite la venta, usando para ello argumentos tales como la libertad de escoger. El bien común queda doblegado ante la denominada libertad individual. Una muy conocida marca de cigarrillos utilizó a la publicidad (con campañas que verdaderamente inducían a fumar) como una de sus principales estrategias para conseguir clientes. Anecdóticamente uno de los modelos empleados en su campaña, fumaba y murió de cáncer de pulmón.

2.1.4.1. Usos de la tabla de comparación

No sobra insistir en la mayúscula utilidad de la tabla de comparación. Y es útil porque permite ver una serie de referentes (medidas) que de otra manera no sería posible. Constituye el elemento clave para identificar componentes relacionados con la enfermedad. Recuerde que el término relación matemáticamente se expresa como una división y de hecho la operación que se hace cuando se buscan relaciones es una división.

2.1.4.1.1. Identificación de riesgos

La tabla de contingencia sin duda es la herramienta esencial para determinar riesgos. Riesgo, es la probabilidad de ocurrencia de un evento en salud. La identificación de riesgos necesariamente se hace tomando como variable dependiente (o de resultado) a la enfermedad en cuestión relacionándola con la presencia o ausencia de factores de exposición.



La búsqueda del RIESGO implica tomar población que padece la enfermedad y usar como referente población que no la padece e indagar por la presencia o ausencia de los factores que se sospechan, están relacionados. La tabla de relaciones (o lo que es lo mismo, comparativa o de contingencia) se construye así:

VARIABLE INDEPENDIENTE = FACTOR DE EXPOSICION	VARIABLE DEPENDIENTE = ENFERMEDAD		Total SI expuesto = $a+b$
	SI tiene	NO tiene	
SI expuesto	SI tiene la variable dependiente y SI esta expuesto = a	NO tiene la variable dependiente y SI esta expuesto = b	Total NO expuesto = $c+d$
NO expuesto	SI tiene la variable dependiente y NO esta expuesto = c	NO tiene la variable dependiente y NO esta expuesto = d	
	Total SI tienen la variable dependiente = $a+c$	Total NO tiene la variable dependiente = $b+d$	

Lo que se compara puede ser:

1. La relación entre el total de expuestos y que les dio la enfermedad.
2. Con el total de los no expuestos que les dio la enfermedad

El resultado de esa relación es la división de la cifra resultante de dividir a los $a/a+b$ dividido entre $c/c+d$.

Dicho resultado (y supóngase que se llama RR) se compara así:

Por cada **1** no expuesto y que le dio la enfermedad, existe **RR** que sí se expuso y le dio la enfermedad. Si el RR da **más de 1**, se hace claro que la exposición favorece la aparición de la enfermedad. Si **el RR** da **1** implica que da lo mismo exponerse o no exponerse para que aparezca la enfermedad. Y si el RR da **menos de 1**, implica que la enfermedad les da más a los no expuestos que a los expuestos (por ende, exponerse protege de la enfermedad).



Para completar, obsérvese estos ejemplos retomando el trabajo de Doll y Hill:

VARIABLE INDEPENDIENTE (EXPOSERSE AL CIGARRILLO)	VARIABLE DEPENDIENTE		
	SI CANCER PULMONAR	NO CANCER PULMONAR	
SI FUMO	647	622	1269
NO FUMO	2	27	29
		649	649

El producto de relacionar a los que se exponen y les dio el cáncer es = 647/1269 con los que no fuman y les dio cáncer = 2/29 provoca una relación:

	RR
Si fumo y Si cancer pulmonar	7,4
No fumo y Si cancer pulmonar	1

Es decir que aquellos que fuman al ser comparados con los que no fuman, les da cáncer 7,4 veces más.

Lo cierto es que los resultados originales no fueron dados con RR sino con OR (que se trata de otra medida de relación) dando una relación de riesgo aún mayor. Ello por el tipo de investigación.

En ese mismo estudio, surgió la pregunta de si ser hombre (comparado obviamente con el de ser mujer) era un factor de riesgo para la aparición del cáncer. La tabla se construyó así:

	Si cáncer	No cáncer			RR
Hombre fumador	647	600	1247	0,518845229	1,00
Mujer fumadora	41	38	79	0,518987342	
	688	638			

Fue necesario hacer una nueva clasificación y separar a los hombres y las mujeres en fumadores y no fumadores (esto se conoce como proceso de **estratificación** y se mencionará más adelante). Lo que se busca es analizar si existe correspondencia entre el género y el cáncer; por ende, se buscó que tanto los hombres como las mujeres que estuviesen expuestos, se relacionaran con el cáncer. Los resultados fueron los expuestos en la tabla (aunque los datos expuestos no fueron los del estudio original, sirven de ejemplo para el aspecto que se está revisando).

	RR
Si hombre y SI cancer pulmonar	1,0
No hombre (mujer) y SI cancer pulmonar	1

Se interpreta que por cada mujer fumadora y con cáncer, existe 1 hombre fumador con cáncer. Es decir que da lo mismo ser hombre fumador o mujer fumadora en torno al cáncer. Resultado por demás lógico.

Es el fumar y no el género el factor de riesgo para el cáncer pulmonar.

Para poder expresar el último ejemplo, se usa un estudio de las favelas del Brasil donde se buscaban factores de riesgo y protección que fuesen útiles para disminuir la incidencia de jóvenes asesinados. El factor estudiado fue si habían asistido a la educación primaria y se encontró:

VARIABLE INDEPENDIENTE	VARIABLE DEPENDIENTE		
	SI asesinados	NO asesinados	
SI estudios primarios	46	168	214
NO estudios primarios	118	48	166
	164	216	

El resultado de la relación fue:

	RR
Si estudios primarios y SI asesinado	0,3
No estudios primarios y SI asesinado	1

Se interpreta que por cada joven que No tuvo estudios y fue asesinado, hay 0,3 jóvenes que si estudiaron y fueron asesinados. Como es menos de 1, claramente expresa que estudiar es un factor que protege a los jóvenes de ser asesinados.

Sea este el momento de afirmar que la medida utilizada en estos ejemplos ha sido el RR o riesgo relativo. Sin embargo, existe otra medida que se puede emplear como es el Odds Ratio que, si bien se interpreta de la misma manera (1, más de 1 y menos de 1), su obtención se hace distinta ($OR = aX_d/bX_c$). En general da resultados con números más grandes que el RR. El uso de una medida o de otra depende del diseño de la investigación.

En general se puede decir que el RR se usa en los estudios de cohorte y en los experimentales. ¿Por qué? Sencillamente porque en los estudios prospectivos lo que se mide son los casos nuevos que surgen, es decir, la incidencia y el RR es precisamente una de las formas de obtenerlo. En los estudios de corte o de casos y controles se utiliza el OR.

En torno a su interpretación y ofreciendo una aproximación cualitativa, se afirma que entre mayor sea el número de 1, mayor será el riesgo. De 1 a 2 riesgo moderado, de 2 a 4, riesgos importantes, pero más de 4 hace sospechar que puede tratarse de un factor causal directo. Por ejemplo, en el estudio de Doll y Hill de casos y controles, el OR del cigarrillo para el cáncer pulmonar fue de 14, resultado que hace sospechar que se está ante una relación causal (ver subtítulo 2.2.). Lo mismo en riesgos menores de 1. Entre más se acerque a 0, más protector. Obviamente, cuantitativamente los valores del RR y el OR son datos numéricos que reflejan perfectamente la interpretación de su valor: si el RR es 1,8, se entiende perfectamente que al compararlo con otra variable que obtuvo OR de 3,3, ésta última en ese estudio es un factor de riesgo mayor.



2.1.4.1.1.1. Presentación de una tabla de contingencia

Una buena presentación de una de tabla de contingencia ha de permitir que el lector, si lo desea, pueda hacer los cálculos de riesgo correspondiente.

		ACTITUDES										PRACTICAS											
		Quiere tener un hijo este año					No cree que papás apoyarían planificación					Tener sexo sin querer			Relaciones sexuales								
		Comparaciones	Si (n)	%	No (n)	%	Odds Ratio	IC95%LI		Valor p	Si (n)	%	No (n)	%	Odds Ratio	IC95%LI		Valor p					
FUNCIÓN FAMILIAR	D. Severa							182	35	345	66			1.411	1.211	1.644	0						
	D. Moderada							475	30	1120	70												
	D. Severa	41	10	369	90		2,232	1,463	3,406	0,0002	182	35	345	66	1,702	1,35	2,145	0					
	Funcional	132	5	2689	95						1046	27	2889	73									
Casos incluidos:		3690					5271																
		PRACTICAS																					
		Tener sexo sin querer								Relaciones sexuales													
FUNCIÓN FAMILIAR	D. Severa	93	10	794	90		1,43	1,168	1,749	0,0005	295	30	685	70									
	D. Moderada	206	8	2543	92						612	21	2363	79									
	D. Severa	93	10	794	90		2,112	1,589	2,805	0	295	30	685	70	2,063	1,718	2,477	0					
	Funcional	363	5	7582	95						1188	14	7262	86									
Casos incluidos:		9822					10519																

Que se compara

Pruebas de significancia

medida

Datos que permiten calcular el OR

En el ejemplo se observa: título: Razones de riesgo (OR) de actitudes y prácticas relacionados con la función familiar de una población adolescente escolar de localidad de Suba en el año 2007.

Como se nota la tabla debe permitir ver qué se está comparando, debe permitir ver, y si se quiere, calcular la medida (en este caso OR) y las pruebas estadísticas. De hecho en una tabla de OR solo deben aparecer los datos significativos estadísticamente, se evita incluir los datos los que no lo son, para así concentrar al lector en los datos significativos.

2.1.4.1.2. Otras comparaciones: sensibilidad y especificidad

La tabla comparativa tiene otra importante función. Se trata de la de ayudar a describir la relación de las manifestaciones (o pruebas de detección) con las enfermedades. Parte de que la detección de las enfermedades se hace a través de la presencia de manifestaciones (o pruebas) que están presentes cuando se está enfermo. Para poder comparar se ha de tener dos poblaciones: una de enfermos y otra de no enfermos (referente necesario) y se deberá indagar por la presencia o ausencia de la manifestación. La tabla comparativa se construye así:

VARIABLE INDEPENDIENTE	VARIABLE DEPENDIENTE		Total SI tienen la manifestación = a+b
	SI ENFERMEDAD	NO ENFERMEDAD	
SI PRESENTE LA MANIFESTACIÓN	SI tiene la enfermedad y SI tiene la manifestación = a	NO tiene la enfermedad y SI tiene la manifestación = b	Total NO tienen la manifestación = c+d
NO PRESENTE LA MANIFESTACIÓN	SI tiene la enfermedad y NO tiene la manifestación = c	NO tiene la enfermedad y NO tiene la manifestación = d	Total SI tienen la enfermedad = a+c
	Total NO tiene la enfermedad = b+d		

La relación que se consigue es del total de enfermos, cuántos presentaron la manifestación. Esto se conoce como SENSIBILIDAD y se obtiene de la relación entre $a/a+c \times 100$ (es una proporción). La otra importante relación es la derivada del total de NO enfermos y que no tienen la manifestación. Se conoce como ESPECIFICIDAD y se expresa a través de $d/b+d \times 100$.

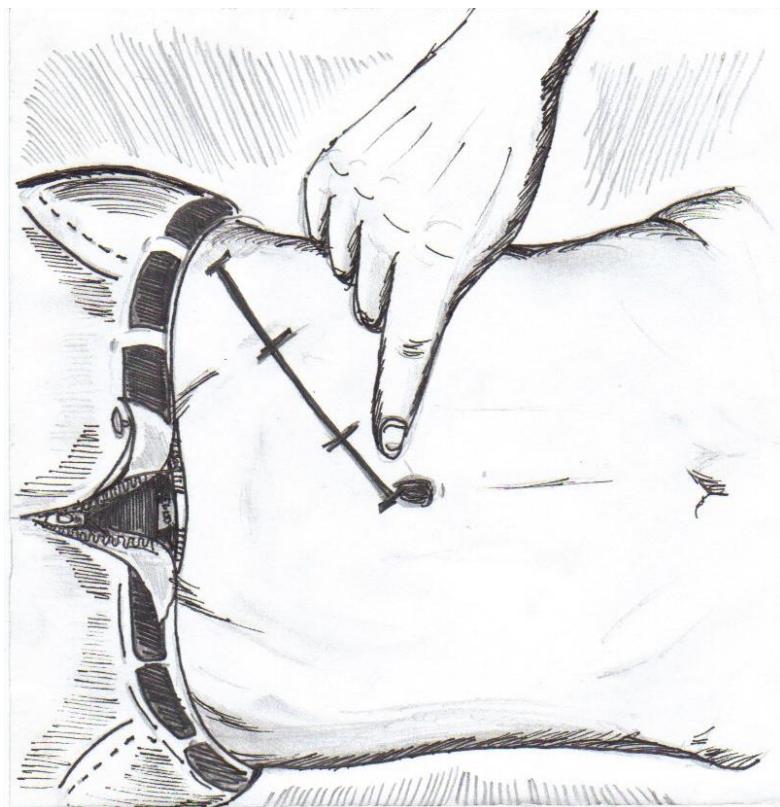


A manera de ejemplo, obsérvese el valor del signo de Blumberg en la tabla resultante de un estudio de probabilidades diagnósticas de apendicitis efectuado en una clínica de Bogotá en el 2002.

		VARIABLE DEPENDIENTE		
		SI	NO	
		APENDICITIS	APENDICITIS	
SI BLUMBERG		128	86	214
NO BLUMBERG		34	249	283
		162	335	497

Lo que expresa es que el signo de Blumberg estuvo presente en 128 pacientes de un total de 162 con apendicitis (si se multiplica el resultado por 100 se obtiene la sensibilidad). Que no estuvo presente en 249 de un total de 335 pacientes que no tenían apendicitis (si se multiplica por 100 se obtiene la especificidad). Expresa:

1. Que la sensibilidad del Blumberg en la apendicitis fue de un 79%.
2. Que la especificidad del Blumberg en la apendicitis fue de un 74%.



La determinación de la sensibilidad y la especificidad es clave en el estudio de la utilidad de las pruebas diagnósticas y hoy son partes esenciales en la determinación de la construcción de probabilidades de la enfermedad a partir de las manifestaciones.

La sensibilidad implica quiénes SI TIENEN ESTO, ESTÁN ENFERMOS. La especificidad, quiénes si NO TIENEN ESTO, NO ESTÁN ENFERMOS. Todo el proceso de diagnosticar enfermedades se hace a través de la sensibilidad y la especificidad. El raciocinio es sencillo.

1. Los médicos, a través de la historia clínica, indagan por la presencia o ausencia de manifestaciones.
2. Con base en ello van construyendo en sus mentes los cuadros clínicos que les permitirán aproximarse a decir que se trata de esta o aquella enfermedad.

En general, y al concluir la historia clínica inicial, el médico ha de tener una aproximación importante a uno o varios diagnósticos. La forma como interpreta lo que encuentra (y lo que no) le da la fuerza a sus impresiones diagnósticas.

3. Luego de la historia clínica y de acuerdo al grado de certeza que tengan, precisarán a través de pruebas más específicas que tiene o no tiene una enfermedad.



Este paso es un complemento del anterior. Existen pruebas paraclínicas que ofrecen de acuerdo a sus resultados una aproximación mayor a la certeza o que tiene o que no tiene la enfermedad. La habilidad del médico está en saber pedir las pruebas de las que obtendrá mayor precisión para la toma de decisiones.

4. Una vez se tiene esa precisión diagnóstica, viene el segundo momento del acto médico que es definir la intervención. Por supuesto, la posibilidad de éxito de la intervención depende de la certeza en el diagnóstico.
5. Finalmente han de esperar el desenlace de la intervención y con base a ello aprender para la siguiente vez.

Con la interpretación de estos dos elementos se construye la EVIDENCIA CLÍNICA de las manifestaciones y de las pruebas paraclínicas de las enfermedades. ¿Cómo? En el ejemplo expuesto esto es lo que se hizo: Se buscó historias de pacientes con dolor abdominal que acudieron a urgencias y se diferenció entre aquellos que tenían apendicitis y quienes no la tenían. Luego se comparó cada manifestación y así se obtuvo la sensibilidad y especificidad. Con estos datos y, a través de sencillas pruebas matemáticas, se puede transformar la sensibilidad y la especificidad en PROBABILIDADES. Con ello se dice: ¿Qué probabilidad

tiene este paciente por tener (o no tener) esta manifestación de tener tal enfermedad?...



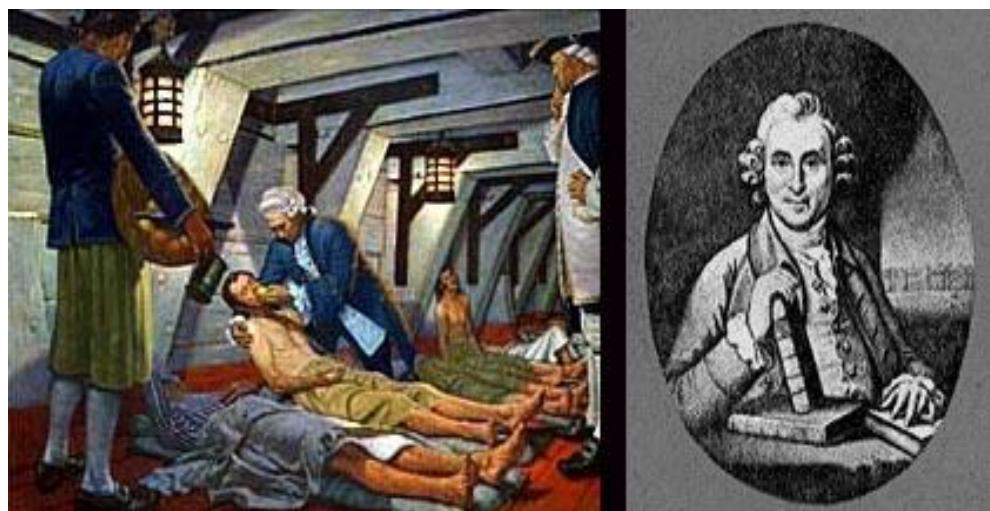
De lo que se trata es de transformar la habilidad médica que ayer se conocía como ojo clínico (donde era la experiencia la base de la habilidad) por la identificación de probabilidades. Dicho cambio implica que la habilidad de ayer, hoy puede estar al alcance de todos los médicos: se trata sólo de que sepan buscar (semiología) y que puedan interpretar el hallazgo (obtener la probabilidad).



2.2. Buscando las causas

Causa es aquello que se considera el fundamento u origen de algo. En salud (y en todo) el detectarla es un paso muy importante (pero no esencial) para poder hacer intervenciones adecuadas. La búsqueda de la causa de la enfermedad ha estado presente a lo largo de la historia de la humanidad de manera extraordinaria. Los paradigmas del pensamiento juegan un importante papel en la búsqueda de la causa. De hecho, y de acuerdo al momento, surgen las hipótesis acerca del por qué de las cosas. Es por ello que en la génesis de la enfermedad se han visto involucradas ideas mágicas, religiosas, empíricas y científicas. Lo cierto es que con base en la causa que la humanidad cree, desarrolla las intervenciones. Si existe relación entre lo que cree y la realidad, entonces las medidas de intervenciones pueden tener éxito. A manera de ejemplo, la infección por el virus del VIH como una porción causal del SIDA ha permitido desarrollar una serie de acciones preventivas en contra de la enfermedad. Pero a veces... sin la certeza de saber, se puede tener éxito. A manera de ejemplo el caso del Dr. James Lind en 1753, cuando puso en evidencia medidas preventivas contra el escorbuto sin ni siquiera entender la enfermedad. Ello es importante, pues de seguro se puede hacer mucho en pos de la salud aún con un grado importante de desconocimiento.

El escorbuto era en los siglos XVII y XVIII el terror de los viajes largos en ultramar. Muchos médicos estaban dedicados a descifrar su origen. James Lind como médico al servicio de su majestad en la marina imperial, escribió un informe donde exponía una investigación de cómo combatirlo. A las seis semanas de estar en alta mar se presentó un brote de escorbuto en 12 marineros.

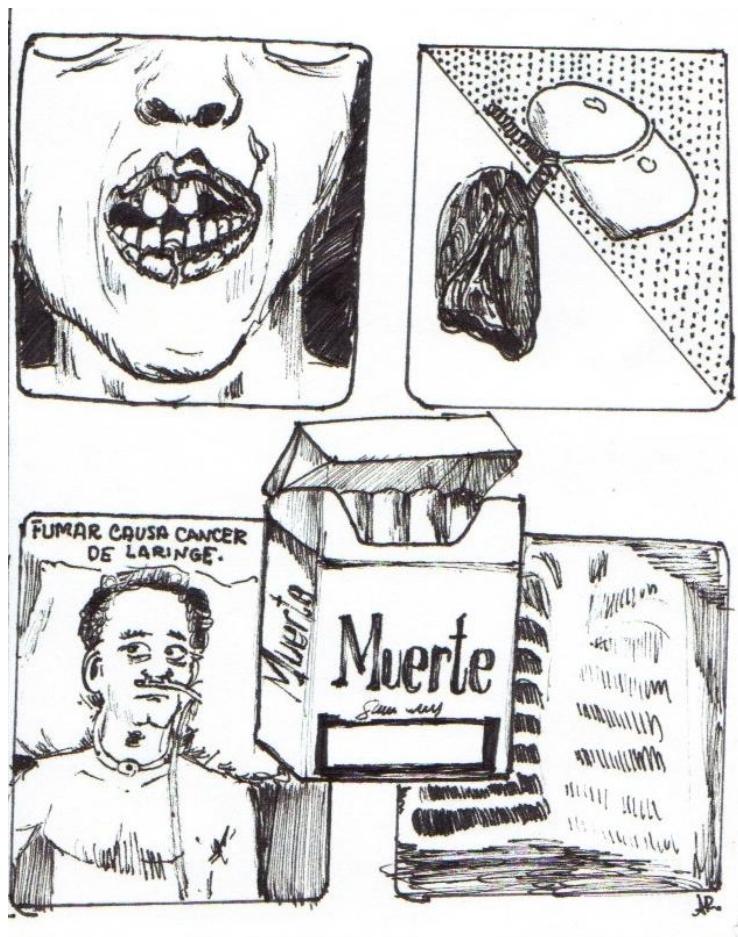


Hizo consideraciones acerca del clima, de la humedad, de los alimentos y del tiempo de estar en alta mar (componentes sobre los que existían hipótesis que los responsabilizaban como causa de esta temible enfermedad) llegando a la conclusión de que en la génesis de la enfermedad la ausencia de alimentos frescos aparece como una probable sospechosa.

Así que dividió en seis parejas a los enfermos (20 de mayo de 1747), los puso en un mismo sitio del buque (descartando así que el lugar pudiese ser un determinante) y a todos les suministró una dieta básica similar, agregándoles de manera diferencial (interviniendo con dietas propuestas por otros médicos que habían escrito al respecto): a una pareja caldo de cordero fresco, a otra, un cuarto de galón de sidra, a la siguiente 25 gotas del elixir de vitriolo (bebéndolo a base de sales de arsénico) tres veces al día, a otra tres cucharadas de vinagre en ayunas; a la quinta pareja, con agua de mar y a la última pareja, dos naranjas y dos limones diarios. En seis días observó una recuperación franca de la pareja que recibía naranjas y limones. Lind descubrió cómo se trataba el escorbuto muchos años antes de saber cuál era su causa. Claro... luego el descubrimiento de Lind permitió saber el papel de la vitamina C en la génesis del escorbuto.

2.2.1 La maravilla de las relaciones y las asociaciones

La identificación de los factores causales se hace con base a la investigación. Una primera fase es la descriptiva, donde se reconoce la enfermedad. Ahí se proponen hipótesis acerca de su génesis. Luego se hacen comparaciones para probar dichas relaciones y finalmente se hacen las intervenciones (o estudios experimentales) para poder demostrar. Las relaciones (obtenidas a través de la tabla comparativa) son un componente indispensable para buscar la génesis de las enfermedades. La fuerza de la relación sugiere y orienta. Por ejemplo, en los estudios de DOLL y HILL acerca de la génesis del cáncer pulmonar y el cigarrillo, las relaciones encontradas (OR de 14) colocan al investigador ante la sospecha de haber encontrado un factor CAUSAL muy importante (más de 4 en el RR y en el OR hacen sospechar causalidad). De hecho, hoy se habla de que el cigarrillo es responsable de al menos el 90% de los casos de cáncer pulmonar. Pero iojo...! no se puede afirmar que sea el responsable del 100%, ya que existen casos de cáncer de pulmón en personas que no fuman (pocos... pero existen). Por ende, el cáncer de pulmón puede ser desencadenado por otro (s) elementos distintos al cigarrillo. Es por ello que hoy se plantea que la enfermedad sea el resultado de una combinación de factores de riesgo (elemento que de estar presente favorece la aparición de la enfermedad). La tabla comparativa es entonces la encargada de detectar a través de las relaciones presentes dichos factores. Respecto al caso del cigarrillo y el cáncer de pulmón vale la pena hacer el comentario del mal uso de la falta de ese 100% de causalidad responsable (así sea 92%).

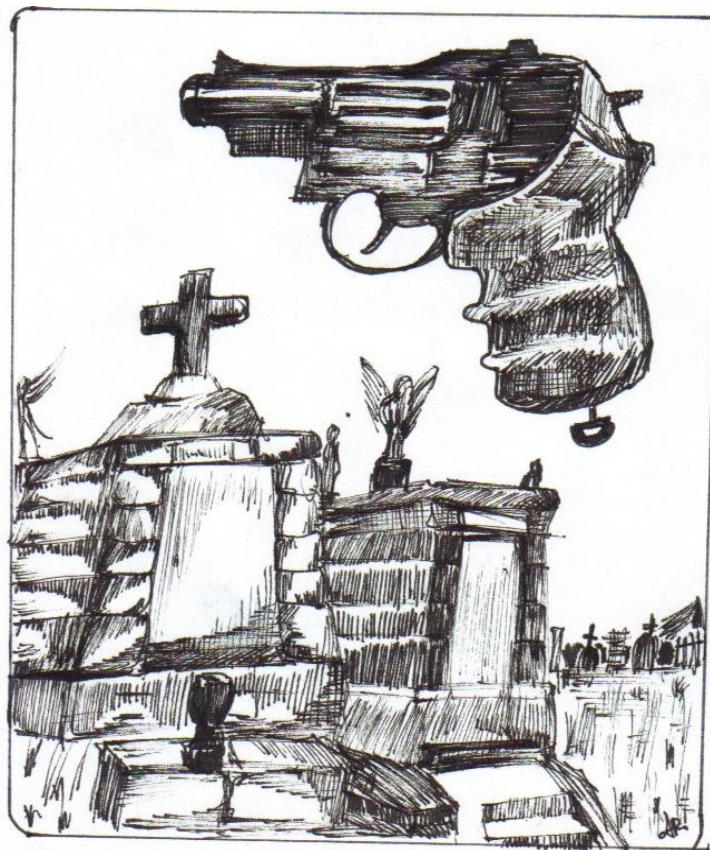


Las grandes tabacaleras han usado esa falta del 100% de la relación para poner dudas acerca de la efectividad de medidas como prohibir fumar, usando su poder para evitar que políticas restrictivas progresen.

¿Qué factores intervienen en la enfermedad? Una vez más se acude a la historia natural. Se considera que pueden ser factores que emergen del huésped, o del agente (s) o del medio. En cada uno de estos aspectos, surgen muchos aspectos (relaciones) a estudiar. Una de las tareas de la epidemiología será buscar estos factores.

Una relación es el resultado de la aplicación de la tabla de comparación y en donde se encuentra que la medida encontrada se ubica o siempre (siempre = resultado obtenido de la estadística) como riesgo o como factor protector, para diferenciarlo de aquellos casos en que el resultado se puede ubicar unas veces en riesgo y otras en protección (resultado de la aplicación de alguna de las pruebas estadísticas, por ejemplo, los

intervalos de confianza al 95%). Cuando se encuentran las relaciones, se le aplican los criterios para saber si son **asociaciones estadísticas** o no (pruebas estadísticas).



Es muy evidente la relación que existe entre muerte y armas de fuego. Esta relación se ve en una tabla comparativa o de contingencia en una población, donde se buscaría: variable dependiente 'Sí muerte' y 'No muerte' y como variables independiente 'Sí uso de arma de fuego' y 'No uso de arma de fuego'. Por supuesto que la medida de la relación (OR) daría muy seguramente un valor muy alto. Si gusta... hágalo.

La asociación estadística es entonces el resultado de una relación que tiende a mantenerse ya en riesgo o en protección en el 95% de los casos en que se combinan las variables de donde se origina. A manera de ejemplo, la asociación entre el cigarrillo y el cáncer de pulmón en el estudio de Doll y Hill fue estadísticamente significativa... es decir, que con esos datos se presentaría al menos 95 veces de 100 (siempre daría

que el cigarrillo se asocia con el cáncer). Detectadas estas asociaciones se verifica si son CAUSALES o no. Acá se hacen una serie de preguntas (**Criterios de Bradford Hill**) que buscan identificar que se puede estar ante la identificación de una de las causas de la enfermedad.

Por supuesto que solo se pueden contestar si el diseño del estudio de investigación lo permite. Doll y Hill en 1950 encontraron la relación entre el cigarrillo y el cáncer (en un estudio de casos y controles). Luego y tomando muchos más estudios se debe analizar cómo se contestan las siguientes preguntas que se hacen en las asociaciones causales. Como se trata de dos variables, entonces las preguntas son referidas a estas dos variables:

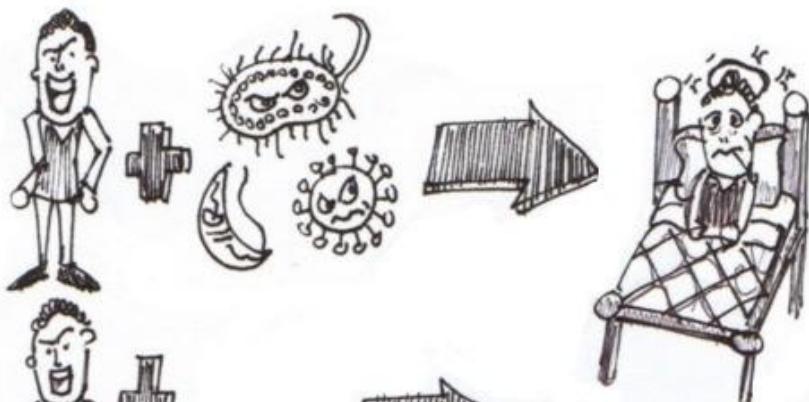
- 1) **El cambio de una de las variables provoca el cambio de la otra:** en el caso de cigarrillo y cáncer: si se deja de fumar (disminución de la frecuencia) se provocan cambios en la frecuencia del cáncer. O por el contrario, si se incrementa la frecuencia de fumadores se incrementa la frecuencia de cánceres del pulmón.
- 2) **Debe demostrarse que la variable CAUSAL precede a la enfermedad.** En esta caso, que no se tiene el cáncer antes de fumar y que aparece después de fumar. Los estudios de cohorte desarrollados confirmaron esta relación y en este sentido.



Acá es necesario, en primer lugar, tener la certeza de que nadie de la población a estudiar tenga cáncer del pulmón y luego (a través del tiempo que va pasando) se van ubicando los que fuman a un lado y los que no al otro y se mide a lo largo del tiempo en qué lado aparece más cáncer. De esa manera se tiene la certeza que primero fue la exposición al cigarrillo y luego apareció el cáncer.

- 3) **Firmeza de asociación (dosis respuesta).** Parecida a la condición numero 1, expresa que existe proporcionalidad en los cambios de frecuencia entre las dos variables. Si fuman mucho más (más dosis), aparecen muchos más canceres o viceversa.
- 4) **Reversibilidad:** Expresa que si se disminuye la variable causal disminuye la enfermedad. Si se deja de fumar disminuye probabilidad de que aparezca el cáncer.
- 5) **Consistencia de la asociación con el conocimiento vigente.** Existen más investigaciones que impulsan la relación. En este caso son muchísimas las investigaciones que muestran la relación entre el cáncer y el cigarrillo. Igualmente puede ser útil la búsqueda infructuosa de investigaciones en donde no se encuentra la relación.

1.



2.

ANTES DE EXPOSICIÓN

3.



4.



5.





Por ejemplo, en el caso de la tuberculosis:

1. Relación entre enfermos (manifestaciones) y presencia del bacilo.
2. En sanos no se encuentra el bacilo.
3. Entre más enfermos se encuentra más bacilo.
4. Si se quita el bacilo entonces se revierten las manifestaciones.
5. Que las investigaciones publicadas muestran al bacilo como el agente etiológico de la enfermedad.

Una aclaración pertinente de las **relaciones causales con las estadísticas** es que pueden existir relaciones estadísticas pero no causales. ¿Cuáles son? Pues las que no cumplen con los criterios expuestos. Por ejemplo, se pueden encontrar una relación estadística entre cáncer de pulmón y dedos amarillos por impregnación de nicotina. Pero no da cáncer por tener los dedos amarillos impregnados de nicotina. El cáncer da por fumar que es lo que provoca en algunos fumadores que los dedos se impregnen de amarillo. Como se observa se encuentra una fuerte relación que incluso se repite estadísticamente, pero no es causal. Esta desconcertante (y frecuente) afirmación, denominada **relación causal espuria** solo pretende colocar ante la perspectiva del lector que el conocimiento es cambiante y que los criterios que hoy usamos para entender los hechos, pueden cambiar mañana (y con ello lo que consideraremos la verdad).

Por ello se debe tener sumo cuidado con las personas que se creen poseedoras de la verdad absoluta. La evidencia dice que suelen estar equivocadas. La verdad científica es muy relativa y puede cambiar en la misma forma que se hacen nuevos descubrimientos.

Otro ejemplo (y para que se vea que es más frecuente de lo que se podría creer), es la asociación que encontró una fuerte relación entre tomar café y el cáncer del pulmón. Estadísticamente estaba presente y de manera notoria. Sin embargo, la realidad demostró que no era el café un factor de riesgo para el cáncer. Lo que sucedió es que en ese estudio, mucha población que tomaba café (tinto) también fumaba, siendo este último elemento el verdadero responsable de la relación. Por supuesto que al separar los que toman café y fuman de los que toman café y no fuman, se observó que el cáncer estaba presente en los que fuman, y no se observaba es lo que toman café y no fuman. El primer resultado (relación entre tomar tinto y cáncer) da pie a lo que se conoce como un **sesgo de confusión**, definido como una relación que, o bien aumenta o disminuye, a otra que sí es el factor de riesgo. Lo que se hizo al separar los que toman café y fuman de los que lo toman y no fuman se conoce como **estratificación**.



El sesgo de confusión es una relación que siendo significativa no se le encuentra alguna explicación plausible. Por ejemplo, en la investigación de Falbo sobre homicidios de adolescentes en Brasil, se encontró que en los hogares donde existía agua del acueducto, existía un OR de 2.8 con (IC del 95% de 1,7-4,5) de muerte de los jóvenes. Ello quiere decir que por cada joven asesinado y en cuyo hogar no existe agua potable, asesinan a 2,8 jóvenes en cuyos hogares existe agua potable ¿Cómo explicar este hallazgo?...

Estratificar es diferenciar y es una excelente forma de buscar la validez de lo que se compara. Por ejemplo, si se quiere medir la eficiencia de un programa de educación sexual en dos colegios y se usará el inicio de la vida sexual entre los jóvenes como medida, resulta muy importante diferenciar a los jóvenes por edades (estratificación por edades), pues es evidente que entre más años, más posible el inicio de la vida sexual. Al hacerlo (estratificar) se separan los jóvenes por la edad construyendo la condición que permite comparar jóvenes de dos colegios y con una misma edad.

Otro criterio a tener en cuenta en las asociaciones son las denominadas asociaciones directas e indirectas, clasificación que únicamente pone bajo la perspectiva los conocimientos vigentes. Por ejemplo: hacia mediados del siglo XX se encontró esta asociación: como tratamiento de la sífilis con salvarsán (medicamento mercurial) se observó una asociación con ictericia, luego de aplicado el tratamiento.

Se podía decir que existía una relación entre el medicamento y la ictericia. Hasta ahí existía relación causa (inyecciones de salvarsán) y efecto (ictericia) evidente. Luego se descubrió que la ictericia no se presentaba por el medicamento: se presentaba por la jeringa! Esto quiere decir que si se obtienen las relaciones entre ictericia y el uso de la jeringa, éstas son mayores que por el uso de salvarsán (podía haber ictericia en personas que recibieron inyecciones distintas al salvarsán).



Y finalmente, luego se descubrió que lo que provocaba la ictericia eran partículas de suero que quedaban en las jeringas (podía haber ictericia en personas que no eran inyectadas sino trasfundidas). Por tanto, era el suero y no las jeringas. Hoy sabemos que no es el suero, sino la presencia del virus de la hepatitis B y el uso de jeringas no estériles (por ejemplo en población de drogadictos a propósito del dibujo) que, junto con la promiscuidad sexual, son los principales factores de riesgo. Lo cierto es que todas las relaciones encontradas, son indirectas (no son las directamente responsables) de la ictericia a excepción de la verdaderamente CAUSAL que es el contagio con el virus de la hepatitis B. Esta es la postura científica vigente del momento... a lo mejor mañana descubrimos que el virus es apenas un transporte de otra particular que es la verdadera responsable de la hepatitis. Otra manera de expresar estos hallazgos es que las asociaciones producto de las jeringas en ese estudio corresponde a un **SESGO DE CONFUSIÓN**.

2.2.2. La red causal

¡Cómo sería de sencillo el mundo médico si la enfermedad fuese el resultado de una sola causa! Todo el actuar de la medicina estaría enfocado a ese aspecto. Pasteur, al iniciar el principio de la etapa bacteriológica, planteaba que a una enfermedad le correspondía un agente bacteriano como causa fundamental para poder explicar su origen. Gracias a ello se desarrolló el pensamiento capaz de enfrentar las temibles infecciones.



Robert Koch (1843-1910) propuso para poder confirmar que un germen es responsable de una enfermedad que debería cumplir con estos cuatro principios:

1. *El microorganismo (o germen) debe estar presente en cada uno de los casos de la enfermedad.*
2. *El germen debe ser aislado y crecer en un cultivo puro.*
3. *El germen cuando se inocula en un animal susceptible produce la enfermedad.*
4. *En el caso anterior (inoculado), el germen debe poderse aislar nuevamente.*

Hoy se diagnostican las infecciones con la búsqueda del germen a través de aislamiento directo o por la detección de su presencia a través de anticuerpos o antígenos, razón por la cual la inoculación en animales se ha disminuido.

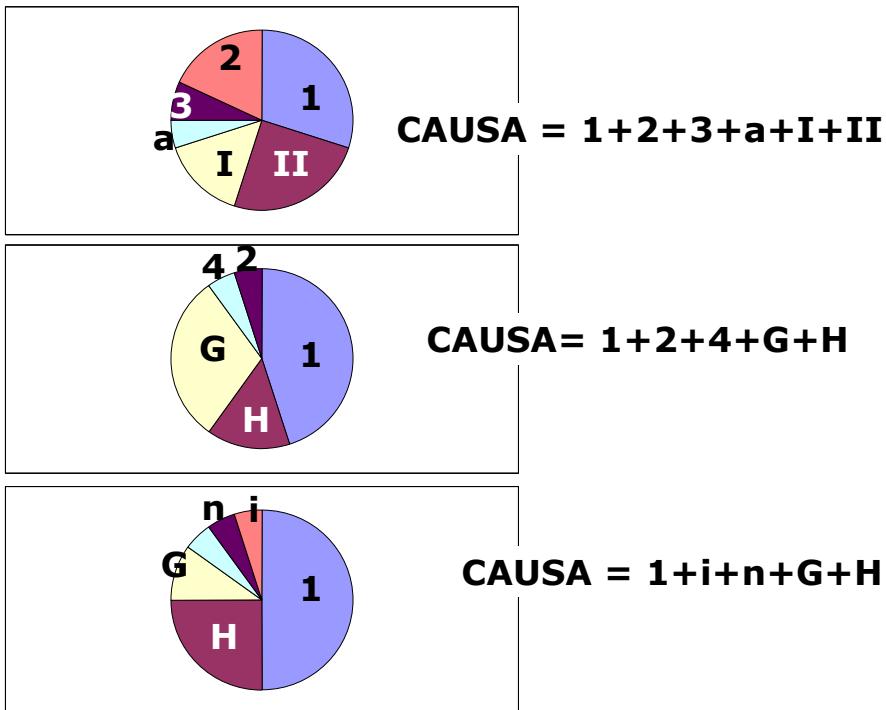
Sin embargo, la realidad pone de manifiesto que son muchas las situaciones que participan en la génesis de las enfermedades. Retomando el ejemplo de la ictericia por el salvarsán, todas las asociaciones encontradas tienen una participación en el proceso de la enfermedad. Si no fuera por la sífilis, no le hubiesen colocado el salvarsán y, por consiguiente la inyección que por no estar debidamente desinfectada, es que ocurre la contaminación, con lo cual, quedan partículas de suero que de otra manera no estarían. Así que esas asociaciones indirectas de todas maneras tienen una participación en la génesis de la enfermedad (de hecho por ello aparecen como relaciones). Con una visión más amplia, téngase en cuenta la triada ecológica. La enfermedad se desencadena por alteraciones en los tres aspectos (agente, huésped y medio). De seguro unos elementos son más evidentes que otros, pero todos participan (**teoría de la complejidad**). Esto se conoce como la red de la causalidad o trama. A través de ella se pretende ver la participación de los distintos componentes para desencadenar la enfermedad.



Una interpretación de la teoría de la complejidad es la de entender a la realidad como una suma de tal cantidad de eventos (relaciones) que se hace realmente difícil poder analizarlos a todos. Es más, con el factor tiempo de por medio, nada es repetible. Cada momento es único y por completo diferente. Una relación se presenta por unas condiciones que no se repetirán.

2.2.2.1. Las causas necesarias y suficientes

El modelo de la red causal quedó insuficiente para efectos prácticos. Se plantea que en la medida que investigamos surgirán más y más elementos causales; por lo cual no aporta tanto como se podría esperar para las soluciones. De ahí que el Dr. Rothamn ha propuesto que las enfermedades sean vistas a través de constelaciones (combinaciones de componentes causales) suficientes para desencadenar la enfermedad. Se podría intervenir sobre dichas constelaciones. Pero, además, propone buscar en las constelaciones (que él expuso como pasteles causales) aquellas causas presentes en todos los pasteles. Propone que sería la causa necesaria (todos los demás serían suficientes solo si se mezclan) y al intervenir sobre ella se puede evitar que el pastel se complete y, por tanto, la aparición de la enfermedad.



La gráfica expresa a tres pacientes en los que se desencadenó la misma enfermedad. A cada uno (superior, medio e inferior) la enfermedad se les provocó por la exposición a diversos factores, pero existe uno que fue común a los tres: ¿Cuál es?... Ese probablemente sea el factor (la causa) necesaria, pero la enfermedad sólo se desencadena si todos se suman. Si los equipos de salud impiden que algunos de los factores se unan, pueden impedir que aparezca la enfermedad (que se complete el pastel causal). ¿Cuál se ha de buscar en el ejemplo? De seguro el componente 1; ¿por qué? Por estar participando en los tres casos... si se logra impedir que se desarrolle (o aparezca), entonces en ninguno de los tres casos el pastel se completará y la enfermedad no aparece.

De las anteriores posturas sólo queda en claro la complejidad del mundo relacional. Hoy se puede mencionar clasificaciones como la siguiente:

Factores predisponentes: aquellos que deben estar presentes para que aparezca la enfermedad. Claro que pueden estar presentes y NO aparecer la enfermedad. Sexo, raza, herencia son ejemplos.

Factores facilitadores: aquellos que pueden favorecer la aparición de la enfermedad. La pobreza o la alimentación deficiente son componentes que favorecen la aparición de casi todas las enfermedades. Es claro que de intervenir sobre ellos se disminuye la posibilidad de que aparezca la enfermedad.

Factores desencadenantes: se refieren a aquéllos que cuando se presentan pueden desencadenar la enfermedad. Un estado de inmunosupresión puede desencadenar la aparición de un herpes zóster. Sin duda, modificarlos es importante para evitar la aparición de la enfermedad.

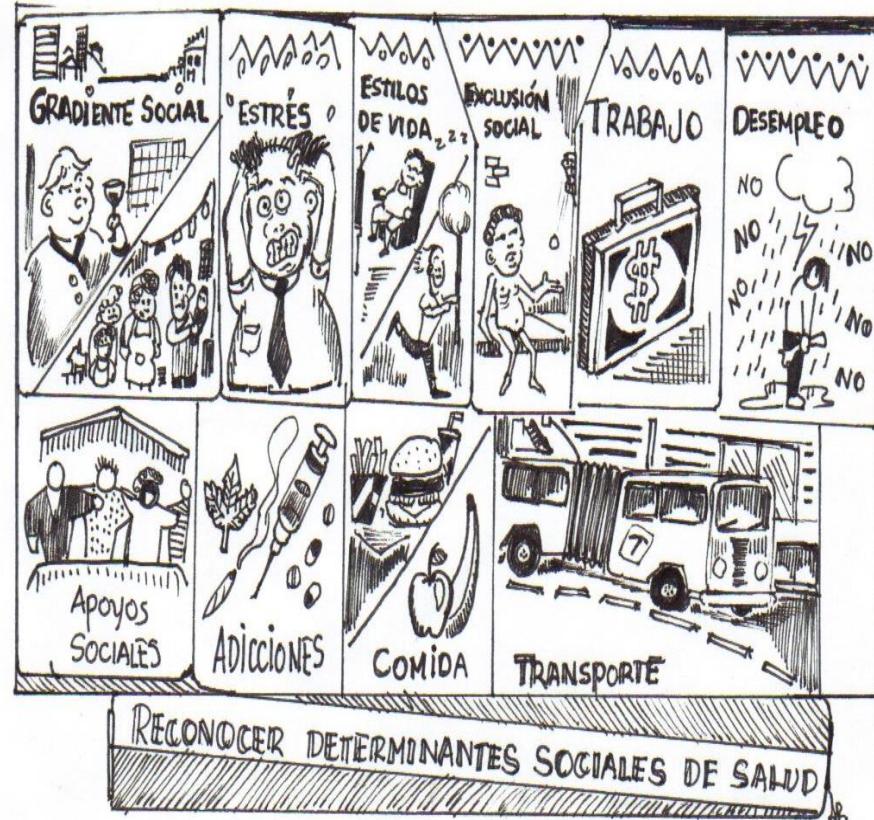
Factores potencializadores: se refieren a componentes que pueden agravar una enfermedad. Por ejemplo, una reinfeción del Virus del VIH puede agravar un cuadro de SIDA en un paciente.

Interacción: se considera que es el resultado de dos factores que actúan simultáneamente, desencadenando un efecto sumado mayor que el de una simple suma. Por ejemplo, fumar y exponerse a asbestos con el cáncer de pulmón. El resultado es una probabilidad mucho mayor de cáncer que el producto de la suma de los dos riesgos independientes.

Exposición al asbestos	Fumadores	Tasas de cáncer pulmonar
NO	NO	11
SI	NO	58
NO	SI	123
SI	SI	602

Un buen ejemplo lo expresa la tabla obtenida de la investigación de Hammond y colaboradores en 1979 en la cual se refleja cómo la interacción entre exposición de asbestos y fumar se relaciona con unas tasas (por 100.000) casi tres veces mayores que las que se obtendrían de sumar cada una de las exposiciones (la suma de sí asbestos=58 + Sí fuma=123 es 181 cifra aproximadamente menor en un tercio a los 602 obtenidos por la interacción Sí asbestos y Sí fumador).

La búsqueda de ese factor sobre el cual de modificarse se evita que aparezca la enfermedad es la dirección que sigue la epidemiología al unirse con la salud pública.



Como un interesante aporte en este sentido, hoy se habla de los **determinantes sociales de la enfermedad**. La inequidad social, el estrés social desencadenado por la inseguridad o por el aislamiento, las privaciones en la infancia en torno afecto o alimentación (pobreza extrema), la exclusión social (estigmatización), la mala calidad del transporte, son todos elementos presentes en los enfermos. Como elementos protectores se mencionan el acceso a la educación, el empleo, la ciudadanía y la pertenencia a redes sociales. El concepto afirma que estos elementos (y demostrado a través de numerosas investigaciones,) participan en el proceso de salud y enfermedad de los seres humanos. ¿Cuánto y cómo? Es lo que hoy muchos están indagando.

2.3. Buscando la salud pública: La vigilancia epidemiológica:

El objeto del conocimiento es servir al ser humano. La epidemiología de manera muy particular ha servido a lo largo de la historia de la humanidad en su lucha contra las enfermedades, en especial contra las infectocontagiosas, pues a través de su detección, impulsó al hombre a tomar medidas para evitar la propagación, a veces, sin éxito, como las medidas tomadas para evitar la peste en el medioevo... pero a la postre, logrando el control de tantas entidades. Hoy esta lucha está ejemplificada en torno al SIDA como representante de las infecciosas y se abren nuevos modelos de enfermedades tales como la diabetes o la obesidad, donde los estilos de vida se convierten en factores desencadenantes. El proceso de detección, de análisis de por qué aparece y de respuesta para contrarrestar la enfermedad son los componentes que forman parte de un sistema de vigilancia epidemiológica.

Aún cuando parece obvio, la vigilancia epidemiológica no es tan simple pues exige tener una **fuente de registros** que permita ver los cambios, circunstancia no siempre presente. Los registros exigen constancia (continuidad) y calidad (veracidad), condiciones no siempre presentes en torno a problemas específicos. Un buen ejemplo es pensar en lo que pasó en 1980 en torno al SIDA. No existía un sistema de registro con la capacidad de identificar a la entidad (de hecho ni siquiera se conocía). Ello implica que podían presentarse los casos, pero no nos dábamos cuenta. Fue gracias a las observaciones de un médico norteamericano (Michael Gottielb) que alertó a los servicios de salud acerca de una enfermedad que mataba a gente joven como consecuencia de una severa inmunosupresión y gracias a ello la humanidad pudo darse cuenta que estaba ante una **PANDEMIA**.

La importancia de los registros de información es evidente. Principios que deben tener: **exactos** (fiel reflejo del evento). **Oportunos** queriendo con ello decir que se llenen cuando sucede. **Fidedignos**, es decir, lo más aproximado a la realidad. **Completos**, llenando todos sus componentes. No sobra insistir en el papel que le corresponde al personal de salud en la calidad de los registros.



Resulta obvio el valor de registrar información que nos sirva de indicadora. El asunto es registrar. Para ello el encontrar el elemento que mejor nos indique es clave. Claro, dicho elemento debe haber sido analizado para poder saber qué pensar en caso de que aumente o disminuya su frecuencia. De ahí surge el término de vigilancia, pues es vigilando la frecuencia de tal elemento que se establecen los sistemas de alertas.

2.3.1. La importancia de los registros

Sin duda el primer paso para resolver un problema es reconocer su existencia. En el caso de las enfermedades el reconocimiento se lleva a cabo a través de la detección de las mismas por medio de sus frecuencias. De hecho el incremento inusitado de alguna de estas frecuencias es lo que se conoce con el nombre de epidemia.

La estrategia para obtener información sobre las frecuencias de las enfermedades surge de los sistemas de registros. Existen una gran variedad de los mismos: registros médicos, registros demográficos, específicos etc.

Las fases de un sistema de registro implican la recolección de los datos, la tabulación de los mismos y la presentación de los informes pertinentes (primera fase). La segunda fase implica el análisis el que se lleva a cabo a través de la descripción encontrada y la comparación que de ahí se deriva. Con base en ello se proponen estrategias para combatir lo encontrado.

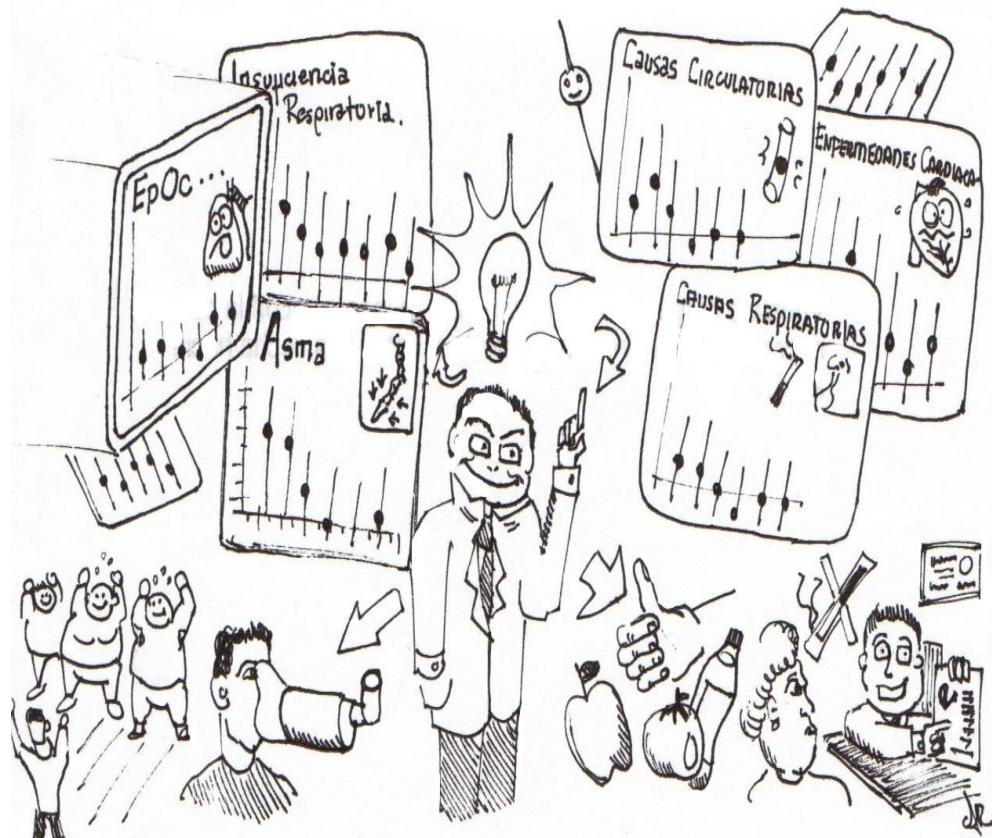
La fuente de la información da pie a distintas clasificaciones de los sistemas de vigilancia. Si la fuente de información se obtiene a través de registros no desarrollados por la epidemiología, se denomina **vigilancia pasiva**. Ejemplo es la detección a través de estadísticas demográficas, obtenidas de fuentes distintas o no a la salud, en donde no existe el criterio de búsqueda específica de algo. Si por el contrario, son las oficinas de epidemiología las que a través de sus sistemas de registro buscan una información y ellas se encargan de recoger los datos, se denominan **vigilancia activa**. Adicionalmente, los epidemiólogos pueden construir un sistema específico para un problema, lo que se denomina **vigilancia epidemiológica especializada**.



En la actualidad numerosas instancias se ocupan de hacer seguimiento a la información. Los comités de vigilancia epidemiológica en las IPS (Instituciones Prestadoras de Servicios de Salud) se encargan de ofrecer la información pertinente a los organismos sanitarios pertinentes. Por ejemplo, toda IPS de segundo nivel en adelante ha de tener un comité de infecciones. Su tarea es analizar las frecuencias de infecciones presentes en el hospital. Pero no los que le llegan a través de los pacientes. Analizan las que surgen dentro del hospital. ¿Por qué es importante? Sin duda por las repercusiones en los pacientes (en caso de que se contaminen, se desencadenan infecciones muy agresivas), pero además, porque a través del trabajo de dicho comité se puede analizar el control establecido dentro del hospital, siendo un INDICADOR de calidad en la atención. La calidad es en este caso el resultado de las acciones que protegen de infecciones nosocomiales a la población que hace uso de dicho hospital.

2.3.2. El arte de la predicción

El objetivo de los sistemas de vigilancia epidemiológica es predecir (antes de que suceda), con base a la información y el análisis, el impacto de las enfermedades en la población y proceder en consecuencia. Para ello, y como estrategia, puede usar **sistemas de vigilancia intensificados**. Por ejemplo, detección de casos de rabia, para así iniciar acciones conducentes a evitar la propagación de un episodio (brote). Cualquier centro que reciba un caso de persona mordida por un canino ha de informar. También puede usar la **detección de eventos trazadores** que indican de manera indirecta la probable aparición de un problema. La detección de perros sin vacunas siguiendo el ejemplo anterior, actúa como evento trazador ya que claramente expone el riesgo sin que se presenten los casos. Otra estrategia son los **observatorios epidemiológicos**, en donde uniendo entidades (intersectoriales) se construyen sistemas de recolección de información a través de los cuales se puede detectar una problemática específica. Por ejemplo, construir con los colegios y los centros de salud un sistema de vigilancia de embarazos de adolescentes. Al hacerlo, se puede actuar y así evitar que aparezcan otros casos.



 La información es clave en la toma de decisiones. Existen algunas enfermedades que sirven para mostrar distintos aspectos de la comunidad (y sirven porque se emplean para comparar). Son entidades INDICADORAS o TRAZADORAS. A través de ellas se puede comparar. Por ejemplo, la mortalidad se puede comparar por años o luego de implementar políticas. Claro, los cambios de mortalidad se toman tiempo. Los ingresos de urgencias o los ingresos hospitalarios son fuentes de registros a través de los cuales se puede obtener mucha información para hacer análisis. La forma como se presente la información es también clave. El ejemplo del dibujo no es claro... sin embargo, pretende mostrar unas gráficas donde se informa sobre el estado de unas condiciones de salud (en este caso enfermedades)...

2.3.3. Apunte de la epidemiología a la salud comunitaria

La epidemiología en sí es comunitaria. Busca contribuir a solucionar problemas del área de la salud y del bienestar de la población. Usa el método científico para enfrentarse a los problemas. Parte de observaciones originadas, entre otros, por los cambios de ocurrencia de las enfermedades y busca detectar qué ocasiona dichos patrones para, a través de ese conocimiento, proponer alternativas de intervenciones que disminuyan las frecuencias. Para todo esto se mezcla con otros conocimientos y lo hace de manera integral. Buscando facilitar el entendimiento muchas veces los seres humanos separamos los conocimientos, pero la realidad es una mezcla de saberes que, de alguna manera, permiten entender el concepto de la complejidad de la que hoy los epistemólogos tanto hablan.

Las sociedades humanas son cambiantes y con sus cambios se enfrentan a nuevas situaciones, muchas de las cuales pueden tener fuertes relaciones con la aparición de enfermedades. Por ello el vigilar los denominados **indicadores de salud** de la comunidad ha de estar presente en cualquier sociedad. A través de cambios en ellos se puede detectar la aparición de problemas. Un indicador es un dato que refleja una situación. La mortalidad infantil es un extraordinario indicador del estado de salud de un país que permite al ser revisado, ser comparado con el de otros y así sacar conclusiones.





Un **indicador** es un dato a través del cual se pueden deducir otras condiciones. Por ejemplo, a través del peso se puede deducir el estado nutricional. En epidemiología existen datos que permiten interpretar el estado de salud de la comunidad.

Índice: indicio o señal de algo.

Más adelante en el numeral 4.2 se ampliará el concepto de indicadores de salud.

Una de las estrategias que más se está usando en los enfoques de salud pública actual es la promoción de estilos de vida saludables, situación que forma parte de uno de los capítulos donde la epidemiología aporta sólidos conocimientos, como es la **MEDICINA PREVENTIVA**. Dicha promoción junto con la reducción de condiciones que favorecen la aparición de las enfermedades (por ejemplo, disminución del consumo de cigarrillo para evitar la aparición de muchísimas enfermedades claramente relacionadas) y acceso a servicios sanitarios, completan el capítulo. La vigilancia de las enfermedades constituye el segundo capítulo del gran aporte que la epidemiología le hace a la comunidad.

Rey Calero sostiene que la estrategia actual de la medicina preventiva se basa en tres aspectos:

1. Promoción de estilos de vida higiénicos.
2. Reducción de condiciones evitables.
3. Provisión adecuada de cuidados sanitarios accesibles a todos.

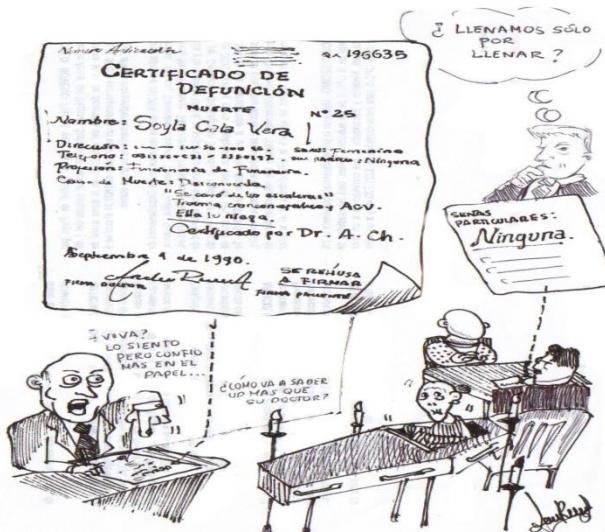
De ahí la importancia de enseñar a muy temprana edad qué son los estilos saludables, buscando que se conviertan en hábitos. La educación, sin duda, es una estrategia válida para ello. Reducir los riesgos impuestos, por ejemplo los derivados del alcohol y del tabaco, es por demás lo esperado. La legislación sanitaria ayuda enormemente y para que las leyes tengan posibilidad, la comunidad debe promoverlas a través de sus posturas. Y, por supuesto que la búsqueda de un servicio sanitario equitativo y eficiente, es la gran ayuda para hacer efectiva la visión de la medicina preventiva. El acceso y el enfoque resultan acá particularmente determinantes.



Hoy como nunca, tenemos los conocimientos que permiten identificar qué provoca enfermedades y qué produce salud. Sin embargo, y ahí una vez más está la paradoja del ser humano, permitimos la promoción masiva de productos que desencadenan enfermedades. ¿Por qué? Sencillamente porque esos productos provocan el muy mal denominado desarrollo económico, donde se nos hace creer que el ser humano se beneficia más de la producción que del daño que provoca lo que produce. Acá los conflictos de interés entre la ambición personal y el bien común se enfrentan y en un inmenso número de casos, la ambición personal usando la corrupción, se impone. Por ejemplo, las tabacaleras quieren hacer creer que fumar o no, es una decisión individual y que se debe permitir a las personas tomar esa decisión. Pero, ellas a través de la nicotina del tabaco, crean adictos, lo que hace que la libertad de escoger se pierda por la adicción que obliga. El punto que se quiere promover es que existe un deformante (y manipulado) uso de la LIBERTAD de escoger.

2.3.3.1. La estrategia de la identificación

Para poder solucionar un problema, primero se debe reconocer. Esta frase lleva implícita una verdad poderosa. Es más probable solucionar lo que se reconoce. Por ello la epidemiología aplicada a la salud pública se esfuerza por tener un apropiado sistema de información a través del cual pueda detectar (reconocer) problemas o potenciales problemas. Para ello utiliza los registros. Un registro es un sistema de recolección de información. El registro se llena con los datos. Con base en los registros se presentan los informes que son consolidados de éstos a veces con un análisis (interpretación de esos resultados). Existe una gran cantidad de registros. Solo por mencionar algunos están los de morbilidad (por ejemplo, los RIPS), están los de mortalidad (certificados de defunción) y claro, están los registros demográficos o vitales (nacimientos, etc.). Lo cierto es que a través de los registros, se puede identificar, observando las frecuencias, cómo es el estado de salud de una comunidad. Esto ofrece el referente necesario para poder establecer modelos de vigilancia o de evaluación de intervenciones.



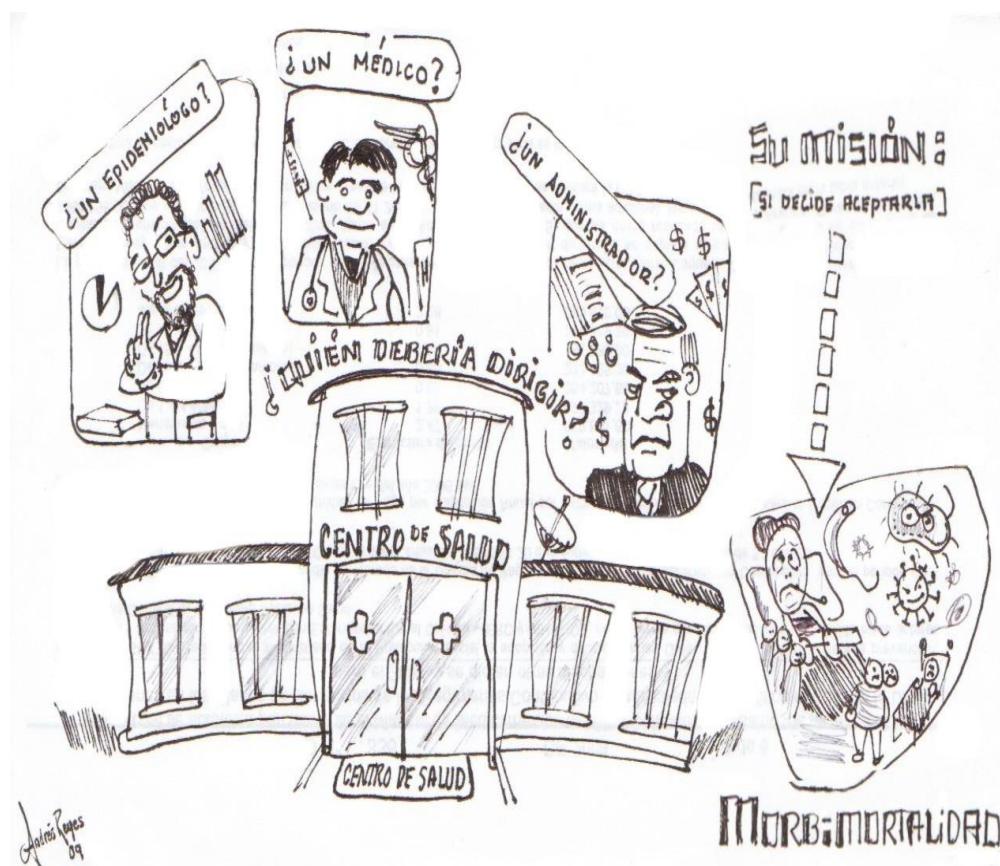
El personal de salud ha de comprender la inmensa responsabilidad que le atañe en torno a los registros. El llenarlos de manera completa y veraz (válida) es esencial para la construcción de la información con base a la cual se tomarán las decisiones. Por ejemplo, al llenar los RIPS. A través de ellos se alimenta el sistema de información de la Seguridad Social. Si se llenan mal los códigos de enfermedades, entonces el perfil mórbido con base en el cual se hacen los análisis será falso y, a partir de ahí, todas las decisiones que se tomen con base a dicha información. Lo propio se puede decir de los certificados de defunción. Con base en ellos se construye la mortalidad del país. Llenarlos de manera veraz es vital para el sistema de información. Claro, no siempre ello es posible. En Colombia muchos certificados se llenan de manera presuntiva (no son los patólogos luego de las autopsias, o en su defecto, los médicos tratantes quienes los llenan). Ello quiere decir que no se tiene la certeza que debe tener un buen sistema de información.

2.3.3.2. Para decidir

Con la información, se debe pasar a hacer un análisis de la misma. Analizar es comparar tomando referentes. Así mismo, se pueden detectar problemas. Por ejemplo, la aparición de epidemias. El resultado del análisis es llegar a la toma de decisiones. Una decisión es una acción propuesta, que se hace con base a una información y que debe solucionar una situación que lo amerita. Por supuesto que si la información es veraz, las posibilidades de éxito de las decisiones al ser ejecutadas son mayores. Para decidir, se han de tener en cuenta diversos elementos. Las decisiones como tales son teóricas (son probables respuestas no probadas) y se espera que como resultado de su aplicación, el beneficio supere los daños probables. Cuando surgen las decisiones se tienen en cuenta diversos componentes, entre los que está el uso de los recursos.

2.3.3.3. Para organizar

El resultado de las decisiones tiene como consecuencia la organización de los recursos disponibles para enfrentar los problemas. La organización de los recursos destinados a la salud da pie a la administración sanitaria. Dicha administración no es otra cosa que la forma como se manejan los recursos para dispensar salud a la comunidad. El medir los resultados se convierte en un importante eslabón de la epidemiología y de la salud pública. Términos empleados bajo este entorno son costo/eficiencia, análisis de coste/beneficio, efectividad.



👉 En una muy hermosa discusión en el texto "El desafío de la epidemiología" el Doctor Milton Terris expone que la orientación de los servicios de salud ha de estar en manos de la epidemiología. Esta disciplina a través de sus métodos y sus medidas tiene la capacidad de poder evaluar los resultados de las decisiones tomadas en torno a la salud y la enfermedad. Plantea que la medicina está basada en la terapia y que la administración solo mira recursos y solo la epidemiología trata de entenderlo todo. Por ejemplo, si se quiere medir la calidad de los servicios de salud, las actividades indicadoras han de ser la morbimortalidad. Los administradores pueden escoger percepción del paciente y los financieros costos del sistema, componentes que sin duda ven, pero, solo la morbimortalidad puede ofrecer una medida integral. Y dicha medida es trabajada por la epidemiología.

Por otro lado, no existe duda que las investigaciones administrativas son un buen reflejo de cómo el método epidemiológico puede aportar. Las investigaciones de procesos o los estudios de costos son un buen ejemplo.

2.4 Hacia dónde va la epidemiología. Mirando la evidencia de la utilidad

Es indudable el inmenso aporte que ha hecho la epidemiología a la humanidad. Su capacidad de organizar el pensamiento para enfrentar los problemas de salud de la población ha mostrado su efectividad a lo largo de la historia. En ello la búsqueda de las causas sin duda ha sido muy importante.



La epidemiología ha utilizado los recursos intelectuales de cada época de manera por demás interesante. A comienzos del siglo XX tomó la teoría microbiológica (o bacteriológica) y la aplicó con fuerza a sus métodos, proponiendo acciones trascendentales que en su momento tuvieron implicaciones formidables en la lucha del hombre contra la enfermedad. El enfoque causal fue ahí indispensable y el pensamiento de los factores determinantes de la enfermedad permitió, entre otros desarrollos, la construcción de medidas terapéuticas específicas. El mejor ejemplo es el desarrollo de los antibióticos en la lucha contra las enfermedades. Pero la realidad ha mostrado que con ellos, la humanidad está lejos de vencer a las infecciones. Muchos otros factores intervienen, factores que pueden tener una repercusión distinta de acuerdo a cada circunstancia en torno a la responsabilidad de la enfermedad (**red causal**). Por ello ha surgido el modelo conocido como la **caja negra** que expresa que prácticamente es imposible identificar todos los factores que pueden intervenir en la aparición de las enfermedades y, por ende, el esfuerzo

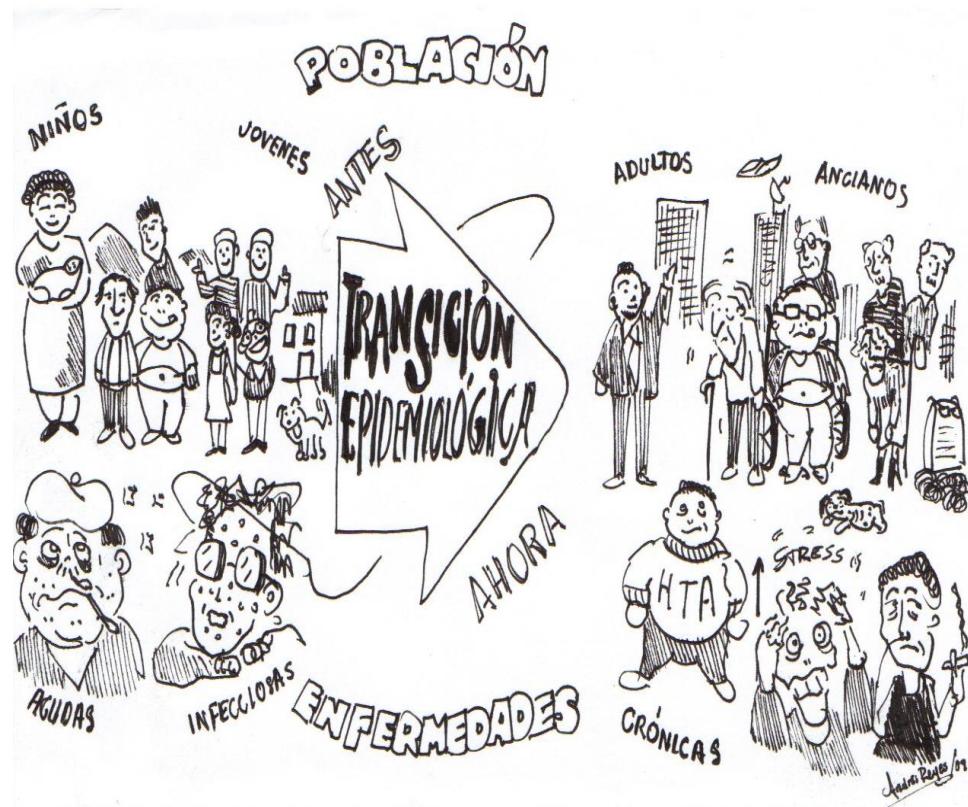
de la epidemiología debe limitarse a la búsqueda de aquellos en donde **la intervención puede romper la cadena causal** de manera efectiva. Pero esta evidente ganancia no es tan real. Al fin de cuentas sin entender en su totalidad muchas veces no ha sido posible detectar cuáles son esos componentes claves. Lo cierto es que más de una vez la epidemiología ha tomado el camino de buscar una inmensidad de factores causales (factores de riesgo) de la enfermedad sin poder influir en el desarrollo de la misma, que en esencia es lo que se busca. Como si fuera poco, la epidemiología se enfrenta a un serio interrogante: ¿los determinantes de la enfermedad de la población son el resultado de los determinantes individuales? Algunos autores, y con razones, creen que no (el todo es diferente a la suma de las partes). Todas estas dudas han permitido que surjan distintas e interesantes posturas. Por ejemplo, ha resurgido la postura de que en la aparición de la enfermedad poblacional, las condiciones socioeconómicas determinadas por el desarrollo político se convierten en factores dominantes (**medicina social**).



La influencia de las condiciones sociales es más que evidente en la salud del ser humano.

Desde el siglo XVII numerosos autores señalan la clara relación entre condiciones de miseria y enfermedades. Con ellos la visión de la salud pública hace su aparición. ¿Cómo combatir dichas condiciones? Dichas condiciones se desarrollan de acuerdo a los modelos políticos que dan pie a los modelos económicos. Unos modelos sostienen que el bienestar se alcanza a través del desarrollo económico privado, otros que se logra a través de una fuerte presencia Estatal. Lo indudable es que la salud de la comunidad se ve influenciada poderosamente a través de cualquiera de estas posturas, pues de acuerdo al dominio de una u otra postura se desarrollan los modelos sanitarios de cada país.

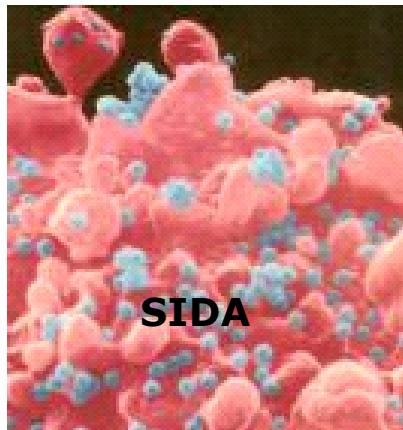
Lo cierto es que con los cambios de los perfiles de las enfermedades en la colectividad, cambian las posturas del pensamiento epidemiológico. Por ello, hoy se menciona que una parte (grande) de la humanidad empieza a padecer con intensidad (mayores frecuencias) de las enfermedades crónicas donde la triada ecológica está siendo sometida a un nuevo enfoque. A manera de ejemplo, ¿cuánto de la hipertensión es debida a la carga genética y cuánto a los estilos de vida?



Hoy se habla de la transición epidemiológica en torno a las enfermedades. Las infecciones que fueron hasta el siglo XX el azote de la humanidad (por su frecuencia) están siendo superadas por las denominadas enfermedades crónicas. El cáncer, las enfermedades cardiocerebrovasculares o las enfermedades respiratorias crónicas junto con la diabetes, ocupan el primer lugar en frecuencia en una población que incrementó su esperanza de vida (más personas mayores de 60 años). Todo esto, implica que se deberán diseñar nuevos modelos de atención...

2.4.1. Una mención de éxitos y limitaciones

La epidemiología ha mostrado significativos éxitos a lo largo de la historia, éxitos que muestran el inmenso poder de la capacidad intelectual para identificar elementos en torno a las enfermedades de la población. Esos éxitos se disminuyen cuando se mide el resultado de dichos pensamientos en la humanidad. Para expresar el punto obsérvese el caso del SIDA.



CON



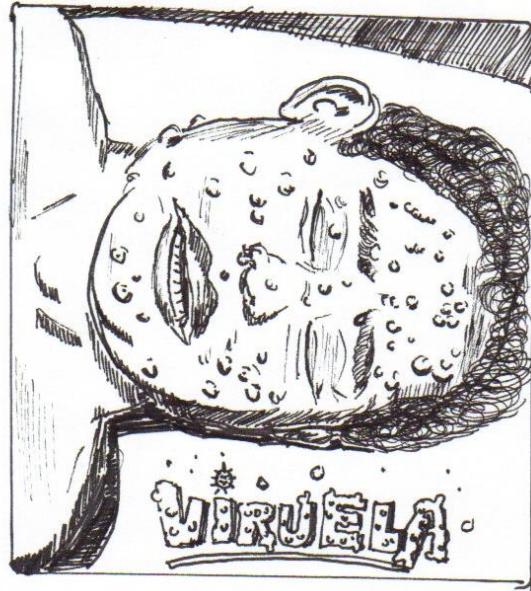
Desde su detección hacia 1980 hasta la certeza de su origen en 1984, la epidemiología orientó el enfoque hacia la sospecha de una enfermedad infecciosa. Estudios de casos y controles dirigidos por el CDC de Atlanta fueron los primeros en relacionar la promiscuidad como un factor de riesgo.

No existe duda del importante avance que se hizo acerca de esta enfermedad con base en los análisis epidemiológicos. Antes de haber identificado el virus responsable, ya se disponían de medidas de salud pública derivadas del uso de la epidemiología que, sin duda si se aplican, pueden dar como resultado el control de la pandemia. Pero, y por ello el éxito es relativo, isi no se aplican como se debería! Existen unas desigualdades sociales que hacen que el trabajo epidemiológico se convierta en más teoría que realidad. Una muestra de esto son los pobres resultados en los proyectos preventivos y el hecho que en la actualidad (2007) tan solo dos pacientes de cada 1000 en el mundo, reciban el manejo de antirretrovirales para controlar el SIDA.



→ Es como si el lema fuera: 'si lo paga, lo tiene'... absurdo postulado ya que la población que padece la enfermedad es en general pobre. Pero si no se controla la enfermedad en dicha población, la propagación persistirá. ¿Será que si los que tienen les financian el tratamiento a los que no tienen, el problema disminuye? La respuesta hipotética es que sí. Pero, es más, al hacerlo la humanidad gana. Un ejemplo, que permite plantearlo es el caso de la viruela.

En contraste, y como éxito mayúsculo de la epidemiología en combinación con la salud pública, está la erradicación de la viruela.



El compromiso político universal junto con el pensamiento epidemiológico provocó la desaparición de esta enfermedad. De hecho, una inversión en su momento de \$ 200 millones de dólares, determinó un ahorro de más de \$ 1.500 millones de dólares de los países ricos derivado al no requerir más de la vacuna.



Otro éxito ha sido el ver la relación entre pobreza y fiebre reumática. Claro, en este caso, las relaciones son más complejas que las derivadas de la intoxicación por metilmercurio, pero no menos importantes. Acá los factores sociales y económicos (pobreza) resultan ser precipitantes de la propagación del estreptococo. El control de dichas condiciones de pobreza determinó un control de esta entidad. Sin duda, el análisis de estas relaciones ha sido esencial en el surgimiento del concepto de los determinantes sociales de la salud.

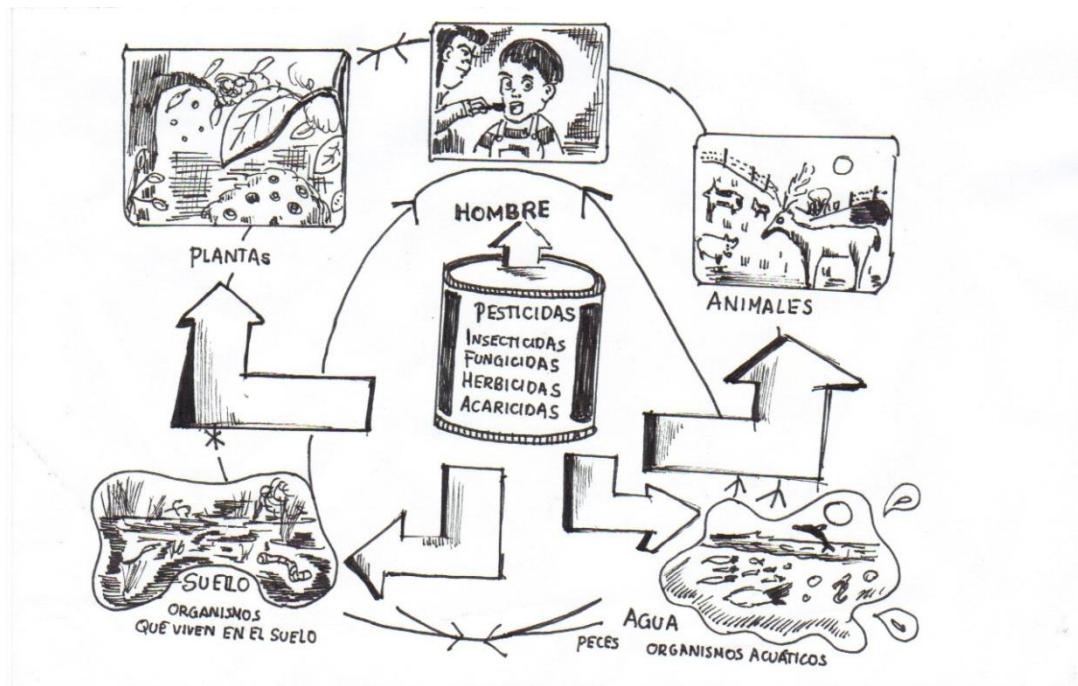
La epidemiología ha contribuido a que se comprenda que lo que se hace en un sitio repercute en otro. Por ejemplo, en los 50, una fábrica en Minamata (Japón) vertía residuos de mercurio en la bahía. El metilmercurio se acumuló en los peces y se desencadenó un

envenenamiento en las personas que comían estos peces. En un comienzo, la sospecha fue hacia una meningitis, pero cuando se descubrió que 121 pacientes residían en la bahía de Minamata (se encontró una relación orientadora de causalidad), el esfuerzo se orientó a analizar qué factor en común se podía encontrar. La detección de que todos eran pescadores dio la búsqueda determinante hacia el origen de la enfermedad. Este trabajo es un hito que hace comprender la salud ambiental.



¿Cuánto en la génesis del cáncer tendrá que ver con el contacto de tanto producto químico al que la humanidad se expone? Los colorantes usados para volver más apetecibles los alimentos, los fertilizantes, los residuos, son apenas algunos de los elementos en torno a los cuales se orientan la investigación epidemiológica en su búsqueda de la causalidad.

Así mismo, se debe reconocer que existan algunos casos en donde la epidemiología no ha logrado demostrar sus sospechas. Por ejemplo, existe evidencia científica del daño de los pesticidas para la vida (sospecha de daño ambiental) y, sin embargo, los modelos epidemiológicos no han podido demostrar que la exposición de los seres humanos (y otros seres vivientes) en las dosis no tóxicas es dañina; por ello se permite su uso. ¿Será que las medidas usadas no son las adecuadas?... O ¿será que falta más tiempo para que aparezca el daño?... esas son preguntas que muestran cómo a la epidemiología aplicada que se usa... le falta!



¿Será que los referentes que se utilizan (medidas) no pueden detectar todo lo que se busca? Por ejemplo, desde hace más de 20 años existe una fuerte sospecha en torno a la relación entre cáncer de cérvix e infección viral. En primera instancia la sospecha recayó sobre el herpes virus tipo II. En aquellos momentos no se podía detectar (medir) el papiloma virus. Cuando la tecnología permitió dicha medida, la sorpresa fue determinar la muy fuerte relación entre este virus y el mencionado cáncer. O tal vez que la **caja negra** permanece cerrada en numerosos áreas. Carol Buck es mencionada al plantear que una de las limitaciones de la epidemiología es el exceso que se le da al método. Plantea que la realidad no es experimental sino cuasiexperimental y, por tanto, esos son los métodos más validos que se han de utilizar.

Aún con esas limitaciones el aporte de la epidemiología es innegable. Hoy como nunca, la humanidad sabe qué se debe hacer en numerosas áreas. En salud se sabe que se ha de prohibir el abuso del cigarrillo o del alcohol o la proliferación de ciertas dietas (comidas rápidas). Las repercusiones deletéreas en la salud están más que demostradas gracias a la epidemiología. Sin embargo, no se hace! (al menos no con el vigor correspondiente). Entre las explicaciones que se pueden mencionar está el poder económico de los productores de aquellos elementos, que usan su poder para promover sus productos.

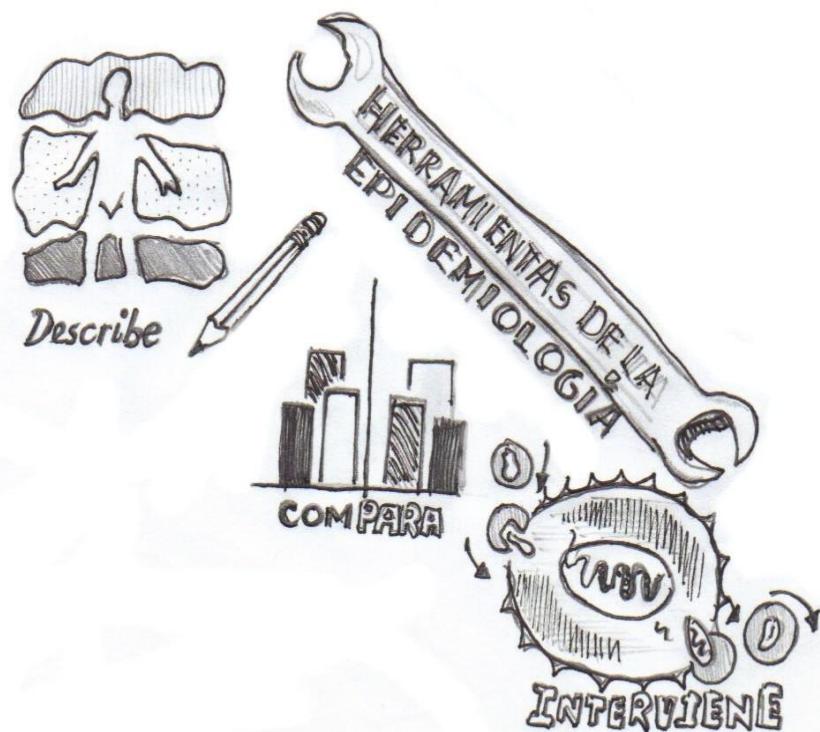


Existe un problema entre el hacer y el deber hacer. La paradoja de la humanidad es que, a pesar de saber que algo nos provoca daño, lo seguimos haciendo. ¿Por qué? Si bien la epidemiología identifica el daño, serán de seguro las ciencias humanas las que nos explicarán por qué lo seguimos haciendo. ¡Ambición y poder! muy seguramente son componentes a analizar.

Interesante, por ejemplo, es ver el papel de la epidemiología en torno a las fracturas de cadera. La epidemiología detectó la evidente relación entre fracturas de cadera y caídas en personas ancianas. Ha mostrado la relación entre la edad y ha propuesto de descalcificación como un factor importante. Ha analizado cómo se produce. Y desde ya, está planteado que se nos avecina como una gran epidemia, a la cual se le deberá dar un enfoque integral... ¿Por qué? Sencillamente porque el ser humano está viviendo cada vez más. En su enfrentamiento, se deberán buscar acciones preventivas (uso de estrógenos, por ejemplo) acciones de

promoción de estilo de vida para ancianos (viviendas dispuestas de cierta forma) y cálculos de costos de servicios de salud para enfrentar el problema.

De todo esto lo que resulta particularmente importante es que todos aquellos que trabajan en salud han de tener en cuenta el valor del pensamiento epidemiológico (y que también son sus herramientas), el cual puede sintetizarse en tres puntos:



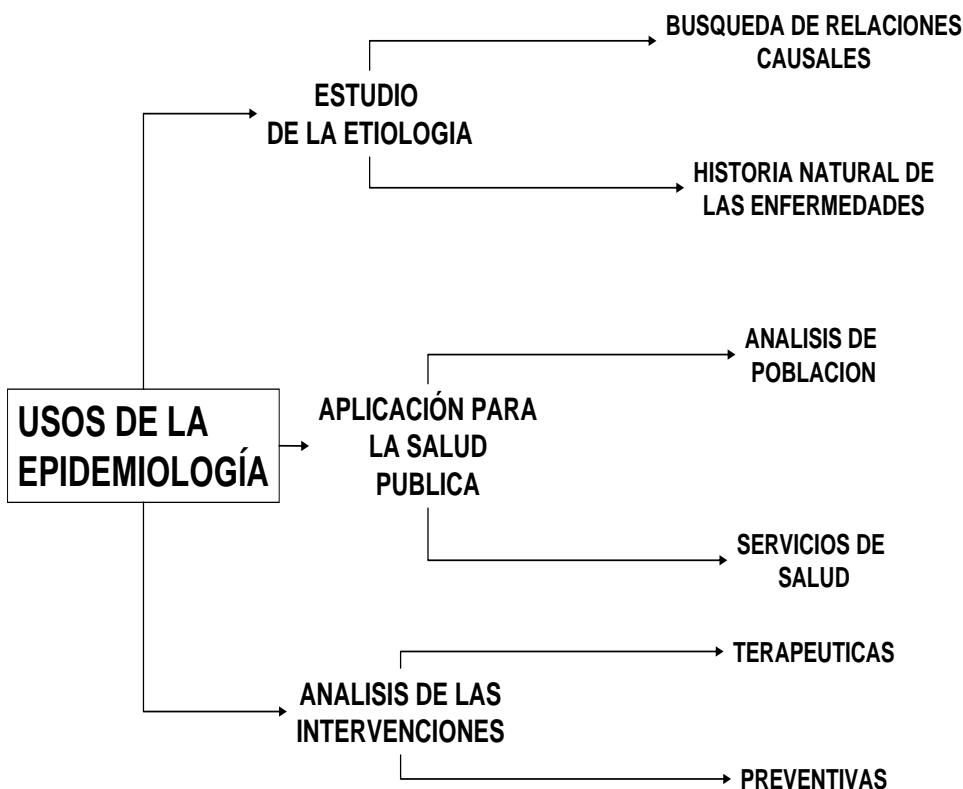
1. Describir lo que ve
2. Compararlo y analizarlo
3. Tomar acciones (intervenir) en pro del beneficio del ser humano y del mundo que lo rodea.

Todo lo anterior a través de la búsqueda de evidencia.

Usos de la epidemiología

Casi a manera de conclusión se puede afirmar que la epidemiología sirve (se usa) para:

- Estudiar la etiología de la enfermedad
- Ser aplicada para ofrecer salud a la población
- Valorar las intervenciones.





Capítulo 3

Las ayudas en la Epidemiología: La Investigación



<u>3.1. El raciocinio científico. La base.....</u>	107
<u>3.2. Los paradigmas de la investigación.....</u>	111
3.2.1. Concepto de sesgos	114
<u>3.3. Las fases de la investigación.....</u>	118
<u>FASE UNO: PLANTEAMIENTO DEL PROBLEMA</u>	121
<u>1.1. Paso 1: Selección del tema.....</u>	121
<u>1.2. Paso 2: Revisión bibliográfica.....</u>	122
1.2.1. Cómo se revisa un artículo	124
<u>1.3. Paso 3: Especificación del problema o de la pregunta de investigación</u>	128
<u>1.4. Paso 4: Identificar las probables respuestas (las hipótesis).....</u>	130
1.4.1. Algo más sobre las hipótesis de investigación	134
1.4.2. Las hipótesis estadísticas.....	137
<u>1.5. Paso 5: Pertinencia de contestar la pregunta de investigación (o importancia de la respuesta de la pregunta).....</u>	139
<u>1.6. Paso 6. Construcción de un marco teórico</u>	141
1.6.1. Definición de los componentes del problema:	142

<u>1.6.2. Qué respuestas se conocen a la pregunta</u>	143
<u>1.6.3. Cómo la han contestado en otros estudios (diseños empleados, variables y manejo estadístico)</u>	144
<u>1.6.4. Recopilación del marco teórico (qué debe decir una tesis)</u>	148
FASE DOS: DISEÑO Y PLANEACIÓN (PLAN PARA CONTESTAR LA PREGUNTA).....	154
<u>2. Paso 7. Selección del diseño que se empleará</u>	154
<u>2.1. Diseños cuantitativos</u>	156
<u>2.1.1. Intervencionales o experimentales</u>	158
<u>2.1.2. Observacionales</u>	162
<u>2.2. Diseños cualitativos</u>	170
<u>2.3. Investigación multi-método</u>	171
<u>2.4. Paso 8. Selección de la población</u>	172
<u>2.4.1. Acceso a la población</u>	173
<u>2.4.1.1. Criterios de inclusión</u>	173
<u>2.4.2. Cumplimiento de principios éticos</u>	176
<u>2.5. Muestras (de población)</u>	178
<u>2.5.1. Muestras representativas</u>	178
<u>2.5.2. Muestras por conveniencia:</u>	180
<u>2.6. Paso 9. Selección de las variables (instrumentos a emplear)</u>	182
<u>2.7. Sesgos</u>	187
<u>2.7.1. Atributos de una medida</u>	189
<u>2.8. Paso 10. Manejo estadístico de las variables</u>	190
<u>2.8.1. Base de datos</u>	190
<u>2.8.2. Paso 11. Construcción del formulario</u>	192
<u>2.9. Paso 12. Plan para el trabajo de campo</u>	194
<u>2.9.1. Descripción de las actividades</u>	195
<u>Consecución de la población</u>	195
<u>Trabajo de campo</u>	197
<u>Análisis de la información</u>	199
<u>Presentación de la investigación</u>	200
<u>2.9.2. Recursos que se necesitan para cumplir con las actividades</u>	202
<u>2.10. Paso 12. El protocolo de investigación</u>	204
FASE TRES: TRABAJO DE CAMPO	208
<u>3. Paso 14. Recolección de la información</u>	208
<u>3.1. Cómo se accedió a la población</u>	208
<u>3.2. Cómo se recogió la información</u>	208
FASE CUATRO: ANÁLISIS DE DATOS	209
<u>4. Paso 15. Registro de información recogida en la base de datos</u>	209
<u>4.1. Paso 16. Procesamiento la información</u>	210
FASE CINCO: DIVULGACIÓN DE LA INVESTIGACIÓN	212
<u>5. Paso 17. Publicación de la investigación</u>	212
<u>5.1. Cómo escribir un artículo</u>	212
<u>5.2. Paso 18. Presentación de la investigación</u>	215



CAPÍTULO 3. LAS AYUDAS EN LA EPIDEMIOLOGÍA: LA INVESTIGACIÓN

La principal ayuda que usa la epidemiología es la investigación. De hecho, es difícil separarlas. **Investigar** es el término que define la construcción del conocimiento. Otra manera de entenderla es como la forma a través de la cual, el ser humano resuelve sus problemas. En el área de la salud la investigación se convierte es una de las principales estrategias para combatir la enfermedad, para consolar al ser humano, para tomar decisiones.



Investigar, es de manera simple, resolver “problemas”. Las dudas que surgen de los procesos de salud y enfermedad en la población son la razón de existir de la epidemiología y la forma como las resuelve es la investigación.

3.1. El raciocinio científico. La base...

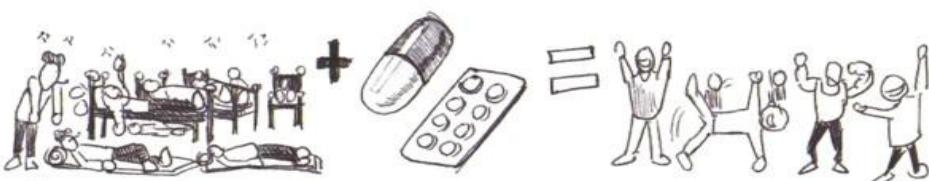
Razonar es pensar. Pensar es enfrentarse a problemas y resolverlos. Un problema, es en sí, una pregunta que amerita ser contestada. Resolver preguntas es investigar. La epidemiología busca resolver problemas de salud y enfermedad en la población, por ende, busca investigar en estos aspectos (salud, enfermedad y población). La base de la investigación epidemiológica es el raciocinio científico. Con ello se quiere decir que se usa la coherencia (relación entre las partes) y el raciocinio lógico, donde la experiencia, las facultades intelectuales y la forma de pensar se integran. Parte de un pensamiento **inductivo** (la mayoría de las veces), donde se busca establecer generalizaciones de lo particular o **deductivo** donde de lo general se busca aplicar a lo particular. Ambas maneras son útiles, pero así mismo, tienen limitaciones establecidas por la coherencia, por la verificación.

El raciocinio científico deriva su nombre del uso del método científico. Como tal, el método exige que se cumplan siempre unos pasos destinados a plantear la solución de un problema. Lo primero es que se debe plantear de manera adecuada el problema. Luego describir al máximo los componentes del problema. De ahí deben surgir las probables respuestas al problema (hipotéticas). Se escogerá de las alternativas una (la más coherente) y se aplicará. Finalmente se debe verificar si con la respuesta escogida ciertamente se resolvió o no el problema. Si es lo primero, entonces se ha conseguido un referente para enfrentar el problema una próxima vez. Si por el contrario no se resolvió, entonces ya se sabe que esa alternativa no es la respuesta y se debe escoger otra de las planteadas.

1. DADO QUE ...

SI DE TODOS

"MEJORÓ"



2 SE DEDUCE QUE ...

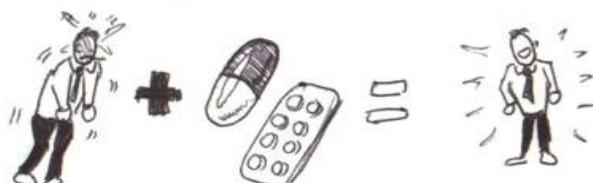


PUEDE SUCEDER ...



1. DADO QUE ...

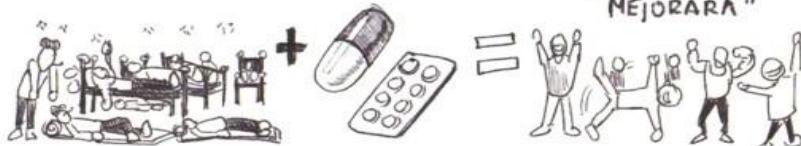
"MEJORÓ"



3. SE INDUCE QUE ...

PUEDE SUCEDER ...

"MEJORARÁ"

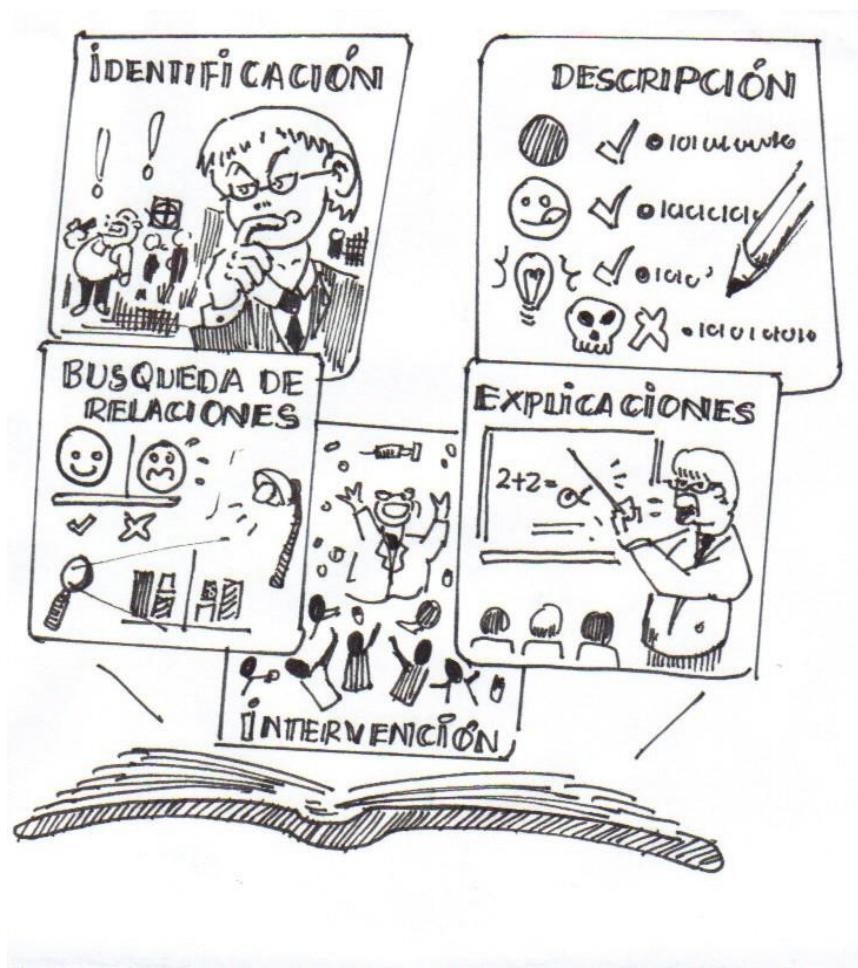


Deducir: Extraer del principio general consecuencias particulares. En investigación es trasladar a un individuo lo que se ha encontrado en la población.

Inducir: Extraer, a partir de determinadas observaciones o experiencias particulares, el principio general que en ellas está implícito. En investigación es trasladar a una población lo que se encontró en una muestra de individuos.

El raciocinio científico exige combatir con particular cuidado los sesgos o errores en sus diseños (de sus investigaciones) buscando con ello aproximarse al máximo a la realidad. Al fin de cuentas el anhelo de las investigaciones es el poder predecir o, lo que es lo mismo, proyectar a una gran población los resultados obtenidos de una muestra. Por ello, el particular interés en combatir los sesgos o errores en las investigaciones. Ellos son la distancia que existe entre la realidad y los resultados y si una pregunta es contestada siguiendo un diseño que se distancia de la realidad, entonces el resultado no podrá aplicarse al resto de la población (se malogra la predicción).

En sí, las grandes áreas de la investigación en salud y, por tanto, de la epidemiología abarcan:



- I. La **identificación** de fenómenos o sucesos parte de la importancia de reconocer lo que nos provoca daño (la enfermedad), para poderlo combatir (nadie combate lo que no conoce). Para esta parte saber lo que es normal y lo que no, es indispensable.
- II. La **descripción** de algún suceso. Acá se buscan criterios lógicos y entendibles que permitan conocer (o identificar), por ejemplo, a la enfermedad.
- III. La **búsqueda de relaciones** de algún fenómeno o suceso. Bajo este parámetro se buscan asociaciones que permitan explicar la aparición del suceso (enfermedad). Como tal es una búsqueda hacia la causa y hacia intervenciones que promuevan que no aparezca (enfermedad) o que aparezca (factores protectores) situaciones pro salud. Claro... esto se hace comparando.
- IV. La búsqueda de **explicaciones** de lo estudiado. Pretende incrementar el conocimiento del fenómeno a través de relaciones, persiguiendo como objetivo máximo la búsqueda de las asociaciones a través de las cuales se pueda controlar el fenómeno.
- V. **Las Intervenciones** o acciones donde se somete a una población a una variable puesta por el investigador (llámese medicamento o vacuna, por ejemplo) la cual ha de provocar un beneficio en la población que la recibe (se cura, se mejora o evita que aparezca la enfermedad).

La visión de las investigaciones epidemiológicas es la predicción. Predecir es poder **controlar** (que dependa de uno) el que aparezca o no un fenómeno. En el caso de la epidemiología la máxima predicción radica en evitar el desenlace de una enfermedad. Por ejemplo, a través de un tratamiento o lo que sería mejor, evitar la aparición de la enfermedad a través de una serie de acciones.

 La historia del SIDA es un buen ejemplo: ante la muerte de una persona joven por una enfermedad rara, un doctor se inquietó y empezó a buscar casos similares en otras partes; al hacerlo, **identificó** una nueva enfermedad. Con base a tal identificación, empezaron a **describirla** siguiendo criterios descriptores (historia natural, por ejemplo) y se dieron cuenta que era una entidad verdaderamente grave (mortal). Entonces empezaron a **buscar relaciones** (comparar). En primera instancia, los homosexuales se convirtieron en la población objeto de la mayoría de los estudios, sencillamente porque eran las personas a las que más les daba. A través de estudios de casos y controles (buscando relaciones) detectaron que se comportaba igual que una ETS. Ello les permitió dar explicaciones hipotéticas, con lo que todos los estudios se dirigieron hacia las infecciones, lo que determinó el descubrimiento del virus. Identificadas las características del virus el siguiente paso fue buscar cómo combatirlo, desarrollando los antirretrovirales. Ya con ellos, lo siguiente es hacer pruebas de medicamentos que buscan evitar la muerte de los enfermos. Al dárselo a un enfermo **predicimos** que no morirá (mortalidad) e incluso, que mejorará (morbilidad)

3.2. Los paradigmas de la investigación...

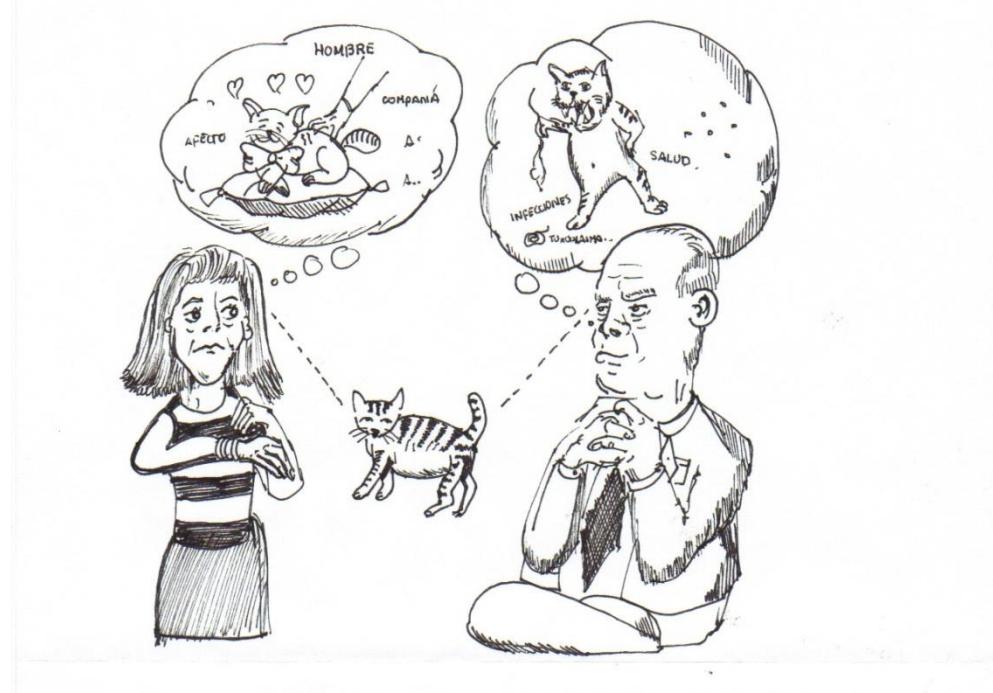


Un **paradigma** es una forma de cómo los seres humanos ven el mundo (cómo se les enseña que vean). La investigación se ha desarrollado bajo dos paradigmas: el **positivista** impuesto y basado bajo los pasos del **método científico y el naturalista o fenomenológico o constructivista**. De hecho, uno y otro paradigma han desarrollado una serie de diseños (forma como se investiga) para el desarrollo de la investigación. El positivismo, los diseños **cuantitativos y el constructivista**, los métodos **cualitativos**. En el positivismo, el investigador busca por sobre todo entender y explicar la realidad partiendo que ella es independiente del observador (es objetiva). Cree profundamente en el **determinismo** que plantea que todo sucede por razones y no de manera fortuita y la búsqueda de esas razones es su objeto, ya que el comprenderlas se podrá predecir. Además, procuran aislar al observador de lo observado, buscando que influya lo menos posible sobre lo estudiado.

El **naturalista (o complejo)** parte de asumir que no existe el determinismo. Que la realidad es cambiante y que depende de los individuos, por ende, todo es relativo. Por ello, y de acuerdo a cada modelo, la investigación se efectúa partiendo de unos diseños de acuerdo a la forma de ver el mundo.

Y ante esto... ¿cuál es el modelo que tiene la verdad? o ¿cuál asemeja mejor la realidad? Eso es algo que cada uno contestará, pero para aquellos que queremos ayudar al ser humano, el entender se convierte en algo por encima del juzgar, porque ese es el primer paso para poder ayudar.

Lo cierto y, verdaderamente útil, es que combinando estas dos formas de ver al mundo a través de sus métodos (diseños), la investigación se enriquece y, sin duda, se acerca más a su importante misión como es ayudar al ser humano.



Un paradigma es un modelo, y un modelo es una forma de hacer o pensar. Bajo el contexto, un paradigma es una forma de pensar usada por una parte de la humanidad. Cada momento de la historia de la humanidad ha tenido un paradigma dominante. En la salud y en la enfermedad la forma de pensar estuvo poderosamente influenciada en la antigüedad por Hipócrates y Galeno. Las analogías en buen parte eran la esencia de la forma de pensar. En la época moderna se introduce el pensamiento anatomoclínico (donde, y a través de las autopsias, se empieza a construir el raciocinio médico), etiopatogénico y fisiológico (donde la búsqueda de las causas y el entender lo que se daña son el origen de las explicaciones médicas). Hoy se está ante el positivismo y el constructivismo. Por supuesto, que cada pensamiento toma de los previos elementos que contribuyen en sus planteamientos. Obviamente los cambios de los paradigmas ocasionan crisis (enfrentamientos entre lo nuevo y lo previo). En los momentos de crisis se enfrentan aquellos tradicionales con los innovadores. La historia muestra cómo la primera generación de innovadores habitualmente es sacrificada. El cambio... no es fácil de aceptar. Tan solo las generaciones siguientes de innovadores ven el cambio con sus consecuencias.

En torno al paradigma positivista, el uso del **método científico** es prioritario. Se parte de un problema bien planteado, donde el caso particular no es lo importante sino más bien lo general (de hecho busca obtener premisas generalizables). Pretende ser **sistemático**, queriendo decir con ello que intenta abarca un todo y para ello cumple con una serie de pasos (**metódico**). Busca **controlar variables** ajena a la pregunta de investigación imponiendo condiciones que disminuyen los sesgos incrementando así la **validez y la precisión**. Usa como punto de partida a la realidad, capturada bajo distintas formas (pruebas empíricas). Se exige, en lo posible, utilizar medidas de referencia objetivas (que puedan ser repetidas). Este modelo, también denominado **biomédico**, ha provocado inmensos avances en torno a la salud del ser humano. Se enfrenta como debilidad a cuestiones éticas y morales importantes que alteran el diseño propuesto y, claro, la debilidad derivada de tener que usar medidas no objetivas en algunas de sus preguntas (por ejemplo, las percepciones).



Cuando los experimentos demostraron el potencial beneficio de los antirretrovirales en los pacientes con SIDA, se pasó a la fase de prueba en poblaciones. Ello implicó suministrárselos a unos pacientes los antirretrovirales y no suministrárselos a otros, ello, sabiendo que lo esperado era que aquellos que NO los recibieran morirían. Ciertamente acá existe un dilema ético... ¿verdad?

Con la investigación naturalista (**paradigma constructivista**) se pretende abordar la realidad con su complejidad. Parte de la premisa que la realidad no está determinada sino que se construye con cada caso, con cada vivencia. Procura con sus diseños ser **integradora** y no **reduccionista** con lo que increpa a los positivistas. Por lo cual, usan diseños flexibles y cambiantes, buscando más explicaciones generales que ayuden a entender que modelos de predicción. Como debilidad se menciona el carácter dominantemente subjetivo de sus diseños.



El positivismo expresa que se debe investigar usando variables objetivas pues de no hacerlo, la validez es cuestionada. Por ejemplo, cuestiona el uso de algunas variables tales como las percepciones para mediar efectos de un tratamiento. La percepción de sentirse bien cuando se usa una medicación en el control de la diabetes, no es una buena medida pues varía: hoy se siente de una forma y mañana de otra y son tantos los factores que intervienen que es difícil atribuírselo al medicamento. Por ende, se busca una variable objetiva y directamente implicada, tal como la cifra de glicemia (por ejemplo). Claro... acá el debate es duro pues el pensamiento complejo que desarrolla el constructivismo dice que la realidad es más parecida a las percepciones que a esa variable (en el ejemplo), es decir, que siempre existen muchísimas relaciones que cambian o influyen en lo que uno mide, a tal punto que no vale la pena ser tan rigurosos como en el positivismo al momento de hacer los diseños de investigaciones.

3.2.1. Concepto de sesgos

Independiente de los paradigmas, ambos modelos buscan entender y aprender, usan o parten de lo que existe en el mundo y pretenden beneficiar al ser humano. Además, un diseño u otro tienen algún grado de error. **El error o sesgo** es la distancia que existe entre la investigación y sus resultados con la realidad.



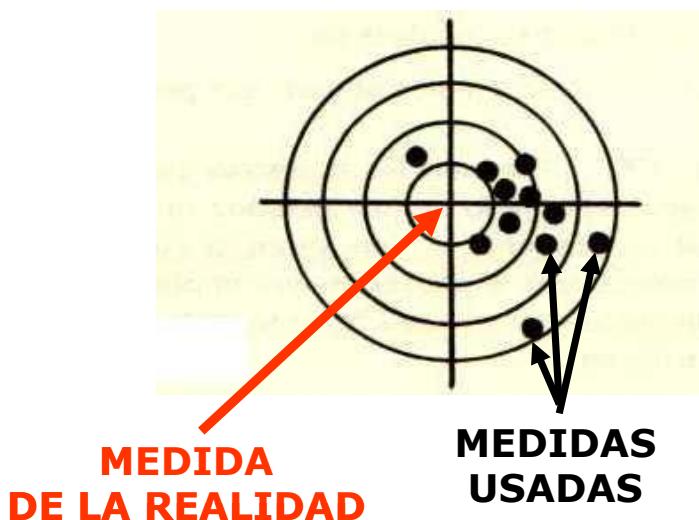
En una investigación existen dos grandes tipos de errores: los **SISTEMÁTICOS** o propios del investigador y los **ALEATORIOS** o propios de la investigación. Los sistemáticos ocurren:

1. O bien por errores en el proceso de recolección de información
2. O por una mala selección de la población objeto de la investigación
3. O por un error en la interpretación de un resultado.

Por ejemplo, si se hace una encuesta y la persona responde sin decir la verdad (**SESGO DE INFORMACIÓN**) o se va a probar un medicamento en una enfermedad mortal y se usan para comparar una población en mejor condición (de salud) que la otra (**SESGO DE SELECCIÓN**) o si ante un resultado se da una interpretación equivocada (**SESGO DE CONFUSIÓN**).

Por ejemplo, en las encuestas de sexualidad practicadas sobre población adolescente se deben proponer diversos mecanismos para evitar que respondan mentiras (sesgo de información). Que los profesores no estén presentes o que las encuestas sean anónimas, son algunos. Un caso de sesgo de selección es si se va a probar un medicamento y se les da a pacientes menos enfermos en tanto que a los más enfermos se les da el placebo. Resultará muy posible que existan más mejorías en los que reciben el medicamento, ¿cierto?... pero, ese resultado está sesgado, no pudiéndose afirmar si la mejoría es porque estaban menos enfermos o atribuírselo al medicamento. Y finalmente, y como muestra de un sesgo de confusión (que es un resultado), se menciona la fuerte relación entre cáncer del pulmón y tener los dedos amarillos. La relación es de confusión pues es por fumar que los dedos se ponen amarillos (impregnación nicotínica) y, por consiguiente, la relación es entre el cigarrillo y el cáncer y no entre los dedos amarillos y el cáncer.

La epidemiología está por completo alineada al paradigma positivista y por ende, al uso del raciocinio científico y como consecuencia derivada de los diseños (forma de contestar las preguntas de investigación) cuantitativos.



Si se supone que la realidad es dar en el centro, la validez de algo está dada por la cercanía que se esté del centro. Entre más distante, más error o sesgo. Y si lo que mido siempre me da en un mismo punto lejos del centro entonces estoy ante un error sistemático.

Claro... una buena medida es aquella que al ser usada expresa la realidad (**validez**). Una medida **confiable** es aquella que al repetirse da en el mismo sitio. Una medida **exacta** es la que es válida y confiable y, por tanto, es el sueño de las medidas.

Si lo que quiero medir es el tamaño de una persona puedo usar como medida cómo le parece a otros: más alto, igual o menos. Esta medida depende de quien la responda (**no es confiable**). En cambio si usa la altura en centímetros el dato es confiable, ya que siempre que lo vuelva a medir, me dará el mismo dato en centímetros y real (**válido**) ya que eso es lo que mide, por consiguiente, se está usando una medida **exacta**. Obviamente saber cuál es mejor para una investigación se cae de su peso... ¿cierto?

Las medidas pueden llamarse en algunos casos **ESCALAS** que son un conjunto de variables unidas y que se interpretan de acuerdo a los datos obtenidos de cada variable y sumadas todas. Tienen un uso particular en las detecciones de riesgo. Por ejemplo, la escala de Framingham usa diversas variables (edad, niveles de lípidos, glicemia, etc.) para que con base en los datos obtenidos (respuesta de cada variable) se pueda identificar el riesgo de que suceda un infarto en determinado tiempo. Una escala perceptiva es el test de APGAR familiar (**Apoyo**,

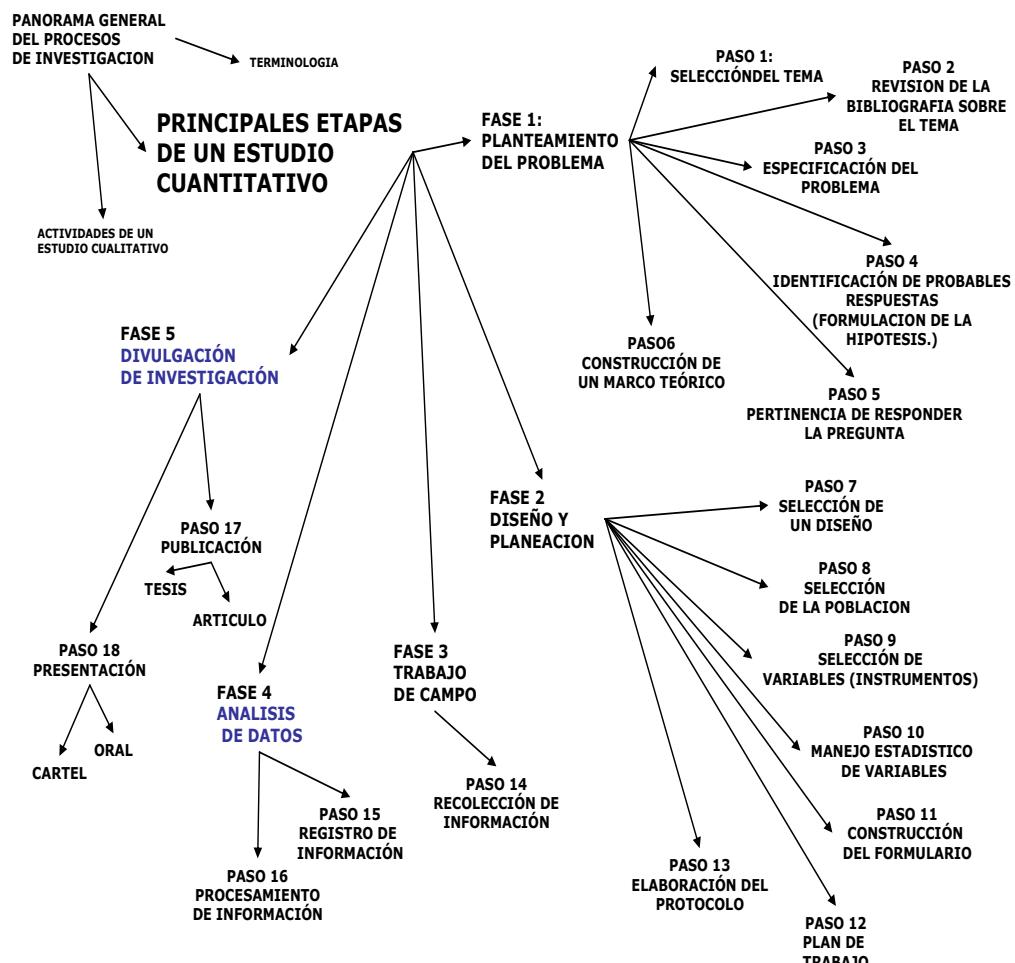
Participación, Ganancias, Afecto, Recursos) en donde a través de cinco preguntas que pueden ser contestadas con unos datos, se mide cómo el que contesta percibe a su familia en torno a la funcionalidad.



Lo realmente interesante es que a través de unir variables y construir escalas en el mundo de la salud se está haciendo un importante esfuerzo por predecir. Por ello la **VALIDACIÓN** de escalas se convierte en una importante área de investigación. Consiste en poder identificar que la escala en cuestión verdaderamente mide lo que se quiere. En los dos ejemplos previos, la escala de Framingham ciertamente mide el riesgo de IAM en los siguientes cinco años, como quien dice que aquel que salió con riesgo positivo, ciertamente tiene una alta probabilidad de que le de un infarto si no modifica su estilo de vida o que el test de APGAR familiar sí mide la funcionalidad percibida del individuo de su familia.

3.3. Las fases de la investigación...

La investigación, de acuerdo a los paradigmas mencionados previamente, puede ser cualitativa y cuantitativa. La cuantitativa sigue unas fases (o etapas) por demás obvias y que se han de tener en cuenta cuando se selecciona alguno de los distintos diseños. A continuación se van a exponer las fases con sus consiguientes pasos. Estos son:



Existen muchas formas de presentar el orden en que se debe hacer una investigación. La gráfica expone uno. El orden que se presente (este u otro) debe cumplir con el requisito de ser coherente, sistemático y racional. Al fin de cuentas de lo que se trata es de aplicar el método científico como modelo para resolver una pregunta.

FASE UNO: PLANTEAMIENTO DEL PROBLEMA

- 1.1. Paso 1: Selección del tema
- 1.2. Paso 2: Revisión bibliográfica
- 1.2.1. Cómo se revisa un artículo
- 1.3. Paso 3: Especificación del problema o pregunta de investigación (qué, cómo, en quién)
- 1.4. Paso 4: Identificar las probables respuestas
 - 1.4.1. Algo más sobre las hipótesis de investigación
 - 1.4.2. Las hipótesis estadísticas
- 1.5. Paso 5: Pertinencia de contestar la pregunta (o importancia de la respuesta a la pregunta de investigación)
- 1.6. Paso 6. Construcción de un marco teórico
 - 1.6.1. Definición de los componentes del problema
 - 1.6.2. Qué repuestas se conocen de la pregunta
 - 1.6.3. Cómo la han contestado en otros estudios (diseños empleados, variables y manejo estadístico)
 - 1.6.4. Recopilación del marco teórico (qué debe decir una tesis).

FASE DOS: DISEÑO Y PLANEACIÓN (PLAN PARA CONTESTAR LA PREGUNTA)

2. Paso 7: Selección del diseño que se empleará
 - 2.1. Diseños cuantitativos
 - 2.1.1. Intervencionales o experimentales
 - 2.1.1.1. Experimento puro
 - 2.1.1.2. Ensayo clínico
 - 2.1.1.3. Ensayo de campo
 - 2.1.1.4. Cuasiexperimental
 - 2.1.1.5. Intervención comunitaria
 - 2.1.2. Observacionales
 - 2.1.2.1. Descriptivos
 - 2.1.2.1.1. Serie de casos
 - 2.1.2.1.2. De corte (descriptivo exclusivamente)
 - 2.1.2.2. Analíticos
 - 2.1.2.2.1. De corte (comparativo)
 - 2.1.2.2.2. Casos y controles
 - 2.1.2.2.3. Cohorte
 - 2.1.2.2.4. Ecológico
 - 2.1.3. Otros diseños: Meta-análisis, Revisiones sistemáticas
 - 2.1.4. Investigaciones de administración
 - 2.2. Diseños cualitativos
 - 2.3. Diseños multimétodos
 - 2.4. Paso 8. Selección de la población (en quién se contestará la pregunta o se resolverá el problema)
 - 2.4.1. Acceso a la población
 - 2.4.1.1. Criterios de inclusión

- 2.4.2. Cumplimiento de principios éticos
 - 2.4.2.1. Principio de beneficencia
 - 2.4.2.2. Principio de respeto a la dignidad humana
 - 2.4.2.3. Principio de justicia
 - 2.4.2.4. Consentimiento informado
 - 2.4.2.5. Consideraciones sobre sujetos vulnerables
 - 2.4.2.6. Comités de ética
- 2.5. Muestras (de población)
 - 2.5.1. Muestras representativas
 - 2.5.2. Muestras por conveniencia
 - 2.5.3. Tamaño de muestra
- 2.6. Paso 9. Selección de variables (Instrumentos a emplear)
- 2.7. Sesgo (Errores)
 - 2.7.1. Atributos de una buena medida
- 2.8. Paso 10. Manejo estadístico de las variables
 - 2.8.1. Base de datos
 - 2.8.2. Paso 11. Construcción del formulario
- 2.9. Paso 12. Plan para el trabajo de campo
 - 2.9.1. Descripción de las actividades (Qué se va hacer, cómo y cuándo)
 - 2.9.2. Recursos que se necesitan para cumplir con las actividades
- 2.10. Paso 13. Elaboración del protocolo de investigación.

FASE TRES: TRABAJO DE CAMPO

- 3. Paso 14. Recolección de información con la población
 - 3.1. Cómo se accedió a la población
 - 3.2. Cómo se recogió la información.

FASE CUATRO: ANÁLISIS DE DATOS

- 4. Paso 15. Registro de información recogida a la base de datos
 - 4.1. Paso 16. Procesamiento de información
 - 4.2. Obtención de características de la población
 - 4.3. Obtención de las frecuencias
 - 4.4. Obtención de la respuesta a la pregunta de investigación.

FASE CINCO: DIVULGACIÓN DE LA INVESTIGACIÓN

- 5. Paso 17. Publicación de la investigación
 - 5.1. Cómo escribir un artículo
 - 5.2. Paso 18. Presentación de la investigación.



Clasificar es una manera agrupar que contribuye con el entendimiento. Sin embargo, esa facilidad no siempre es válida, y con ello se quiere decir que el ser humano a veces usa clasificaciones que no son reales o lo que es lo mismo, que no expresan la realidad tal cual: son apenas expresiones de cómo unos entienden algo. En torno a los pasos de una investigación, existen múltiples y variadas clasificaciones. Acá se expone una de tantas...

FASE UNO: PLANTEAMIENTO DEL PROBLEMA

El término problema en investigación se refiere a la inquietud que el investigador pretende resolver. A pesar de lo obvio que parece, el identificar claramente lo que se quiere indagar no es precisamente algo que el ser humano haga con frecuencia. El objetivo de esta primera fase es que partiendo de una inquietud el investigador llegue al planteamiento de una pregunta de investigación bien elaborada.

A continuación se ofrece al lector una serie de pasos que pueden serle útiles en el momento de plantear problemas de investigación.

1.1. *Paso 1: Selección del tema*



En primer lugar, el investigador ha de escoger un TEMA o área en la que DESEA investigar. Distintas razones orientan la selección del tema, siendo la principal la **motivación** del investigador. Es un requisito sin el cual es poco probable que la investigación concluya, ¿por qué?, pues sencillamente porque pocas veces el adagio que lo que se quiere exige esfuerzo resulta tan válido (válido = real). La investigación exige esfuerzo, así como tiempo, condiciones ambas que implican que la actitud del investigador ha de ser coincidente con ellas; por eso, el gusto del investigador ha de estar presente, para que más que un trabajo sea un placer. Dado que la investigación como pocas actividades humanas exige trabajar con otros, el que los investigadores compartan el interés por el tema es así mismo una deseable condición. Por ello se habla de la necesidad de establecer contacto con una comunidad científica, que no es otra cosa que investigadores con los que se comparte el tema o el problema, por lo cual no solo es fácil hablar con ellos sino que resulta estimulante. Otro factor que puede influir al escoger el tema o incluso el problema es la **EXPERIENCIA** previa del investigador. Ello facilita enormemente la labor del investigador, porque ya conoce (camina sobre terreno conocido). Además, no es infrecuente que los investigadores se dediquen a áreas de sus intereses (de ahí surgen los expertos). Claro... otro aspecto a tener presente es la **FACILIDAD** o **FACTIBILIDAD** del tema. Es evidente que existen temas o áreas que son más fáciles para unos u otros. Al momento de seleccionar el tema el tener esto en cuenta siempre importa. Por ejemplo, el acceso a BIBLIOGRAFIA acerca del tema, siempre es una condición que facilita enormemente la labor del investigador. La **MODA** también es un importante motivador al momento de seleccionar un tema. Existen temas que el mercado solicita en determinado momento y al abordarlos el investigador puede estar asegurando su subsistencia. Pero tal vez el criterio de factibilidad más importante es el **ACCESO A LA POBLACIÓN** objeto de la investigación (al menos en la investigación epidemiológica o poblacional). Saber que se puede contar con ella es, sin duda, uno de los criterios más importantes a tener en cuenta.

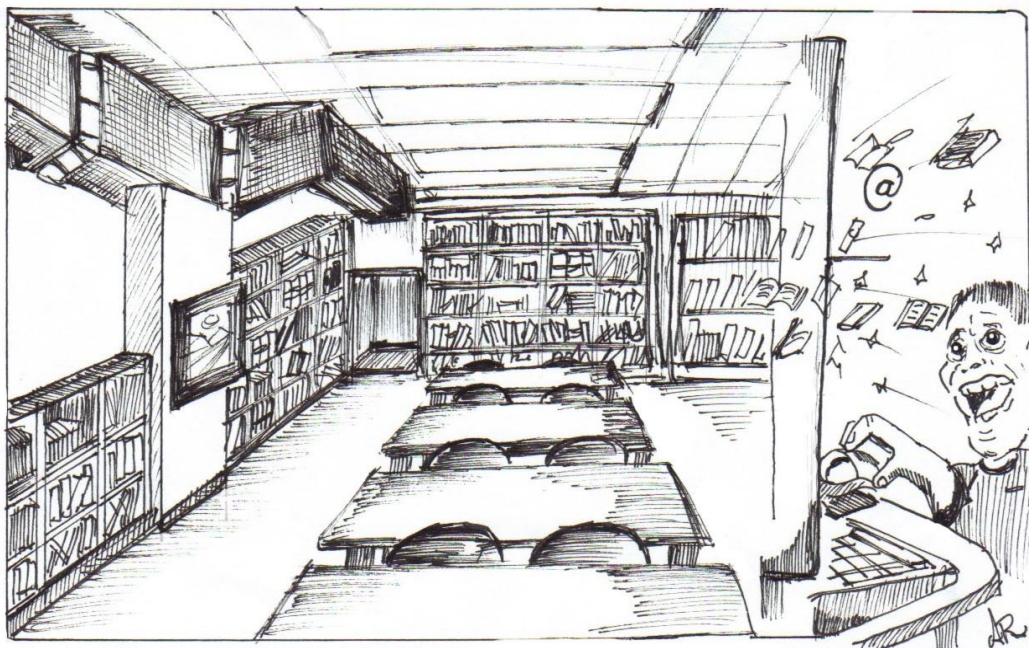
1.2. Paso 2: Revisión bibliográfica

Seleccionado el tema el conocer qué se ha publicado al respecto es clave (definiendo si es posible qué se quiere saber y en quién), esto es muy importante para precisar la pregunta de investigación. Para ello, saber buscar será una importante herramienta que el investigador deberá afinar. Al hacerlo el investigador podrá ver cómo otros accedieron a la población o qué diseños o instrumentos se emplearon en otras partes. Es más... a lo mejor se puede dar cuenta si la pregunta que le inquieta fue ya contestada y, así, cambiar de objetivos.

Saber buscar información resulta muy trascendente. Una gran parte de las investigaciones son publicadas en revistas interesadas en el tema. Esas revistas pueden ponerse al alcance de todos a través de su inclusión en bases de datos a las que se puede tener acceso. Las bibliotecas adquieren esas bases de datos y mediante ellas los consultantes. Para poder buscar los artículos, los organizadores de información promueven el uso de unas palabras denominadas términos MeSH pretendiendo que siempre sean estas las que se usan (en las bases de datos) y no sinónimas, facilitando así la clasificación de la información. Probablemente PUB MED sea la base a través de la cual se debe iniciar la búsqueda de información respecto al tema, por tratarse de la mayor base de datos del mundo biomédico.



Hoy existen cursos de búsqueda de información en muchas partes. Sin duda vale la pena hacer alguno y aprender a buscar pues ello determinará en su momento la solidez de la investigación que se obtiene al confrontar lo indagado con lo revisado.



El concepto de biblioteca como tal se ha modificado. De ser un espacio real hoy ha pasado a incluir los espacios virtuales en el mundo de las redes de informática.

La mayoría de la información publicada son investigaciones convertidas en artículos. Con el fin de facilitar su interpretación la mayoría de las revistas han venido definiendo unos formatos que faciliten su revisión.

1.2.1. Cómo se revisa un artículo

Las revistas, hoy por hoy, publican las investigaciones utilizando un formato, el que se debe conocer para así saber qué buscar y en dónde. Los autores se han comprometido a publicar de acuerdo a las siguientes pautas:

1. TITULO DE LA INVESTIGACION: Debe decir QUÉ SE HIZO, en QUIÉN se hizo, CÓMO se hizo y CUÁNDO se hizo. El cumplimiento de estos aspectos permite que los artículos se clasifiquen e ingresen a las bases de datos. Acá usan palabras reconocidas por los términos MeSh.

1.1. AUTORES: Ha de mencionar los apellidos y nombres de los autores, así como sus títulos académicos. Hoy se incorpora la declaración de si existen o no conflictos de interés.



Se entiende por **conflicto de interés** cuando los autores tienen algún compromiso con los resultados. Un ejemplo de conflicto de interés es cuando el investigador en un estudio sobre medicamentos, trabaja para el laboratorio que produce el medicamento. Si los resultados son adversos (por ejemplo) puede tener un conflicto derivado de su condición laboral para expresar el resultado.

1.2. RESUMEN. No mayor de 250 palabras. La idea esencial es que acá se encuentre la información básica para que otros investigadores identifiquen si les sirve o no el artículo original. Debe tener:

1. 2.1. Objetivo: se menciona el objetivo de la investigación.
- 1.2.2. Material y métodos (o Metodología): Donde expone tipo de diseño de la investigación.
- 1.2.2.1. Población objeto: Se menciona sobre quiénes se efectuó el estudio.
- 1.2.2.2. Variables utilizadas: Se menciona cuáles son las variables empleadas y cómo se interpretan.
- 1.2.2.3. Manejo estadístico: Se expone cuál fue el manejo estadístico de las variables.
- 1.2. 3. Resultados: Se menciona la respuesta a la pregunta de investigación.
- 1.2.3.1. Conclusión: Se expone el principal uso de los resultados.
- 1.2.4. Palabras claves: Se colocan las palabras claves de la investigación que aparezcan como términos MeSH.

RESUMEN

Objetivo Estimar la relación entre conocimientos, actitudes y prácticas (comportamientos) de riesgo en la sexualidad juvenil y la estructura y percepción de funcionalidad familiar.

Metodología Se realizó un estudio observacional de corte trasversal entre 13 699 adolescentes (10 a 19 años de edad) de 12 colegios públicos de la localidad de Suba. Se aplicó una encuesta autodiligenciada donde se interrogó por conocimientos, actitudes y comportamientos sexuales y la estructura y percepción del funcionamiento familiar.

Resultados Edad media de los adolescentes 13,5 años. Prevalencia de relaciones sexuales de 17 %. El 8 % de los hogares presentan disfunción severa y 39 % son nucleares incompletos. Las mujeres tienen menos relaciones sexuales que los hombres (OR 0.53; IC 95 %; 0.47-0.59). Se presenta más relaciones sexuales cuando los adolescentes perciben de manera disfuncional a su familia (OR 2.06; IC 95 %; 1.72-2.48).

Conclusiones Es más importante la percepción de función familiar que la estructura como factor de riesgo en la sexualidad de los jóvenes.

Palabras Clave: Adolescentes, sexualidad, conocimientos, actitudes y prácticas, relaciones familiares, colegios.

 Los resúmenes resultan muy importantes cuando se hace la revisión bibliográfica pues es lo que el investigador lee (porque se publica en todas las bases de datos) para saber si el artículo es o no de su interés. Si lo encuentra de su interés será su misión leer el original siendo para ello necesario que sepa cómo buscarlo. En el ejemplo, un resumen acerca de sexualidad, adolescencia, relaciones familiares...

2. INTRODUCCION: El lector debe identificar:

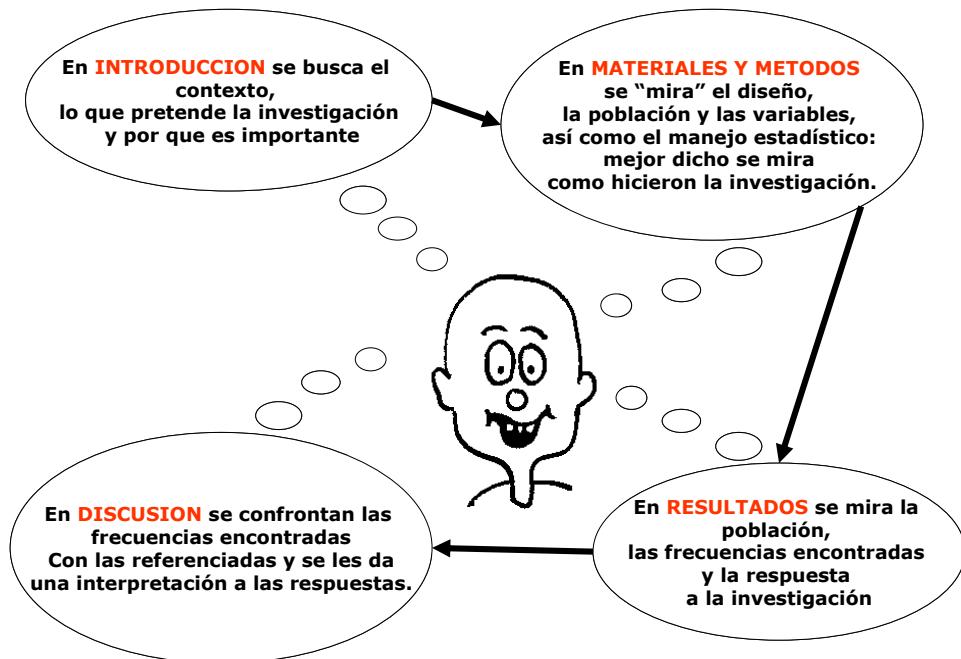
- 2.1. Contexto del tema
- 2.2. La pregunta de la investigación y/o los objetivos
- 2.3. La importancia de contestar la pregunta (pertinencia).

3. MATERIAL Y METODOS. Se identificará:
 - 3.1. Diseño de la investigación: Mencionar tipo de estudio
 - 3.2. Población objeto: Expone en quiénes se efectuó la investigación Cómo se escogieron y cómo se cubrieron los aspectos éticos
 - 3.3. Cuál fue el instrumento empleado, explicando cómo se manejaron las variables
 - 3.4. Cuál fue el manejo estadístico efectuado
 - 3.5. Cómo se realizó el trabajo de campo.

4. RESULTADOS. En general siguen este orden:
 - 4.1. Descripción de las características de la población estudiada: Total, sexo, edades y demás variables indagadas (para saber en quiénes se efectuó la investigación)
 - 4.2. Descripción de las frecuencias encontradas: Mencionar las prevalencias encontradas (de las variables estudiadas)
 - 4.3. Resultado de la pregunta efectuada: Identificación de la respuesta a la pregunta a través de las variables empleadas para ello
 - 4.4. Aplicación de las pruebas de estadística a los resultados (busca ver la inferencia en torno a población y a las respuestas).

5. DISCUSION. Acá es donde está la esencia de la investigación. El lector debe identificar:
 - 5.1. Identificación de SESGOS: Busca encontrar errores que no permitan estar seguros de los resultados y ver la capacidad de inferir en torno a lo encontrado. Esto se conoce también como LIMITACIONES de la investigación leída
 - 5.2. Comparación de las frecuencias encontradas con las referenciadas con las explicaciones que lo expliquen
 - 5.3. Respuesta a la pregunta de investigación con las explicaciones ofrecidas.

Es probable que muchos artículos no cumplan con todo esto. Ahí el lector decide la calidad y el valor que le da a los artículos leídos. Lo cierto es que conocer este orden es importante para el objetivo final que es entender lo que se lee.



En la década del 70, los editores de las revistas biomédica más prestigiosos se reunieron en Vancouver y propusieron una forma de presentar las investigaciones para su publicación. El **formato IMRD** o Introducción, Materiales y métodos, Resultados y Discusión fue el seleccionado. Esto se conoce como el **Protocolo de Vancouver**. Surge de la reunión de editores de revistas biomédicas y ofrece las recomendaciones que han de cumplirse en una publicación. Las revistas que accedan a aplicar los Requisitos Uniformes (unas 500 aproximadamente) se les solicita que citen el documento de 1997 en sus normas para los autores. Es importante destacar lo que estos requisitos implican.

- En primer lugar, los requisitos uniformes son instrucciones a los autores sobre cómo preparar sus manuscritos así se tenga que tener en cuenta algunas recomendaciones específicas de algunas revistas.
- En segundo lugar, si los autores preparan sus manuscritos según el estilo especificado en estos requisitos, los directores de las revistas acogidas al mismo no devolverán los manuscritos para que se realicen cambios de estilo. Sin embargo, en el proceso editorial las revistas pueden modificar los manuscritos aceptados para adecuarlos a su estilo de publicación.
- En tercer lugar, los autores que remitan sus manuscritos a una revista que participe de esta normativa, deben seguir los Requisitos Uniformes.
- Los autores seguirán también las instrucciones de cada revista con respecto a qué temas son pertinentes y el tipo de artículos que admite: por ejemplo, originales, revisiones o notas clínicas.

Además, es probable, que en dichas instrucciones figuren otros requisitos específicos de la publicación que deban seguirse, tales como el número de copias del manuscrito, los idiomas aceptados, la extensión del artículo y las abreviaturas admitidas. Se espera que las revistas que hayan adoptado estos requisitos indiquen en sus instrucciones para los autores, que sus normas siguen los requisitos uniformes para manuscritos enviados a revistas biomédicas y se cite una versión publicada de los mismos.

Una vez revisado el artículo, el investigador debe aprender a referenciarlo. Prácticamente todas las revistas mencionan cómo se debe hacer y hoy, los autores científicos se han puesto de acuerdo que ha de ser así (protocolo de Vancouver):

Primer Apellido y luego las iniciales del nombre del autor. Título de la investigación. Revista (con iniciales de identificación internacional, si las tiene). Año, volumen, y páginas. Observe los ejemplos:

- Haquin C, Larraguibel M, Cabezas J. Factores protectores y de riesgo en salud mental en niños y adolescentes de la ciudad de Calama. Rev. chil. pediatr 2004; 75 (5): 425-433.
- Bonomo Y, Proimos J. Substance misuse: alcohol, tobacco, inhalants, and other drugs. BMJ 2005; 330-333.

Como se observa, la referencia permite que el lector pueda, si lo desea, consultar el artículo (por ende, la información que ofrece se lo debe permitir). Si se quiere profundizar acerca de cómo se cita, se recomienda ubicar en la red las normas de Vancouver. Lo cierto es que de estas referencias provendrá la base de comparación para los resultados y así se construye la discusión (comparando lo hallado con lo revisado).

1.3. *Paso 3: Especificación del problema o de la pregunta de investigación*

Escogido el tema o área de interés, el plantear el problema en forma lógica y entendible es muy importante. Se afirma que un problema bien planteado está parcialmente resuelto, queriendo con ello insistir en la importancia de entenderlo. Como método para plantear problemas, se propone que cumpla con los siguientes aspectos:



1. Redactarlo a manera de pregunta. Así se puede intuir la respuesta que se busca y de ahí se desprende la utilidad de dicha respuesta.
2. Asegurarse de que gramaticalmente esté bien escrito. Ello ayuda a entenderlo a cabalidad. ¿Cómo se resuelve lo que no se entiende? Así mismo, el que sea breve ayuda (entre más palabras, más probable es la confusión).
3. Verificar que los componentes de la pregunta coincidan con los aspectos más mencionados en la observación que tiene el investigador. De hecho, usar frases donde se tomen varios tópicos del tema, puede ser de gran utilidad para precisar cuál o cuáles son los componentes que más inquietan al investigador (las que más se repitan son las que más se indagan).



Ejemplos de preguntas son:

¿Cuál o cuáles son las características...? Utilizada para describir.
¿Qué relaciones existen entre...? Utilizada para buscar asociaciones.
¿Qué efecto... o utilidad tiene...? Utilizada para medir intervenciones.

La pregunta debe llenar como requisitos: **Qué** se quiere indagar, **cómo** se va a indagar y **en quién** se va a indagar.

1.4. Paso 4: Identificar las probables respuestas (las hipótesis)

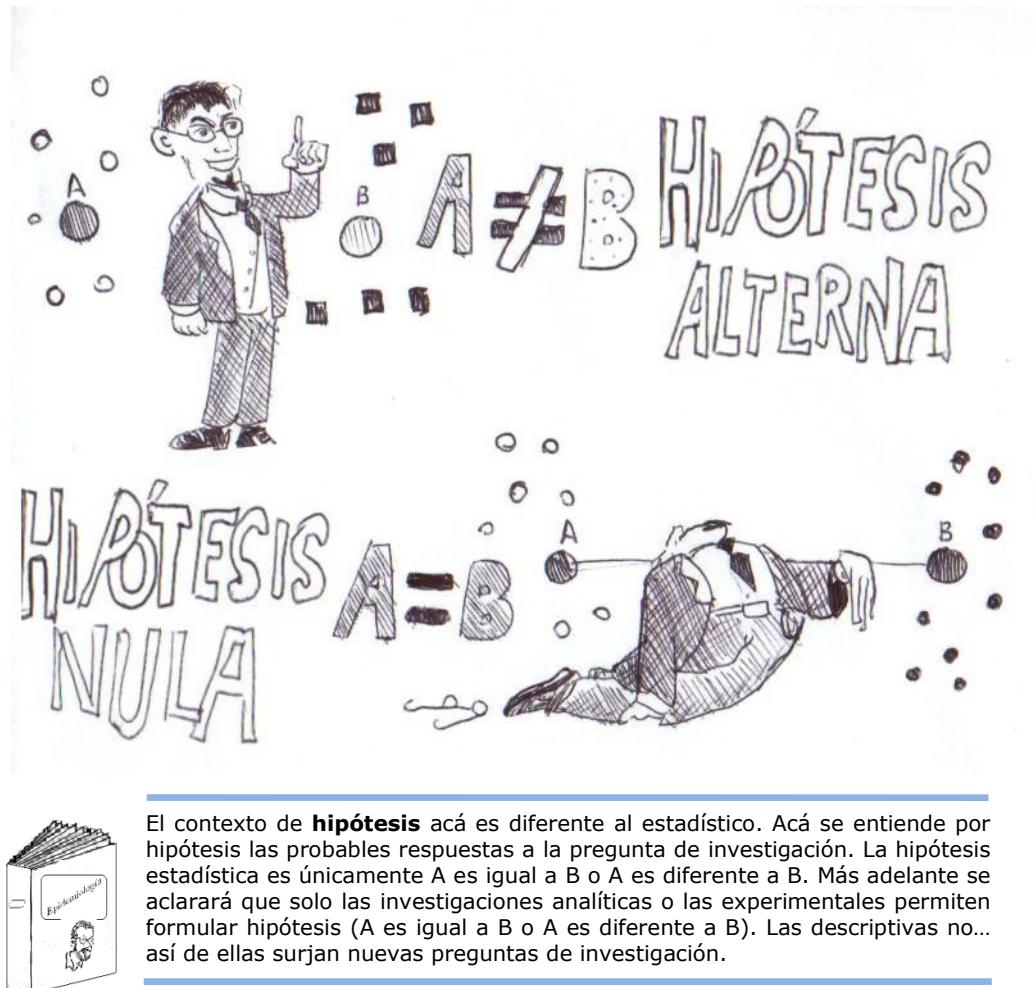
Por hipótesis se entiende las probables respuestas que se pueden encontrar al hacer una pregunta. Como respuestas que son, describen o explican. Cuando describen no existe interpretación distinta a lo hallado. Cuando comparan, pueden encontrar dos posibilidades: que se relaciona o que no lo comparado (o si existe o no, diferencia entre lo comparado).

Es muy importante tener en cuenta de qué se trata la hipótesis. Implica saber a dónde se puede llegar. Dicho de otra manera, cuando el investigador se hace una pregunta, ha de saber cuáles son las probables respuestas, no la respuesta (pues si la supiese entonces para qué va a hacer la investigación). Esto tan obvio ha de tenerse en cuenta pues un inconveniente que puede ocurrir es que en algún momento el investigador se pierda entre tanta información obtenida.



A manera de ejemplo: Cuando Alicia en el país de las maravillas se encontró en un punto de donde salían ocho caminos, le preguntó al conejo: '¿Por cuál camino sigo?' Él le respondió: '¿Para dónde va?' Ella le dijo: 'no sé'. Entonces el conejo le dijo: 'vaya por cualquiera'. Esta metáfora se usa para exemplificar que cuando no se sabe qué se busca, entonces no se sabe cómo o qué se hallará (por dónde buscar)...

De acuerdo al tipo de pregunta y al diseño seleccionado, se podrán encontrar respuestas descriptivas o comparativas. En las primeras se obtienen frecuencias. En las segundas, alternativas (uno u otra respuesta de una alternativa previa).



El contexto de **hipótesis** acá es diferente al estadístico. Acá se entiende por hipótesis las probables respuestas a la pregunta de investigación. La hipótesis estadística es únicamente A es igual a B o A es diferente a B. Más adelante se aclarará que solo las investigaciones analíticas o las experimentales permiten formular hipótesis (A es igual a B o A es diferente a B). Las descriptivas no... así de ellas surjan nuevas preguntas de investigación.

Las hipótesis (probables respuestas) buscan contribuir con el conocimiento. Los diseños descriptivos no plantean hipótesis distintas (probables respuestas) a la descripción en mención. Son solo los diseños analíticos y experimentales (como se verá más adelante) los que plantean hipótesis alternativas. Estas en esencia se identifican a través de relaciones de una variable con otra (o con varias). De hecho cuando se hace la pregunta de investigación, en los métodos se escogen variables a través de las cuales (de acuerdo a lo que respondan) se identifica la respuesta. Por tanto, resulta particularmente importante que al elaborar la pregunta de investigación, se identifiquen con claridad las probables respuestas y que esa respuesta surja de los resultados de unas variables entendibles y con medidas así mismo entendibles para lo investigado.

De hecho, a veces y como un lamentable error en investigación, para responder preguntas de investigación se emplean variables que no responden la pregunta o se usan variables que no tienen criterios objetivos de medida lo que no permite identificar válidamente la respuesta.



**Esto es lo que
quiero construir**



**Estas son las
Variables
que escojo**

No valido

**Esto es lo que
"construyo"**



La selección de las variables a través de las cuales se puede responder la pregunta de investigación no siempre se cumple como debería: por ello lo más útil suele ser el emplear variables ya usadas en otras investigaciones y, por ende, debidamente validadas. La gráfica expresa una situación por demás frecuente como es, querer mirar algo (contestar una pregunta de investigación), escoger unas variables inadecuadas para ello y al final obtener una respuesta NO VÁLIDA, de seguro porque las variables seleccionadas no fueron las adecuadas. Un ejemplo puede ser que se esté estudiando si existe relación entre ser adicto al alcohol y la violencia en jóvenes. El error será usar variables donde se pregunta por consumo de alcohol en los jóvenes, PERO no se puede precisar si se es o no adicto. Entonces encontrar que existe relación con comportamientos violentos no permite precisar que es por ser adicto. La solución será usar variables que identifican no solo el consumo sino por sobre todo, la adicción al alcohol.

Las respuestas descriptivas suelen divulgarse usando la tabla de frecuencia. Al publicar los resultados (respuesta a la investigación), el describir las **características de la población** estudiada así como las **frecuencias** de las variables estudiadas, constituye dos componentes siempre presentes y que, sin duda, ayudan a entender. Si la investigación es analítica o, lo que es lo mismo existe una pregunta de comparación, se debe agregar otra tabla donde se muestre qué se confronta con las medidas de comparación (**tabla comparativa** de OR o

RR). Como se observa, las investigaciones comparativas siempre describen y además comparan (tienen las tres tablas de resultados); en tanto que las descriptivas solo tienen las dos tablas de descripción (la de características de población y la de frecuencias).

CARACTERÍSTICA	Nº	%
Sociodemográficas		
Edad		
De 10 a 12 años	4 481	33
De 13 a 16 años	8 029	59
De 17 a 19 años	1 009	8
Estratos		
Estratos 1 y 2	10 460	79
Estratos 3	2 654	20
Estratos 4 o más	76	1

VARIABLE	Nº	%
Conocimientos y actitudes		
Ha recibido información sobre cómo planificar		
Sí y la entendió	4 768	36
Sí, pero tiene dudas	3 295	25
No ha recibido	5 051	38
Quién le enseñó sobre planificación		
Colegio	5 236	38
Padres	4 179	31
Por su cuenta	969	7
Servicio de salud	849	6
Amigos	793	6

En la investigación cuantitativa las tablas descriptivas resultan particularmente importantes y aprender a hacerlas es clave. Buscan dar una información precisa de ciertas características de la población estudiada. En el ejemplo, se muestran dos tablas (ambas de frecuencia) de una investigación con adolescentes escolarizados: una con características socio-demográficas (edad y estrato) y otra de frecuencias de lo encontrado (de lo que se les preguntó) en adolescentes escolarizados. Como se observa solo se usa las frecuencias y los porcentajes.

VARIABLE	SI Nº	SI %	NO Nº	NO %	OR	IC 95% LI	IC 95% LS	Valor P
Conocimientos y actitudes								
Ha recibido información sobre cómo planificar								
Femenino	2 437	52	2 284	48	1,4	1,2	1,6	0,000
Masculino	2 317	46	2 726	54				
Casos incluidos:	7 630							
Cree que el condón protege contra el Sida								
Femenino	1 505	37	2 509	63	0,4	0,3	0,4	0,000
Masculino	2 383	57	1 829	43				
Casos incluidos:	6 409							

La tabla descrita muestra el diseño de una tabla comparativa. Ella debe mostrar las variables que se quieren comparar (por ejemplo, 'ha recibido información sobre planificación'), entre quienes se quiere comparar (femenino y masculino), los datos que se comparan ('sí' y 'no' en número y porcentaje), la medida de la comparación (OR) y la estadística usada en esta comparación (IC95% con sus límites inferior y superior, así como el resultado de la prueba p del Chi cuadrado).

Lo cierto es que los resultados (respuesta a la pregunta de investigación) de cualquier investigación cuantitativa serán en general presentados a través de tablas.

1.4.1. Algo más sobre las hipótesis de investigación

Dado que la pregunta y las probables repuestas se encuentran a través de la revisión teórica (marco teórico que se revisará más adelante), es indispensable buscar en ellas (las respuestas) explicaciones lógicas, razonadas y justificables buscando sean congruentes con las investigaciones efectuadas (buscar criterios de Bradford Hill y consistencia de asociaciones con conocimientos vigentes).

Las hipótesis pueden ser inductivas (de lo particular generalizan), donde se busca predecir de lo visto de manera particular. El otro mecanismo es el deductivo donde la probable respuesta surge de una teoría que pretende ser particularizada. La investigación es el mecanismo que se usa para comprobar esta o aquella respuesta (hipótesis) y, a pesar de que la investigación ofrezca al finalizar una respuesta, es importante mantener cierto escepticismo cuando se trata de generalizar. Se vuelve a insistir en que la duda se debe mantener en lo posible en todo lo que hagamos. Buscar sesgos o errores en el diseño, o en la población, en fin... lo que se quiere decir es que nunca se podrá estar por completamente seguro de una respuesta en investigación y ello es una actitud sana y que impulsa al ser humano.

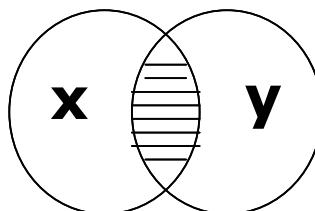
Las probables respuestas (en los diseños analíticos o de intervención) en una investigación surgen o bien, de comparar lo hallado con lo revisado pudiéndose afirmar o, que lo encontrado es igual o diferente a lo revisado.



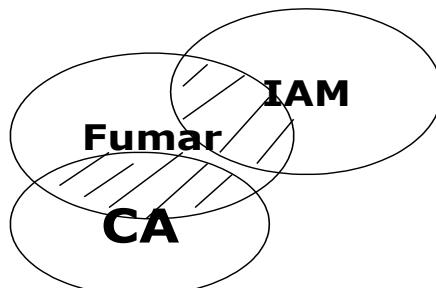
Por ejemplo, cuando se encuentra que la edad prevalente de inicio de relaciones sexuales en una población adolescentes escolarizados de estrato 1 y 2 es de 13,5 años, al comparar en las revisiones bibliográficas con poblaciones similares la misma pregunta se podrá afirmar o que lo hallado es igual o es distinto, formando parte de la discusión la probable explicación de la similitud o la diferencia.

Las **hipótesis** (en las investigaciones analíticas) pueden ser **SIMPLES** que son las que relacionan una variable con otra variable. Pueden ser **COMPLEJAS** cuando se relacionan una o varias variables con otras. Una simple es el plantear una relación entre el cigarrillo y el cáncer del pulmón. Una compleja es plantear la relación entre el tabaquismo y el cáncer del pulmón y la enfermedad coronaria. Más compleja puede ser la relación en los adolescentes entre la deserción escolar y la disfunción familiar con la delincuencia cuando lleguen a adultos.

En investigación las hipótesis surgen de la comparación entre las variables. Si se relacionan solo dos variables (una dependiente y otra independiente) se está investigando para obtener una hipótesis simple.

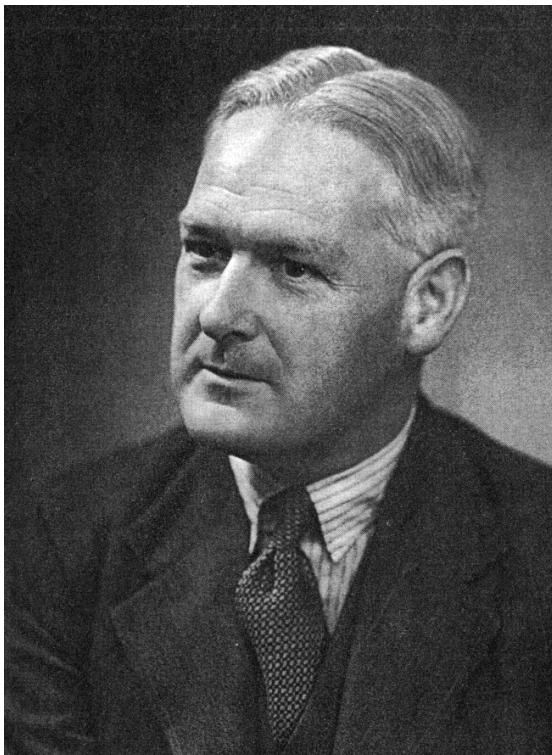


Por ejemplo, si se quiere probar (hipótesis) que los ancianos se caen más que los jóvenes, la variable independiente (x) será la edad de los pacientes y la variable dependiente será las caídas. El área sombreada en la figura, establece que sí existe relación (entre más área de convergencia, más relación).



Las hipótesis complejas plantean relaciones entre más de dos variables. Por ejemplo, el tabaquismo (variable dependiente) se relaciona con el cáncer y con el infarto (variables independientes ambas). La complejidad la provoca la relación entre más de dos variables, y, sin duda, nos acerca más al mundo real donde los eventos se encuentran relacionados con muchas variables.

Otro aspecto interesante al plantear hipótesis es que se pueda **establecer la dirección** (qué fue primero, en los criterios de Bradford Hill) hacia dónde se dirige la relación. El diseño seleccionado para contestar una pregunta resulta particularmente importante. Por ejemplo, algunos diseños (un estudio transversal, por ejemplo) pueden establecer que existe una asociación entre fumar y cáncer pulmonar, pero no se puede identificar qué es primero el cáncer o el fumar; solo se sabe que existe una fuerte relación entre estas dos variables. En cambio, y usando un diseño de cohorte donde partiendo de población que no fuma y que no tiene cáncer y que con el paso del tiempo empieza a fumar y luego le empieza a aparecer el cáncer, se puede establecer que es el cigarrillo el que provoca el cáncer (y no que el cáncer hace fumar a los pacientes).



Ciertamente los estudios observacionales tipo corte o casos y controles establecen relaciones, pero no pueden identificar per se qué fue primero (así el sentido común nos indique cuál fue primero). Los estudios de cohorte sí, por ello también se denominan **estudios CAUSALES** pues permiten establecer causalidad de las enfermedades. El asunto es que es más fácil operativamente un estudio transversal o de casos y controles. Las cohortes implican seguimiento y ello les confiere una mayor planeación. Un excelente ejemplo de seguimiento científico fueron los doctores Doll y Hill quienes en 1950 hicieron una investigación de casos y controles relacionando el cáncer del pulmón y el tabaco. En 1951 iniciaron una cohorte donde indagaron a una población (médicos) si fuman o no. Ulteriormente hicieron un seguimiento (durante los siguientes diez años) de causas de enfermedad y muerte encontrando que el tabaco fue primero (causal) que la enfermedad (cáncer de pulmón).

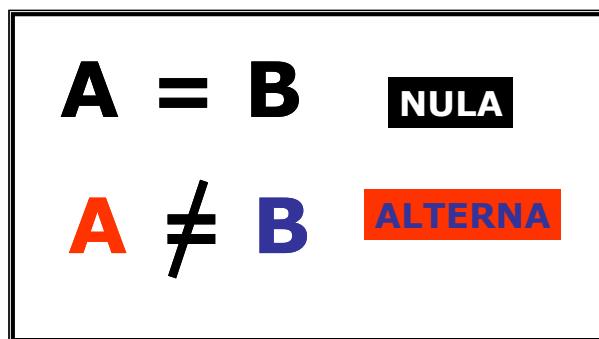
Pero, además, detectaron que fue factor desencadenante de enfermedad coronaria, de enfermedad pulmonar obstructiva crónica. En la foto aparece el Dr. Doll, célebre por sus importantes investigaciones sobre la relación entre cigarrillo y cáncer de pulmón.

1.4.2. Las hipótesis estadísticas

Al plantear las hipótesis, y de acuerdo a los diseños empleados, el proponerlas de manera estadística es importante. Cuando se dice que se espera ver la relación entre el cáncer de pulmón y el cigarrillo, se dice que se está haciendo una declaración de la hipótesis. El plantearla de acuerdo a la estadística requiere dar dos alternativas: **SÍ** existe relación o **NO** existe relación y esto resulta particularmente importante, en especial para demostrarlas. El término de hipótesis NULA es el que habitualmente expresa la ausencia de relaciones. La hipótesis ALTERNA es la que plantea la presencia de relaciones. Expresado de esa manera sería:

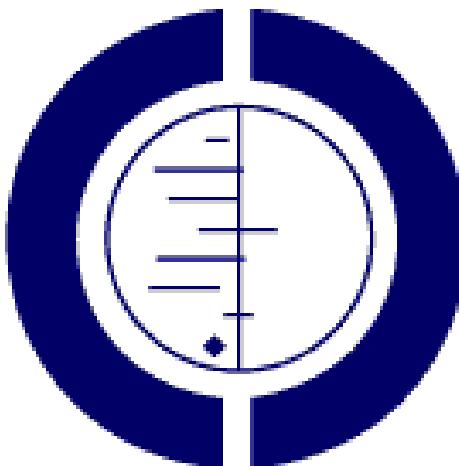
A = Exponerse al cigarrillo
B = No exponerse al cigarrillo

La hipótesis nula es que es igual exponerse que no exponerse en el desarrollo del cáncer pulmonar ($A = B$). La hipótesis alterna es que exponerse al cigarrillo provoca diferencias en torno a la aparición del cáncer de pulmón al compararse con no exponerse (A es diferente a B).



Las hipótesis, como predicciones que son, encuentran o no relaciones entre hechos descritos con variables. En general, sus hallazgos son el resultado de teorías que casi siempre el investigador **revisó a través del marco teórico**. En esencia, las hipótesis fomentan el pensamiento crítico y claro.... facilitan el entendimiento del mundo (buscan hacerlo). Las investigaciones que describen SIN RELACIONAR no son susceptibles de emitir hipótesis. Sin embargo, son en esas descripciones donde los investigadores buscan las relaciones que plantearán ulteriormente. Idealmente toda hipótesis parte de argumentos coherentes, razonados y sólidos, productos de otras investigaciones. Pueden ser **INDUCTIVAS** o que proponen generalizar de lo particular o **DEDUCTIVAS** que pretenden particularizar de lo general. Cuando se plantea una hipótesis se debe mostrar con nitidez las dos posibilidades, expresadas bajo estas dos formas: existe o no existe relación o, según sea el caso, es igual o es diferente.

El planteamiento hipotético parte de que se encontrará diferencia entre lo revisado y lo hallado. Como necesita de un planteamiento alterno (no se encontrarán relaciones o diferencias) se propone la **HIPOTESIS NULA** que plantea lo contrario de la **HIPOTESIS ALTERNA**, que dice que sí se encontrarán relaciones o diferencias de acuerdo a las variables utilizadas (NULO y ALTERNO son términos derivados de la estadística).



Por ejemplo, cuando se prueba un medicamento para una enfermedad que provoca la muerte, obviamente la respuesta se verá en el grupo que recibe el medicamento (A) y se confronta con el grupo que no (B) que recibe un placebo.

El planteamiento hipotético es que los que reciben el medicamento mueren menos que los que no lo reciben (B). Entonces lo esperado es que A sea diferente a B. La hipótesis nula es $A = B$ que significa que no existe diferencia entre los dos grupos (recibir medicamento vs. recibir placebo). El logo que aparece es el de la Fundación Cochrane que se ha propuesto revisar sin conflictos de interés los medicamentos. El logo en sí mismo representa una escala gráfica de intervalos de confianza.

Las hipótesis estadísticas se confirman (obviamente) mediante pruebas estadísticas de hipótesis. Como se vera en el capítulo correspondiente, la prueba p (o del Chi cuadrado) o la de los intervalos de confianza son las usadas con mayor frecuencia.

1.5. Paso 5: Pertinencia de contestar la pregunta de investigación (o importancia de la respuesta de la pregunta)

Planteado el problema, el investigador debe preguntarse la importancia de contestar la pregunta (**pertinencia**). Explicar por qué es importante hacer la investigación, expresando con claridad lo que se espera (aportes) de encontrar la respuesta, es una de las exigencias que hacen muchas entidades antes de financiar las propuestas. Preguntarse (y expresar) si se espera que los resultados aporten nuevos conocimientos, o si permiten conocer más a la población, o si modificarán conductas sociales, son elementos que permiten ver la pertinencia. Por supuesto que la definición de este aspecto puede resultar importante para los entes financieros.



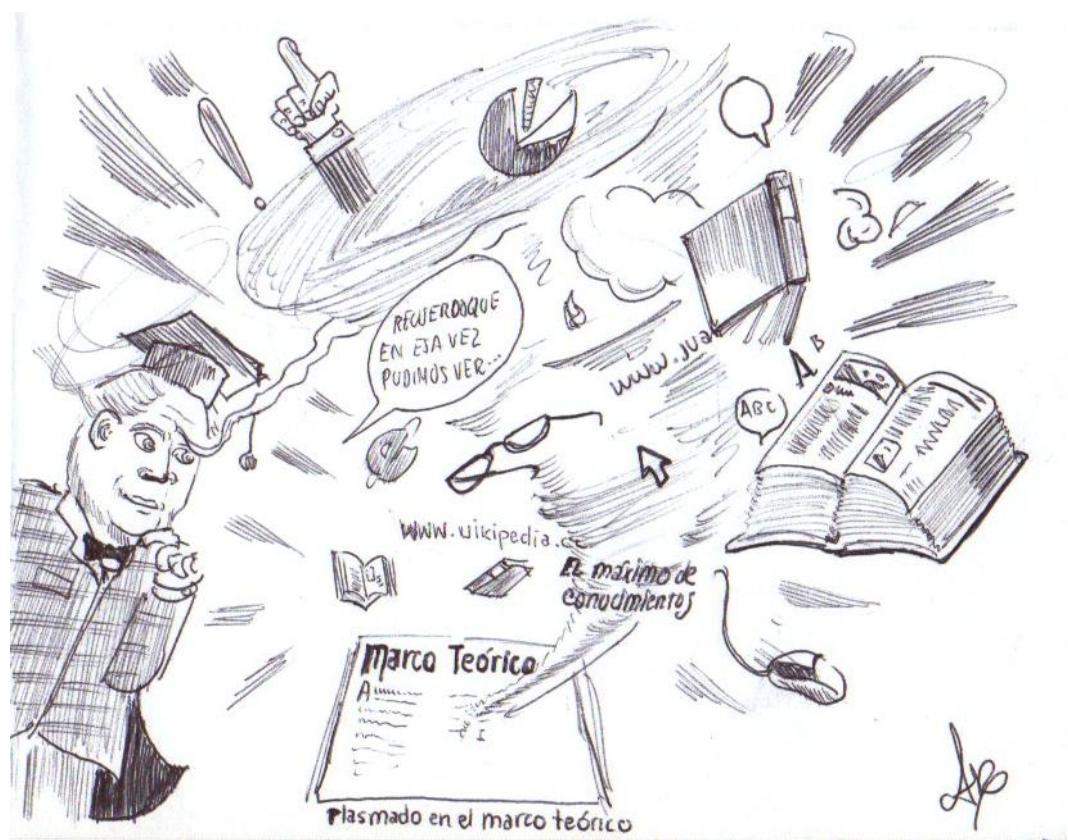
Existen entidades que promueven la investigación otorgando los recursos y solicitando la presentación de protocolos en donde se demuestre que lo que investigará está en sus líneas de interés. El estudio de nuevas moléculas con interés terapéutico, por ejemplo, es una línea de las farmacéuticas o los estudios de prevalencia para construir historias naturales pueden ser del interés de entidades gubernamentales. Lo cierto es que de acuerdo a la pertinencia que se vea de contestar la pregunta de investigación surgirá la posibilidad de obtener una fuente de financiación. La imagen es el logo del Instituto Nacional de Salud de los Estados Unidos, probablemente el ente que más investigaciones biomédicas financia.

Para identificar la importancia de responder una pregunta de investigación siempre ha de tenerse en cuenta cuál es el aporte que le hace a la población objeto de la investigación. En salud se debe

asegurar que la población que nos ayuda se beneficie (por supuesto, nunca, nunca, se le puede ocasionar daño siendo clave verificar los riesgos existentes). En segundo lugar, identificar qué vacíos existen en la literatura (que el investigador revisó) y qué va a resolver la investigación. Para esto se hace indispensable buscar qué se ha escrito al respecto (del problema). Esta búsqueda es esencial para darle consistencia (peso) a la pregunta y, por tanto, a la investigación. Busca como objetivos ver cómo otros han abordado el problema. El planteamiento de los objetivos de la investigación (lo que se pretende alcanzar) permite ver en buena parte la pertinencia.



 Es muy importante, cuando se piensa en la investigación, que se tenga en claro la forma como se beneficiará la población objeto. Es el caso de investigaciones en grupos poblacionales, tales como colegios, donde la población adolescente escolarizada es la que será estudiada. Los investigadores deben pensar en cómo podrán beneficiar a la población que participa. Por ejemplo, en las encuestas de detección de prevalencias de riesgos, ofrecer consulta para orientar aquellos participantes que así lo deseen, puede ser una forma. Por supuesto que la presentación de los resultados en los colegios con sus comunidades (estudiantes, profesores o padres) se da por descontado.

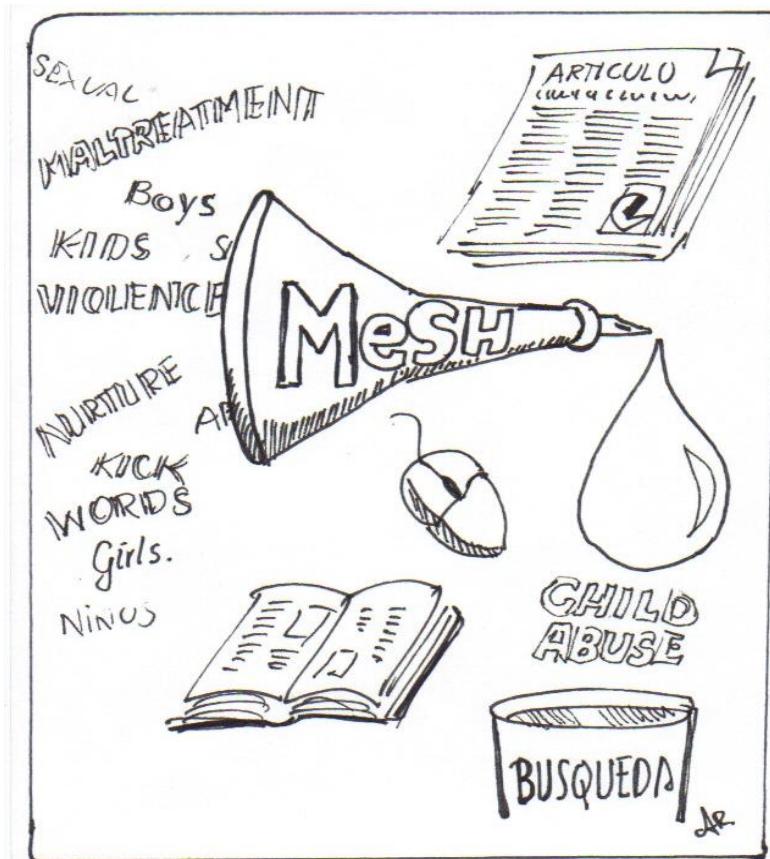


1.6. Paso 6. Construcción de un marco teórico

El marco teórico es donde el investigador busca y describe el máximo acerca de su investigación. De lo profunda que sea su investigación acerca de lo escrito y descrito sobre su problema de investigación, se construye la solidez de su trabajo. Dado que exige el esfuerzo de buscar (revisión bibliográfica) y harto... es clave (y se repite) que al investigador le guste el tema y el problema, pues lo que suele suceder es que al finalizar sus revisiones, se ha empezado a convertir en un experto. Claro... la ejecución de la investigación le da práctica cerrando el círculo de la solidez del conocimiento. Un buen marco teórico deberá abordar de cada componente, cómo se define, qué relaciones o factores causales existen en torno a tal componente, qué variables lo miden y qué datos se usan.

1.6.1. Definición de los componentes del problema:

Si ha sido posible construir una pregunta de investigación, **decidir si la pregunta elaborada será o no la definitiva**. Al revisar sobre el tema en más de una ocasión se enfrenta el investigador al cambio del problema o una mejor elaboración de la pregunta. La revisión permite establecer qué **SE SABE** y, con base en ello, precisar lo que se quiere saber, por ende precisará más la pregunta de investigación. En esta especificación de la pregunta de investigación puede resultar importante descomponer la pregunta en sus partes y definir cada una con nitidez.



Como se expone, son los componentes de la pregunta los referentes que se usan para hacer la búsqueda. Es muy importante buscar cómo se describen médica y científicamente esos términos (términos MeSH). Por ejemplo: no aparece si es frecuente o no tal cosa, sino cuál es la prevalencia (prevalencia es el término que usa las fuentes de información de salud).

La definición de los términos debe permitir una revisión de la población objeto y, por sobre todo, surgen las variables que se pueden emplear para contestar la pregunta.

1.6.2. Qué respuestas se conocen a la pregunta

De la revisión se debe prever unos resultados de acuerdo a investigaciones similares revisadas, resultados que permiten identificar unas frecuencias (que comparará con lo encontrado) y si el diseño lo permite, esperar unas relaciones (por ejemplo: *este aspecto influye más en estos que en aquellos*) o proyectar el resultado de una intervención (como: *los que reciben esto viven más que los que no lo reciben*) lo que se conoce formalmente como HIPOTESIS DE LA INVESTIGACIÓN o predicción de lo que el investigador cree que encontrará. Estos aspectos resultan claves en el momento de la DISCUSIÓN, que es donde el investigador confronta la revisado (y, por consiguiente, lo esperado) con lo encontrado.



Cuando nos hacemos una pregunta de investigación a lo mejor tenemos una respuesta preconcebida. Pero puede que ella no sea la respuesta verdadera. Por ello, resulta particularmente importante revisar cómo otras investigaciones la han contestado. Así ampliaremos nuestro panorama de la realidad y entenderemos la importancia de encontrar la respuesta en la población seleccionada (o, lo que es lo mismo, hacer la investigación).

1.6.3. Cómo la han contestado en otros estudios (diseños empleados, variables y manejo estadístico)

Una buena búsqueda debe permitir observar cómo otros investigadores abordaron la pregunta, lo que resultará importante para seleccionar el diseño de la investigación. Ahí, se hace especial atención al abordaje de la población, al manejo ético, al manejo del tiempo etc.... Lo cierto es que el investigador podrá confrontar sus recursos con otros diseños para así seleccionar cómo contestará su pregunta (aprende de los otros). Se debe identificar qué cuestionarios (variables) se han empleado, así como su transformación en medidas que permitan ver la respuesta. Así mismo, de cada medida empleada (variable con su dato correspondiente) observar el manejo estadístico empleado (cómo describirá y comparará cada medida).



iOjo!, esta búsqueda de variables es particularmente importante. Una tendencia ingenua es INVENTARSE el cuestionario. Al hacerlo, incurrimos en la necesidad de validar tal instrumento o, lo que es lo mismo, demostrar que ciertamente mide lo que se pretende. Por ello, es mejor usar instrumentos ya validados (usados en investigaciones previas en donde se demostró que miden lo que se busca). Además, y como otra ventaja, al usar variables ya empleadas en otros trabajos, se puede apreciar el manejo estadístico otorgado a la variable, lo que sin duda favorece su empleo. Por supuesto que el investigador puede tomar variables de distintas experiencias investigativas y diseñar un cuestionario original, obteniendo la validez de las revisiones originales de cada variable.

Lo cierto de todo esto es que gracias a la revisión efectuada, el investigador debe precisar:

1. Qué diseño va emplear
2. Cuáles variables va a emplear
3. Cuál será el manejo estadístico de cada variable



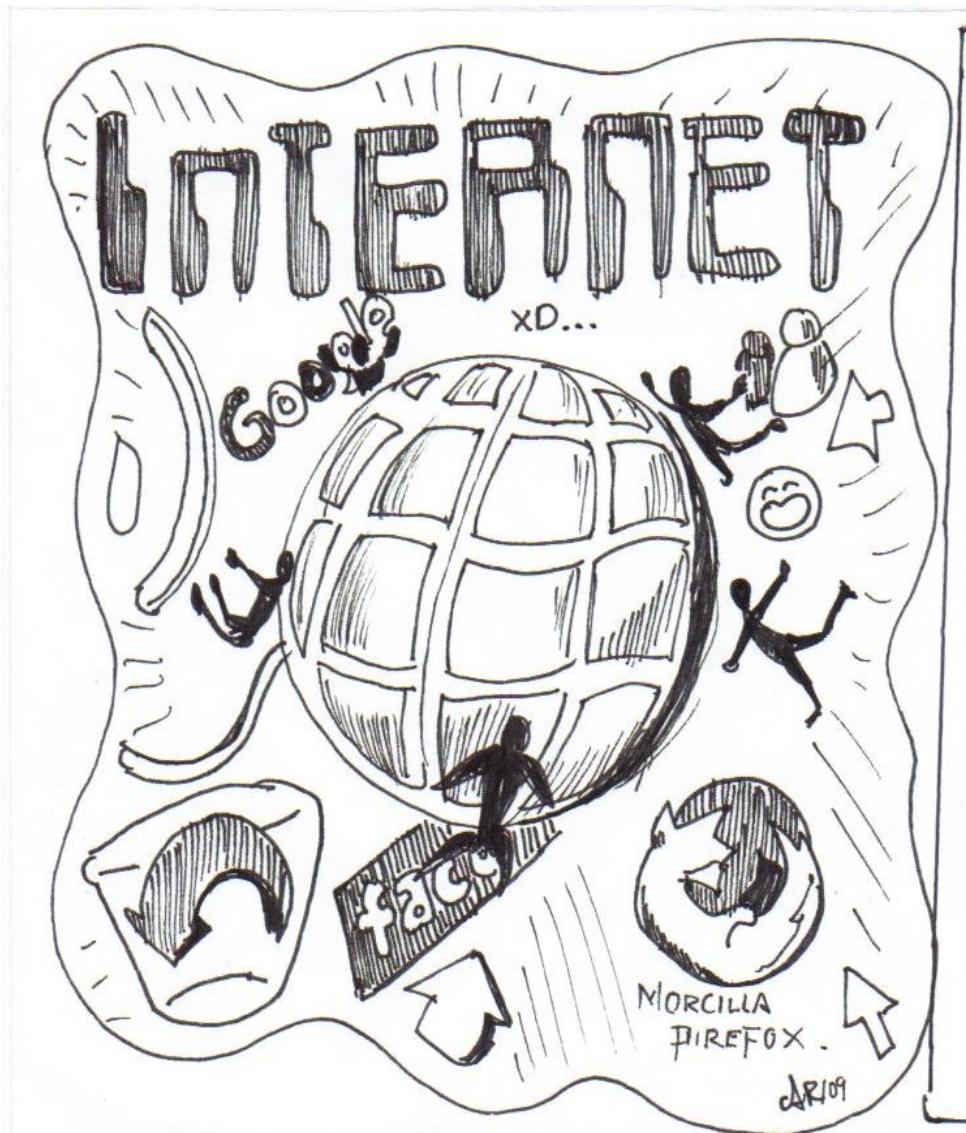
En una sociedad del conocimiento donde abunda la literatura, prácticamente es improbable que no se encuentre algo acerca del tema o de la pregunta. Sin embargo, la real dificultad la establece la capacidad de búsqueda del investigador. Buscar investigaciones previas, teorías acerca de la pregunta (o de sus componentes), diseños de investigación, opiniones e incluso anécdotas son fuentes de riqueza para la estructuración de la investigación. Las tesis doctorales en general son documentos que pueden estar al acceso de la comunidad científica, en donde, y por principio, las revisiones a profundidad de algunos de los componentes de la pregunta se presentan. Los artículos, si bien son más breves, permiten identificar a través de las citas bibliográficas las disciplinas interesadas y las revistas que lo abordan. Por supuesto que la capacidad del investigador en hacer una lectura crítica es clave. Para ello, el identificar las pregunta que se hace en la investigación, el diseño escogido para contestarla, la población objeto del estudio, las variables empleadas, los resultados obtenidos y el manejo estadístico junto con la discusión (aportes al conocimiento), resulta invaluable. La validez de lo leído está dada por la distancia entre lo presentado y la realidad y el

análisis de los sesgos (errores) se convierte en una actitud que debe mantener el investigador. Será una tarea importante del investigador el hacer resúmenes de lo que lee, con el fin de fortalecer su conocimiento y, además, el darle respaldo a su estudio.

Se conoce como **fuentes primarias** a los estudios originales, donde se presentan investigaciones a cargo de quienes las hicieron (en general los artículos publicados). Fuentes secundarias son los resultados de análisis de otros que, si bien no hicieron la investigación, efectúan una discusión de lo encontrado por los investigadores.



*La búsqueda de los estudios o de las publicaciones es, hoy por hoy, un proceso al alcance de todos. Las bases de datos electrónicas son sitios en donde se pueden encontrar una inmensa cantidad de artículos, clasificados de distintas y ordenadas formas. Una base de datos puede agrupar una inmensidad de revistas. Las bibliotecas donde están habitualmente tienen acceso a la totalidad del artículo; sin embargo, existen muchas en donde solo se presenta el resumen del artículo y la revista donde está, debiéndose pagar para poder acceder al estudio completo. La base de datos de la **National Library of Medicine** probablemente sea la más completa. Por ello, el aprender a buscar información en ella es una herramienta muy útil en las revisiones. El acceso a los resúmenes (summary) es gratuito.*



Una de las diferencias que existían entre los investigadores de países desarrollados y no desarrollados eran las bibliotecas. Los primeros tenían enormes ventajas en torno al acceso a la información. Hoy, gracias a las bases de datos y a las conexiones de Internet, el acceso a la información es más equitativo.

1.6.4. Recopilación del marco teórico (qué debe decir una tesis)

El paso 6 o construcción de un marco teórico no es otra cosa que la prueba de la revisión que ha hecho el investigador. Recopilar esta revisión en un documento sin duda es un esfuerzo válido, obligatorio en la modalidad de investigación donde se hace la presentación de tesis y no tanto cuando se publica el artículo para una revista. Si bien existen muchos formatos de cómo presentar un marco teórico en las tesis, uno puede ser el siguiente:

1. Introducción. Compuesta por:

1.1. Identificación del problema o, lo que es lo mismo, una explicación debidamente referenciada de la importancia del tema, donde se plantea lo que hace falta saber al respecto y los problemas que de acá se derivan en el mundo. Claro, tal introducción ha de estar dirigida a llegar a la pregunta de investigación.

1.2. La pregunta de investigación. Al hacerlo: Definir cada uno de los términos que componen la pregunta con la debida referencia.

2. Propósitos y objetivos. Se expone:

2.1. El objetivo general o mención de lo que pretende lograr la investigación.

2.2. Objetivos específicos o desglose del objetivo general. Claro, estos objetivos son los que se debe haber cumplido al finalizar la investigación.

3. Revisión teórica. Donde se exponer el producto de la revisión de cada uno de los componentes de la pregunta. Está compuesto por:

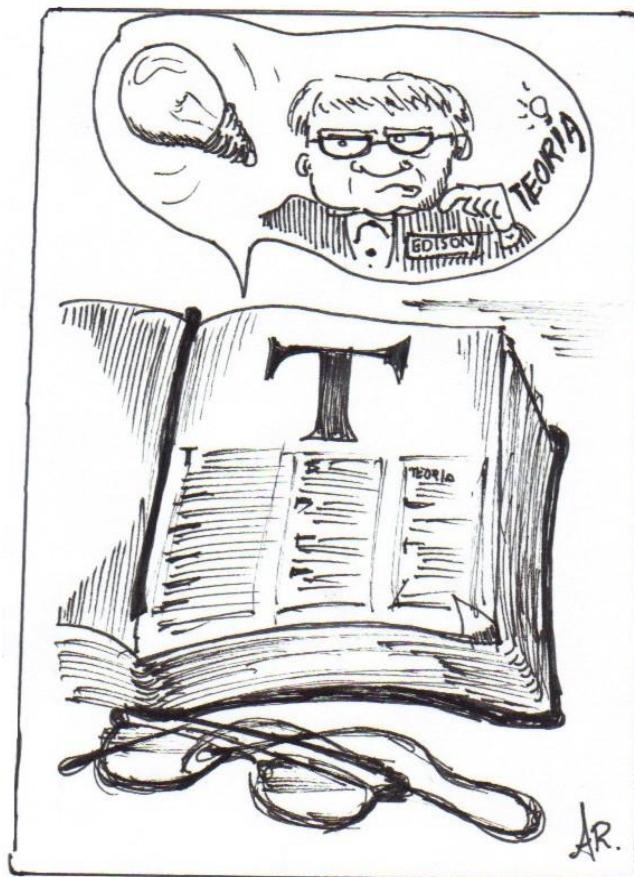
3.1. Qué se sabe hasta el momento de cada uno de los componentes del problema, claro... debidamente referenciado. Se debe explicar muy bien cómo están unidos en otras investigaciones los componentes, explicar modelos (explicaciones de cómo o por qué funcionan las cosas). Mostrar frecuencias que sirvan para comparar después. Qué respuestas ha investigado en torno al problema, mencionando cómo los otros investigadores interpretan esos resultados (discusión).



Por ejemplo, si el problema es '¿Cuál es la relación entre los comportamientos de riesgo y la familia (estructura y función) en la población adolescente escolar?' Los componentes son: Adolescentes escolares, comportamientos de riesgo, género y familia. Entonces, el marco teórico no debe olvidar la integración de estos tres aspectos. Una propuesta de su desarrollo puede ser el siguiente:

1. Describir a la adolescencia, clasificarla (pues en la investigación de seguro se ordenará). De esta revisión surgirán las variables con los que la investigación describirá a los adolescentes. Por ejemplo, los clasificará de 10 a 12 años como púberes, de 13 a 15 años como exuberante y de 16 a 19 años como tardíos [Clasificación de Carvajal. Adolescencia: la aventura de una metamorfosis. Una versión psicoanalítica de la adolescencia. Bogotá: Tiresias].
2. Explicar qué es el riesgo, cómo se clasifica, cuáles son los comportamientos de riesgo que asumen los adolescentes, explicarlos, mencionando qué teorías existen en torno a ello. De esta revisión surgirán las variables de comportamiento de riesgo.
3. Definir familia, explicar cómo se clasifica y cómo se mide en torno a su estructura y en torno a su función. Explicar qué relación existe entre los adolescentes y la familia y la relación entre comportamientos de riesgo y estructura y función familiar.
4. Mencionar qué estudios han abordado la pregunta de investigación y describir resultados y explicaciones. Acá mencionar cómo han manejado población, qué tipo de diseños y claro... las variables empleadas.
5. Describir lo que se sepa de la población objeto.
6. Cuestionario: Se debe dejar una parte del marco teórico para describir las variables que se emplearán, debidamente referenciadas.

Casi es hacer un libro acerca de la pregunta, de hecho muchos libros surgen de las investigaciones y, el marco teórico es el responsable del mayor desarrollo. Mucho del marco teórico surge de la revisión de artículos, por tanto, aprender a hacerlo resulta particularmente importante.



Un buen marco teórico debe identificar **teorías** acerca de lo que se quiere indagar, entendiendo una teoría como una descripción detallada de un fenómeno que permite proyectar la trascendencia de lo que se investigará. Las teorías científicas relacionan conceptos y, a través de esas relaciones, buscan deducir, intentando llegar a la predicción, que en últimas es una de las razones esenciales de la investigación positivista.

 Por ejemplo, en la búsqueda del origen del SIDA, surgió de las descripciones en torno a la población que lo padecía, la investigación de una relación común. La teoría descriptiva en su momento mostró cómo la población homosexual, prostitutas, drogadictos, hemofílicos y pacientes trasfundidos, constituyían la población que lo padecía con mayor frecuencia. La pregunta que se hizo fue, y partiendo de la sospecha de tratarse de una enfermedad infecciosa, ¿cuál es el elemento común que interviene en la cadena de la infección de estos pacientes? De ahí surge la sangre como un probable responsable y con base a esta sospecha se toman medidas, prediciendo que con dichas medidas se puede controlar en algún grado la propagación en aquel momento de la supuesta infección.

Lo cierto es que una **TEORÍA** en sí es una exposición donde se describen los fenómenos como conceptos y se demuestran relaciones que permiten explicar cómo y por qué y, por lo tanto, ser verificada a través de lo que se ve (empirismo). Lo que se busca es que, entendiendo tal fenómeno, se pueda **PREDECIR** su probable comportamiento.



A pesar de que se pueda creer que las personas piensan de manera diferente, la realidad grupal (grupal = personas que viven con características similares) es que la mayoría piensa de una misma forma. Esto es por demás fácil de entender, pues una persona en general se desarrolla con otras que lo afectan poderosamente, influencia que se traduce en similitudes en la forma de ver las cosas. Claro... en el dibujo se expresa ante un mismo objeto dos maneras de pensar distintas, pero es que se trata de dos personas que pertenecen a grupos diferentes. ¡Y cómo es de difícil que se pongan de acuerdo dos personas que piensan distinto...!

Así mismo, un buen marco teórico referencial debe permitir identificar los **MODELOS CONCEPTUALES** existentes en torno a la pregunta. Un modelo es otro intento de relacionar conceptos igual que las teorías, pero, a diferencias de las mismas, acá no existe tanta solidez, lo que

permite un manejo más flexible. En esencia, los modelos no son verificables (lo que sí debe tener una teoría). Por ejemplo, los modelos de salud, son explicaciones, donde se relacionan conceptos y, a través de ello, se pretende proponer un actuar. Sin embargo, la propuesta no ha sido probada (que por las relaciones y conceptos expuestos, se intuye que puede ser...).

De estos dos aspectos, surgirán las hipótesis (probable respuesta) de investigación. Los investigadores buscarán, o probar o negar las teorías o los modelos conceptuales revisados ó aportar a través de la descripción, nuevos fenómenos de interés.

Hoy, y como un importante aporte, vale la pena incluir el **MARCO DE REFERENCIA** entendido como la descripción del método escogido por el investigador para hacer la búsqueda de información bibliográfica.

Pero, ¿se pueden hacer investigaciones sin tomar estos aspectos como referentes? La respuesta es SÍ... pero resulta evidente que sin la revisión, medir la trascendencia de lo investigado es menos probable y resulta por demás obvio, que el peso de un investigador que conoce el tema a investigar es mayor (mucho mayor...) de aquel que no.



El marco teórico no suele aparecer en los artículos publicados, pues éstos buscan ser lo más concretos posibles; dicha indagación se deja a las publicaciones tipo tesis o textos.

Como se observa, se propone que los investigadores definan su investigación luego de la construcción del marco teórico. Solo después de esta revisión de conceptos, teorías, modelos y términos de referencia, el investigador tendrá la capacidad de saber en definitiva si la búsqueda de la respuesta a la pregunta efectuada es a la que quiere dedicarle el esfuerzo por venir. Con estos elementos tenidos en cuenta, la identificación de un **PROPOSITO** resulta un esfuerzo relativamente fácil. Y el desmenuzarlos en objetivos específicos es, así mismo, más factible y sencillo.

Siguiendo con el formato tesis, luego de la construcción del marco teórico y partiendo que ya se concluyó la investigación, en el protocolo sigue:

4. MATERIALES Y MÉTODOS: Acá se expone:

- 4.1.** Tipo de estudio efectuado
- 4.2.** Características de población estudiada
- 4.3.** Variables empleadas
- 4.4.** Recolección de información.

5. RESULTADOS

- 5.1.** Características encontradas de la población estudiada
- 5.2.** Frecuencias encontradas de las variables estudiadas (con el manejo estadístico pertinente)
- 5.3.** Respuesta a la pregunta de investigación (con el manejo estadístico pertinente).

6. DISCUSIÓN

- 6.1.** Exposición de limitantes de la investigación (sesgos, inferencias)
- 6.2.** Comparación de frecuencias encontradas con revisadas y su explicación
- 6.3.** Explicación de la respuesta a la pregunta encontrada, comparando con otras y ofreciendo su debida explicación.

Cómo se observa, esta presentación no difiere gran cosa del artículo publicado en cuanto a sus partes. La gran diferencia lo establece la extensión. En la tesis el investigador lo pone todo pues con base en ello será evaluado. En el artículo se debe reducir al máximo, pues solo se quiere contar lo que se encontró y no demostrar la capacidad del investigador.

FASE DOS: DISEÑO Y PLANEACIÓN (PLAN PARA CONTESTAR LA PREGUNTA)

Identificada la pregunta viene el importante proceso de planear. Planear es predecir que si se hace algo se obtendrá un resultado específico. Seleccionar el diseño y verificar los recursos necesarios para hacer el trabajo de campo constituyen el objetivo de esta fase.

2. Paso 7. Selección del diseño que se empleará

Todo problema o pregunta de investigación puede ser contestado de distintas formas. De acuerdo a la manera como se haga, se podrá o no hacer inferencias a otras poblaciones distintas a la investigada. La forma que se decida es lo que se denomina DISEÑO de investigación (TIPOS).

Como se mencionó en el capítulo introductorio, existen dos grandes tendencias en torno a la forma de contestar las preguntas, agrupadas bajo los términos CUALITATIVO y CUANTITATIVO. La primera es especialmente empleada en las disciplinas de las ciencias humanas. La segunda, en las ciencias naturales. En el caso concreto de la salud, la disciplina dominante es la epidemiología y emplea diseños cuantitativos. Una importante diferencia entre los tipos cualitativos y cuantitativos, lo establece la flexibilidad en los diseños cualitativos y la rigidez cuantitativa. De hecho, el diseño cualitativo es eminentemente exploratorio y busca entender, en tanto que el cuantitativo busca inferir para con ello predecir.

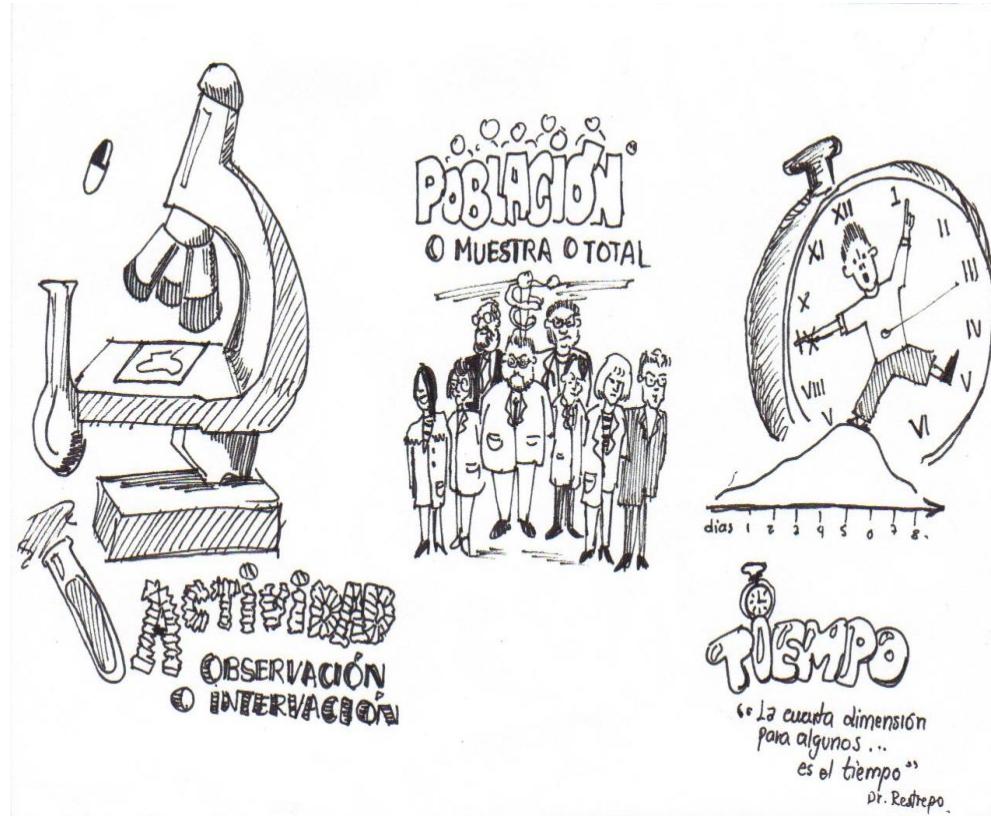
Para escoger una forma de contestar la pregunta de investigación, se ha de tener en cuenta que el diseño sea **PERTINENTE** (que se relaciona) con la pregunta y dado que una misma pregunta se puede contestar de distintas maneras, el investigador deberá escoger cuál usar a sabiendas que los que más asemejen la realidad, son los que tienen mayor validez. Así mismo, ha de tener en cuenta la posibilidad de **ERRORES** (distancia con la realidad) de cada diseño. Tener en cuenta a la población o a los cuestionarios que va a emplear, cómo los va a aplicar y cómo el diseño procura garantizar que no existan selecciones de población sobre las que no se pueda sacar conclusiones o garantizar que las respuestas obtenidas sean ciertas. La búsqueda de un diseño que permita **precisión** (si se repite se puede obtener la misma respuesta) también es importante. Otro aspecto es conseguir la potencia en los estudios. La **POTENCIA** es la capacidad de un estudio en detectar relación entre las variables y eso se consigue con incremento en los tamaños de la población estudiada. Y claro, definitivamente la **FACTIBILIDAD** (que se pueda hacer) entra a jugar un papel primordial al momento de escoger. Existen unos diseños más fáciles que otros por la logística que implican.



Por ejemplo, si se quiere saber si el cigarrillo provoca cáncer se ha de escoger un diseño diferente al de indagar si existe relación entre el cáncer y el cigarrillo, ya que el solo hallazgo de la relación no precisa qué fue primero. Ahora bien, si se toman personas que no tienen cáncer y se observa a través del tiempo quiénes empiezan a fumar y luego comparar entre los que fuman y los que no, en qué grupo se presentan más casos de cáncer, se podrá precisar que fue el cigarrillo el responsable de la aparición del cáncer (y no el cáncer el responsable de que se fume). Claro... este último diseño tiene problemas de logística importantes, pues se debe observar por mucho **tiempo** (de hecho los años suficientes para que aparezca el cáncer), debe pensar en cómo **mantener la población**, cómo medirla (**seguimiento**) siendo apenas algunos de los problemas a enfrentar. En el primer ejemplo, hacer una encuesta en la población escogida es, sin duda, más sencillo... ¿cierto? Estos diseños tendrán mucho que ver con la validez, siendo los más válidos los que se asemejan más a la realidad. Las cohortes se acercan más a la realidad que los estudios de corte o de casos y controles... ¿cierto? Pero resultan más difíciles de elaborar (más adelante se aclara más).

2.1. Diseños cuantitativos

Para definir un diseño cuantitativo se ha de tener en cuenta tres aspectos: la **ACTIVIDAD** a realizar, el **TIEMPO** en que se hará y la **POBLACIÓN** sobre la que se hará.



En torno a la actividad, los diseños cuantitativos permiten tres acciones:

- Describir. Por ejemplo, describir las características de una enfermedad o de un fenómeno.
- Comparar. Por ejemplo, ver si se relaciona o no una variable con otra.
- Intervenir. Por ejemplo, medir si un medicamento es útil o no en un grupo de pacientes.

En torno al tiempo, se deberá escoger entre hacer una revisión hacia el pasado o un estudio de este momento o un estudio hacia el futuro. Términos sinónimos son:

- Estudios retrospectivos. Por ejemplo, ver lo que los pacientes presentaban en historias clínicas.
- Estudios transversales o de corte: Por ejemplo, hacer una encuesta y obtener información ya o del momento.
- Estudios de prospectivos: Donde se hace evalúa a los pacientes en distintos momentos (para poder comparar). Ejemplo son los estudios de cohorte.

Finalmente, en torno a la población objeto de la investigación, se debe identificar:

- Sobre quién: Las investigación puede ser sobre **individuos** o sobre **grupos de población**.
- En quién: Puede ser sobre el total de la población (**UNIVERSO**) objeto o sobre una **MUESTRA**.
- Criterios de selección. Cómo se escoge la muestra: o de manera **ALEATORIA** (que cualquiera de la población objeto pueda ser escogido) o por **CONVENIENCIA** (que no fue aleatoria).

Teniendo esto de presente, una forma de clasificar a los diseños cuantitativos puede ser:

A: EXPERIMENTAL

- A.1. Experimental puro
- A.2. Ensayo clínico
- A.3. Ensayo de campo
- A.4. Cuasiexperimental
- A.5. Intervención comunitaria.

B. OBSERVACIONAL

- B.1. DESCRIPTIVOS. Series de caso
- B.2. ANALITICOS
 - B.2.1. Corte
 - B.2.2. Casos y controles
 - B.2.3. Cohorte
 - B.2.4. Ecológicos.

2.1.1. Intervencionales o experimentales

La actividad esencial de la investigación experimental es la **INTERVENCIÓN**, queriendo con ello decir que existe una variable colocada por el investigador sobre una población y cuyo efecto se pretende medir. En segundo lugar, todos los estudios experimentales son manejados en torno al tiempo de manera prospectiva, donde se mide antes de la intervención a la población objeto y después mínimo una vez, pero usualmente varias veces después (y así poder identificar hasta cuándo aparece o desaparece el efecto). Por supuesto que para poder medir ese antes y ese después, se requiere de un grupo de comparación. Así que se mide sobre la población **INTERVENIDA** y sobre otra de **CONTROL** (que es la de comparación). Obviamente en lo posible no debe existir diferencias entre una y otra población antes de la intervención, para luego poder ver si se presentan diferencias. Son entonces estos los parámetros que se han de tener en cuenta en torno a los estudios experimentales. Los tipos son:

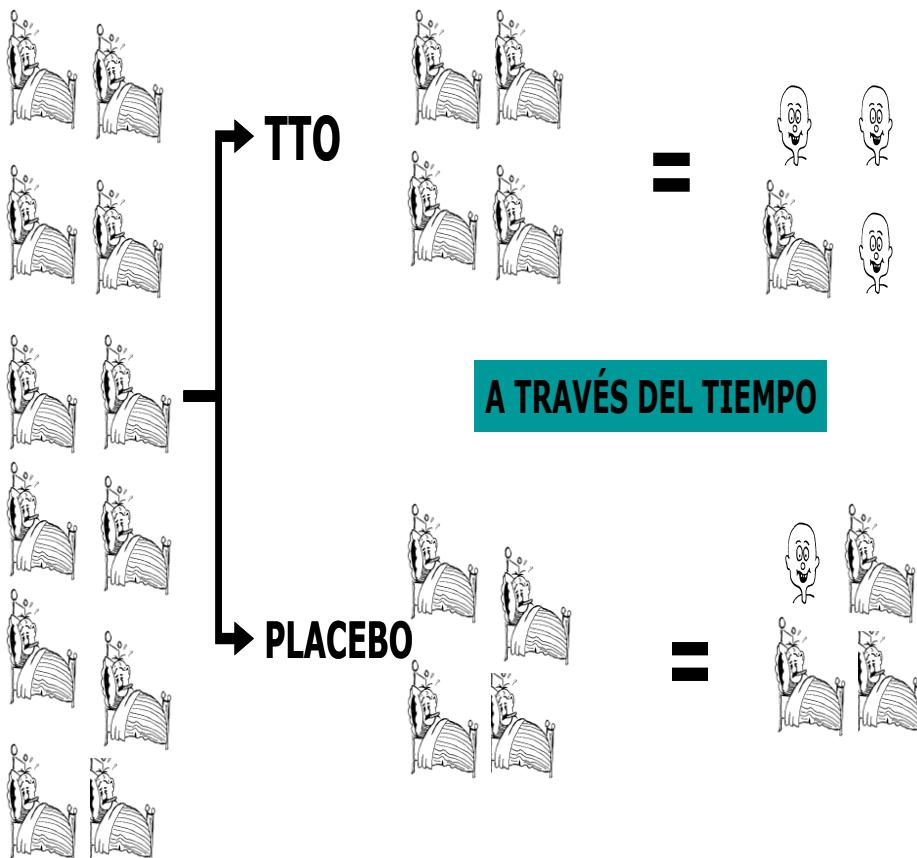
2.1.1.1. Experimental puro: La mejor forma de entenderlos es pensar de una vez en la experimentación derivada de la prueba de los medicamentos. La búsqueda de moléculas benéficas para el ser humano se lleva a cabo en una primera fase en los laboratorios experimentales (con tecnología de punta), donde se busca controlar todas las variables. Por supuesto que son prospectivos y la población objeto son animales de laboratorio. En general, el principio de la investigación se basa en la comparación. Entonces el comparar efectos en quienes reciben (intervención) con grupos que no reciben es la clave, así como el máximo control sobre las demás variables (para estar seguros de que los efectos son por la intervención y no por otros elementos). Otra característica que exige la experimentación es el garantizar que las comparaciones se hagan en grupos donde no exista diferencia entre el intervenido y el no intervenido. Claro... de esa manera se busca garantizar que, si existe efecto en el intervenido, fue por la intervención y no por otra causa. Para ello, la selección aleatoria de los grupos es la clave.

ENsayo CLINICO

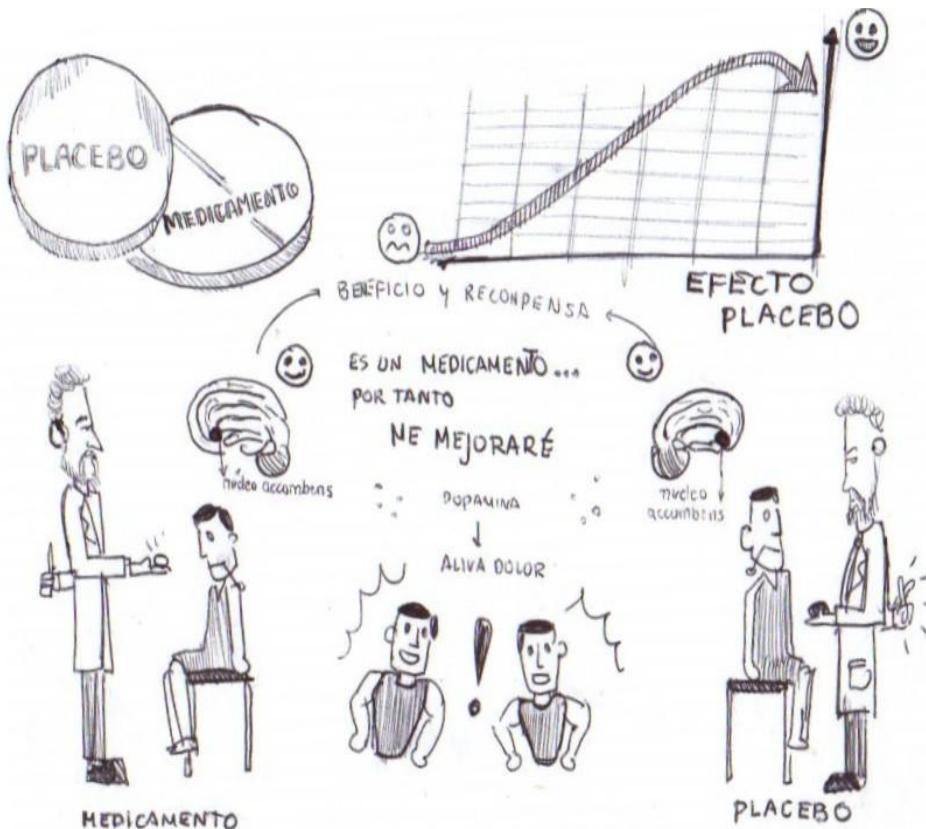
ENFERMOS

ALEATORIO

MEDIDA
DE MORBIMORTALIDAD



2.1.1.2. Ensayo clínico aleatorio: Sin duda, es la investigación estelar en el área de la salud. Constituye los diseños de la etapa avanzada en la investigación de medicamentos. En esencia lo que se busca es medir la efectividad de un medicamento (intervención) en un grupo de enfermos. Son prospectivos por excelencia y se mide a la población objeto antes de recibir el medicamento y en varios momentos después (siempre a la misma población). De hecho se mide el efecto buscado y otros aspectos (especialmente efectos indeseables) hacia el futuro (entre más dure las medidas en torno el tiempo, mayor la posibilidad de detectar efectos indeseables). Exigen garantizar la selección aleatoria de la población intervenida y la de control (grupo de comparación).



Sucede que cuando una persona recibe algo en lo que tiene esperanzas, puede momentáneamente mejorar. Si lo que recibe es el placebo y se recupera transitoriamente se está ante un **efecto placebo**, el cual ulteriormente desaparece. Esto se vio cuando aparecieron los antirretrovirales en pacientes con SIDA que no sabían que estaban recibiendo el placebo y creían estar recibiendo la droga milagrosa. Momentáneamente se mejoraron, así después, la mortalidad fuese la esperada.

Con el fin de evitar el sesgo del efecto placebo en los ensayos clínicos se propone que la población del experimento no sepa cuál es control y cuál será intervenida.

Lo anterior es lo que se conoce como **estudio CIEGO**. Si el que reparte la medicación (placebo e intervención) tampoco sabe cuál es cual, entonces se denomina **DOBLE CIEGO**. Y si el que analiza los resultados no sabe cuál es cual y solo compara a dos poblaciones sin saber cuál es la intervenida, se denomina **TRIPLE CIEGO**. Claro... los estudios con menos sesgos son los triples ciegos. En inglés estos estudios se denominan Randomized double blind trial.

- 2.1.1.3.** El **ensayo de campo** es casi igual al ensayo clínico pero se diferencia en que no interviene a población enferma sino sana. Pero el resto de características se debe cumplir. Por lo cual, son los diseños para medir la efectividad de las vacunas, por ejemplo, pues se parte de población sana. Una muestra fue la vacuna SALK contra la poliomielitis, donde participaron cerca de un millón de niños.
- 2.1.1.4.** En el **diseño cuasi-experimental** la diferencia lo establece la dificultad para trabajar con grupos de intervención o de control seleccionados aleatoriamente. Acá, definida por algún criterio la población sobre la cual se va hacer la intervención, se escoge una población lo más parecida (pero no de manera aleatoria) de control. Por ejemplo, una intervención educativa donde se toman jóvenes de un mismo curso de dos colegios. Los jóvenes se escogen por estar en el curso y no como resultado de un proceso aleatorio.
- 2.1.1.5.** En la **intervención comunitaria** no se escogen personas, y sí población. Por ejemplo, la población de un municipio que será intervenida con agua fluorizada y la de otro municipio cuya agua no será fluorizada (control). Claro... las poblaciones de cada municipio han de ser lo más semejante posible.

Es indudable que un estudio INTERVENCIONAL debe presentar en su descripción a la población antes de la intervención para que el lector pueda ver que si son comparables. Debe explicar muy bien cómo se repartió y explicar cómo se mantuvo el criterio de ocultamiento o enmascaramiento. La respuesta a estas presentaciones demuestra el grado de validez (o semejanza con la realidad).

2.1.2. Observacionales

La actividad fundamental de los diseños observacionales es DESCRIBIR y/o COMPARAR (pero por sobre todo NO intervenir). Su objetivo es describir (por ejemplo, una entidad) o buscar relaciones entre variables (y, así, proponer asociaciones causales). En torno al tiempo estos diseños son y pueden ser versátiles: tomar eventos pasados y a través de la documentación existente describirlos y compararlos. También pueden, usando instrumentos, identificar variables (de descripción y/o comparación) en el momento o hacerlo de manera prospectiva (medir varios momentos). Respecto a la población pueden trabajar sobre universos (población objeto total) o sobre muestras que pueden ser por conveniencia o aleatorias. Se clasifican en:

2.1.2.1. Descriptivos: Se denominan series de casos y la principal condición es seleccionar una cantidad de casos definidos por el investigador y describir la presencia o ausencia de una serie de variables. Pueden hacerse tomando historias clínicas, por ejemplo, estudios retrospectivos y a través de las mismas se describen las variables de interés. Pueden hacerse con muestra o con población total. Es muy usada para entender los comportamientos de las enfermedades. Sirven para construir la historia natural de las enfermedades.



Un ejemplo puede ser la descripción que se efectuó en la década de los 80 en torno al SIDA. La recolección de información, con su ulterior clasificación permitió ver las características de la población, de las manifestaciones, el pronóstico etc.



 Las series de caso permiten aprender muchísimo. Probablemente sean en medicina la mejor actividad descriptora. Los hospitales deberían tener abiertos sus archivos para poder describir las manifestaciones, tratamientos y demás de los pacientes que allí acuden. Así, de seguro surgirían interesantes planteamientos para hacer nuevas investigaciones en beneficio de los pacientes. En las investigaciones, la tabla de frecuencia es la encargada de presentar las descripciones. Por ello, saber construir una tabla descriptiva es un deber de los que anhelan ser investigadores.

En una serie de casos se pueden hacer numerosas comparaciones a través de las cuales se puede aprender. Por ejemplo, en una clínica, el tomar todos los casos de dolor abdominal de un año ya pasado (retrospectivo) y observar en qué casos se diagnosticó apendicitis, permite al compararlo con los que no, identificar la sensibilidad y la especificidad de las manifestaciones clínicas de la apendicitis. Es obvio que tal investigación puede contribuir a mejorar los procesos diagnósticos...

2.1.2.1.2. De corte (descriptivo exclusivamente). Se trata de describir unas variables, para lo cual se recogen los casos (población objeto) en el tiempo en que van llegando. Por ejemplo, describir las características del dolor torácico en un servicio de urgencias.

2.1.2.1.3. Los estudios analíticos aparte de describir, comparan poblaciones. Se denominan también diseños correlacionales, pues lo que buscan es ver si una variable cambia al cambiar la otra. Este hallazgo (la relación) en salud tiene enormes implicaciones, pues a partir de dicha correlación se pueden hacer los planteamientos (sospechas) de la causalidad. Pueden ser:

2.1.2.1.4. De corte: También denominado **cross sección** o estudio de **prevalencias**. En este diseño se aplica a una población un instrumento (variables) y se describen las frecuencias, las que se pueden comparar en esa población (los que tienen una variable con los que no la tienen). La condición esencial lo establece el hecho de que a toda la población objeto se la identifica una vez (se obtiene la información ahora). El que se trabaje con muestra o con población objeto total lo define el investigador de acuerdo a lo que busque, siendo muy obvio que la muestra representativa le permite a este diseño hacer inferencias. Los instrumentos más característicos son las encuestas que se hacen con tanta frecuencia (es el diseño más aplicado). Un ejemplo son los estudios de prevalencia como saber cuál es la frecuencia de fumadores de tal sitio o cuál es la frecuencia de factores de riesgo en una población.



Ciertamente estos estudios son particularmente simples y sencillos de hacer. Por ello, una gran cantidad de investigaciones son de este tipo. En los cursos cortos de investigación son los que más se realizan por las facilidades del diseño. Exigen acceso a la población, con la que solo se tiene contacto la vez que se le hace la encuesta. Clave es saber hacer y presentar el cuestionario...

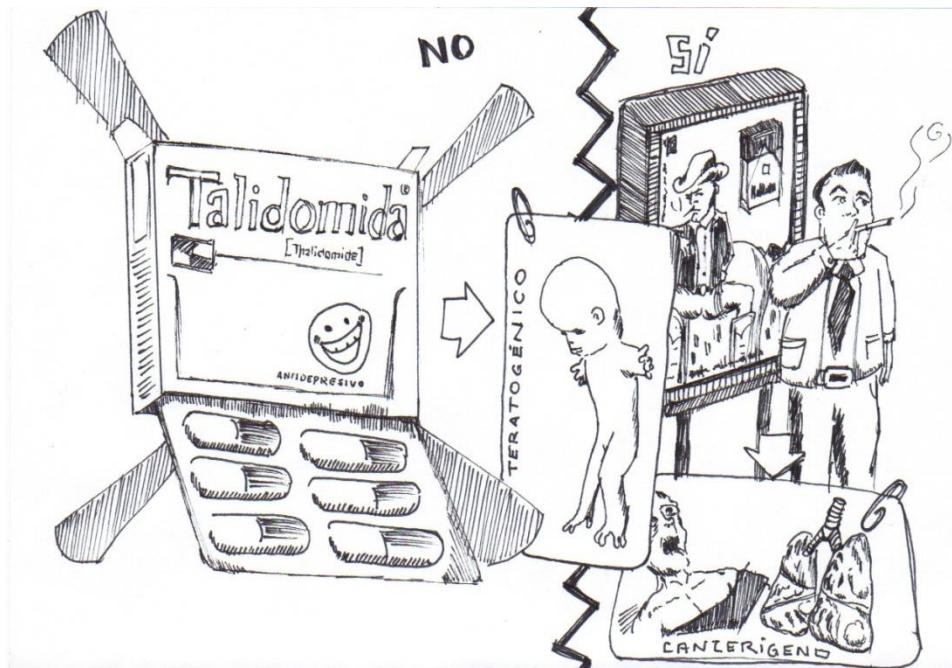
2.1.2.1.5. Estudio de casos y controles: En este diseño la característica esencial está en la población objeto. Se escogen dos poblaciones: Una identificada como CASOS y otra como CONTROL. El caso debe tener sin duda la condición que se estudia (habitualmente una enfermedad). En el control se debe tener la garantía de que no tiene la condición de caso; pero, por lo demás, ha de ser lo más parecido al caso. Con esa condición se aplica un instrumento

para comparar los casos con los controles buscando encontrar diferencias en las variables. En general, se hacen en torno al tiempo de manera transversal y se busca como mínimo comparar un caso por un control, dándole más potencia si se comparan dos o más controles por caso.



Recuerde que la potencia lo establece el número de participantes. Entre más participantes, es más probable encontrar más relaciones.

Un ejemplo fue el estudio de Doll y Hill acerca del cáncer del pulmón (casos) y del cigarrillo (exposición). Partiendo de pacientes con cáncer del pulmón se buscó población similar pero sin cáncer y se observó (a través de relaciones OR) si fumaban o no. Se encontró una relación expresada en OR de 16 (por cada persona con cáncer y que no fuma se encontró 16 con cáncer y que fuman).



Evidentemente cuando se encuentran relaciones tan fuertes, el paso siguiente es hacer estudios para ver la CAUSALIDAD. Por ejemplo, en la década de los 50 se encontró en Alemania un incremento del nacimiento de niños con malformaciones. Se efectuó un estudio de casos y controles donde se halló una fuerte relación con la ingesta de Talidomida (un antiemético recomendado en el embarazo). Ello dio pie a la suspensión inmediata del producto para el consumo humano. Ni siquiera se planteó hacer un estudio de cohorte. ¿Por qué se actuó de esta manera con esta droga y no con el cigarrillo?... iah, acá está el poder del dinero!

2.1.2.2.1. Estudios de cohorte: Así como el ensayo clínico es el estelar de los diseños experimentales, los estudios de cohorte son lo máximo de los estudios observacionales (no existe intervención). Son ideales para describir y comparar. En estos se selecciona una cohorte de personas a los que se les medirá a lo largo del tiempo (prospectivos). Son estudios que buscan causalidad y la condición esencial es garantizar que la población a la que se la va hacer el seguimiento, no tenga la condición que se quiera indagar (no hace falta que sea población aleatoria). Luego, y a lo largo del tiempo (mucho tiempo...) se hacen mediciones buscando ver si aparece la condición y, cuando lo hace, se revisa a qué se expuso, buscando relaciones entre lo causado y lo expuesto. Por ejemplo, luego de encontrar la relación entre el cáncer del pulmón y el cigarrillo, se planteó hacerle seguimiento a una población de la que se sabía no tenía cáncer. En dicha población, algunos empezaron a fumar y otros no y se pudo medir la comparación que con el tiempo determinó el evento de fumar y la aparición del cáncer. Pero no solo encontraron que los que fumaban les daba más cáncer. También pudieron medir que aparecía en los fumadores más casos de enfermedad coronaria, de enfermedad pulmonar obstructiva y cánceres en otros lugares. Claro... el seguimiento fue por más de 10 años.



Los estudios de cohortes se proyectan para hacerse mínimo durante el tiempo que se toma una enfermedad en aparecer. Por ejemplo, en los estudios de cáncer, la duración mínima será 10 años, pudiendo ser proyectados a más de 20 años. En la actualidad existen muchos estudios de cohorte indagando sobre las distintas exposiciones del hombre actual. Exposiciones a radiación de los celulares, daño de audición por el uso de audífonos (para música), exposiciones a químicos en alimentos etc. Claro, en el momento en que en las medidas se encuentren anormalidades estadísticamente significativas, tales estudios se publican para impedir que se continúe con las exposiciones dañinas.

2.1.2.2.2. Estudios ecológicos: Este diseño es como tal muy eficiente ya que busca comparar situaciones con datos ya recogidos. Además, se trabaja con grupos poblacionales. La comparación, y ahí está lo difícil de interpretar, se hace entre un momento (pasado) con otro momento (presente) lo que determina precisamente su debilidad; cada época es diferente de manera compleja y, por consiguiente, la comparación está sujeta a más variables que las previsibles. Un ejemplo puede ser el comparar la frecuencia de caries en 1980 en un municipio y ahora (2007). Claro, lo que se desea medir es el impacto que pudo tener la fluorización del agua en la frecuencia. Para ello se usan registros previos y se comparará con registros actuales. Dado lo distinto en las condiciones de la población que se compara, a veces los resultados no son ciertos. Eso se conoce como la **falacia ecológica**.

Uno de los retos de estos estudios (analíticos) es presentar la respuesta en las tablas de comparación. Obviamente, se originan de la tabla de contingencia, pero presentar las tablas de resultados (consolidadas) de comparación es todo un arte.

COMPORTAMIENTOS	HABER SIDO GOLPEADO EN EL ÚLTIMO AÑO						HABER GOLPEADO A ALGUIEN EN EL ÚLTIMO AÑO						LLEVAR UN ARMA					
	SI	NO	OR	IC 95%		SI	NO	OR	IC 95%		SI	NO	OR	IC 95%				
				LI	LS				LI	LS				LI	LS			
13 a 16 años	2 432 (35)	4 459 (65)	1,41	1,28	1,56	2 512 (37)	4 346 (63)	1,59	1,44	1,77	746 (11)	6 122 (89)	2,27	1,88	2,75			
10 a 12 años	923 (26)	2 639 (74)				907 (26)	2 633 (74)				165 (5)	3 423 (95)						
17 a 19 años						356 (37)	615 (63)	1,36	1,14	1,61	137 (14)	836 (86)	2,54	1,94	3,32			
10 a 12 años						907 (26)	2 633 (74)				165 (5)	3 423 (95)						
Masculino	2 062 (37)	3 574 (63)	1,65	1,51	1,8	2 405 (43)	3 194 (57)	2,66	2,44	2,92	752 (13)	4 864 (87)	3,06	2,63	3,57			
Femenino	1 639 (28)	4 224 (72)				1 399(24)	4 447 (76)				305 (5)	5 584 (93)						
Triste	446 (50)	446 (50)	2,15	1,81	2,55	414 (47)	473 (53)	1,93	1,62	2,3	141 (16)	744 (84)	1,96	1,51	2,53			
Feliz	938 (25)	2 752 (75)				1 030 (28)	2 635 (72)				273 (7)	3 425 (93)						
Disfunción Moderada	1 064 (41)	1 538 (59)	2,02	1,82	2,23	1 102 (42)	1 510 (58)	1,75	1,58	1,94	299 (11)	2 311 (89)	1,51	1,28	1,77			
Funcional	1 836 (25)	5 418 (75)				2 010 (28)	5 209 (72)				524 (7)	6 737 (93)						
Disfunción Severa	520 (59)	362 (41)	4,06	3,46	4,75	474 (54)	404 (46)	3,09	2,64	3,63	152 (17)	732 (83)	2,37	1,9	2,96			
Funcional	1 836 (25)	5 418 (75)				2 010 (28)	5 209 (72)				524 (7)	6 737 (93)						

Por ejemplo, en una investigación que buscó relaciones de edad, género, ánimo y estructura y función familiar con comportamientos violentos por parte de adolescentes escolares, se presentó esta tabla de resultados: en la primera columna están las variables dependientes (edad, género, ánimo y estructura familiar con funcionalidad familiar y las variables independientes son los comportamientos violentos (haber sido golpeado en último año, haber golpeado a alguien en último año y llevar un arma). Observe la tabla y se dará cuenta que los datos descritos permiten que el que la lea sepa de dónde viene cada resultado (y si quiere lo obtenga). Mírela con cuidado... ¿la sabe interpretar? Como ejemplo observe la relación más fuerte que es el OR de 4.06 que se encuentra al comparar la percepción funcional con la disfunción severa en la variable haber sido golpeado el último año. Se expresa así: Por cada adolescente que **SÍ** fue golpeado el último año y percibe funcionalmente a su familia, existen 4.06 adolescentes que **SÍ** fueron golpeados y perciben disfuncionalmente a su familia. Esta relación es significativa ya que los intervalos de confianza del 95% así lo demuestran.

2.1.3. Otros tipos de investigaciones cuantitativas

Una investigación cuantitativa que se está convirtiendo en el fuerte de la epidemiología es el **META ANÁLISIS**. En este se buscan estudios con diseños similares (todos los que se puedan...) y con el menor número de sesgos. Juntados todos los datos de cada una de las investigaciones seleccionadas, que son resultados de distintos investigadores y en distintas partes, se hace un **nuevo análisis** como si se tratase de una sola investigación. Lógicamente, tienen exigencias muy específicas, pero una vez se unen, la validez de los resultados se potencializa y aumenta la validez de lo encontrado. Por ejemplo, tomar las bases de datos de los ensayos clínicos de un medicamento y construir una nueva base de datos y hacer todo el proceso de análisis, da sin duda una aproximación mayor en torno a lo que se encuentre con la realidad.

Las **REVISIONES SISTEMATICAS** (que ya son en sí mismas investigaciones de un alto grado de evidencia) son exploraciones eminentemente bibliográficas en donde se busca todas las investigaciones publicadas en determinado tiempo acerca de un tema específico (problema) y se hace una comparación de sus resultados con el consiguiente análisis. Sus partes tienen criterios tales como explicar cómo se hizo la búsqueda, qué criterios de selección de investigaciones se tomaron, descripción de los estudios tomados, resultados y discusión de éstos, finalizando con las conclusiones de los autores (quienes hacen la revisión).



Existe en el mundo una tendencia a hacer de los investigadores y de sus investigaciones una gran comunidad científica libre de los conflictos de interés, donde todos podemos participar y unir nuestros esfuerzos, buscando que lo que se indague pueda unirse y, así, aproximarse más a la realidad. Por ejemplo, si se trata de estudiar la eficacia de un medicamento, la Fundación Cochrane propone que, sin importar el sitio del mundo, los investigadores se vinculen a los grupos de estudio y trabajen como si fuera una sola investigación. Para ello, todos los interesados en el estudio de tal medicamento deben cumplir con un protocolo (que busca disminuir al máximo los sesgos) y una vez efectuada la investigación, compartir los resultados con los demás miembros de esa comunidad científica. Así se podrán unir bases de datos y hacer meta análisis.

2.1.4. **Investigaciones administrativas:** Otros tipos de investigación y de tendencia cuantitativa, donde la clasificación (con los criterios de actividad, tiempo y población) no es tan sencilla y que se mencionan más con carácter informativo, proponiendo que el interesado las busque, son las investigaciones administrativas. Estas, en general combinan actividades, tiempos y poblaciones de una forma no fácil de clasificar en los diseños epidemiológicos. Una aproximación a su entendimiento sería:

La investigación de procesos: término que abarcaría a todos esas investigaciones donde se implementan procesos obligatorios. Por ejemplo, los requisitos de habilitación para una IPS o la implementación de las normas de ISO en una IPS. Para ejecutarse se debe: describir condiciones del momento (diagnóstico previo), evaluación de recursos, diseñar procesos de implementación para cumplir las normas, etc. Estas se miden por indicadores.



Como tales, son muy integrales, queriendo decir con ello que se acercan mucho a la realidad: de hecho, y por ejemplo, en una investigación para implementar la habilitación de una IPS, si no se observa todo, la realidad es que la cierran!

Las investigaciones de costos. Son aquellas que, siguiendo criterios cuantitativos, se toman los recursos como medida (en general valores de dinero) y, a través de ellos, se establecen las descripciones y las comparaciones. Son muy útiles para medir eficiencia.



Su mejor ejemplo son los estudios de costos por actividad en donde, y a través de ello, se pueden comparar la eficiencia de las EPS, de las IPS, de las actividades de las empresas etc.

2.2. Diseños cualitativos

Constituyen la otra gran tendencia en el mundo de las investigaciones. La característica más importante de los diseños cualitativos es la flexibilidad de los mismos. De hecho en la medida que el estudio avanza, el diseño se construye. De ahí el término de diseño emergente. Su razón esencial es **entender** los fenómenos que estudia. Por ende, el análisis de la información ocurre junto con cada paso, de los que a su vez, van surgiendo los subsiguientes.



Los diseños cualitativos sostienen que la realidad no puede encajarse en los modelos cuantitativos. Refieren que es cambiante y no rígida, por lo que sus diseños procuran ser tan flexibles como la realidad se puede presentar.

Una clasificación más con carácter informativo, puede ser la siguiente:

1. Investigaciones que buscan entender **CULTURA** de la población:

- Etnografía. Busca entender de manera holística la cultura.
- Etnociencia. Busca entender significados que moldean el comportamiento.

2. Investigaciones que buscan entender **experiencias**:

- Fenomenología. Entender experiencias de los individuos.
- Hermenéutica. Entender experiencias bajo el contexto sociocultural.

3. Investigaciones que buscan entender **conductas y acontecimientos**:

- Etiología. Entender conductas o comportamientos.
- Psicología ecológica. Entender conducta influida por el ambiente.

4. Investigación que busca entender **ambientes sociales**:

- Teoría fundamentada. Busca entender los procesos psicológicos y estructuras sociales del entorno social.
- Etno-metodología. Entender cómo o por qué se comportan los individuos de una u otra forma socialmente.
- Interacción simbólica. Entender cómo se interpretan las interacciones sociales.

5. Investigación que busca entender la **comunicación humana**:

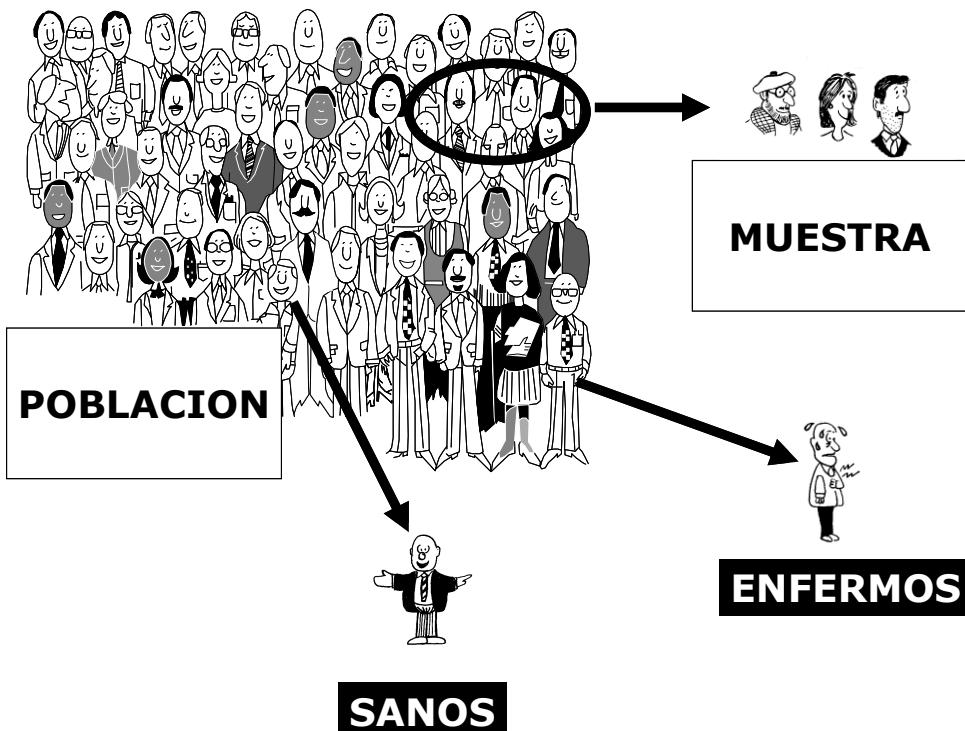
- Análisis del discurso. Entender las formas y reglas de la comunicación.

6. Investigación **histórica**. Recolección de datos e interpretación de los mismos.

2.3. Investigación multi-método: Combina los diseños cuantitativos y los cualitativos. Por ejemplo, hacer un estudio de prevalencia acerca de sexualidad en un colegio (diseño cuantitativo), procesar los resultados y después reunir a un grupo de adolescentes para, a través de una metodología cualitativa, tratar de entender las respuestas encontradas.

2.4. Paso 8. Selección de la población

Los estudios pueden hacerse sobre población total (o de referencia) o sobre muestra (de la población de referencia). El concepto de **población total** se establece cuando se puede definir con claridad el universo (total de población) y el investigador define que puede acceder a él. Por ejemplo, toda la población adolescente de un colegio. Claro, una población total de un colegio no permite hacer inferencias de lo encontrado o, lo que es lo mismo, generalizar lo encontrado a la totalidad de la población adolescente escolarizada de la ciudad. **La inferencia** en general es una de las razones de los estudios cuantitativos. Solo existe una manera de alcanzarla y es trabajar con una **muestra de población representativa**. La **representatividad** se expresa como la capacidad que se tiene de inferir de la muestra al universo (o población total de similares características). La investigación cualitativa no usa en general muestras representativas pues lo que buscan es entender y no inferir y, para ello, cualquier población sirve.



Una población es el agregado total de casos que cumplen con una serie determinada de criterios. Población accesible son los casos de población que cumplen con criterios y que son accesibles al investigador. Población blanco es la población sobre la que se quiere hacer inferencias.

2.4.1. Acceso a la población: El identificar claramente cómo se puede llegar a la población es particularmente importante cuando se va hacer una investigación. En este punto, es clave el seleccionar unos criterios que debe cumplir la población para ser escogida y asegurarse de cumplir con principios éticos sobre dicha población.

2.4.1.1. Criterios de inclusión: También denominados CRITERIOS DE ELEGIBILIDAD, son las condiciones que debe tener la población para poder ser seleccionada. Buscan disminuir al máximo la posibilidad de escoger población sesgada. Indudablemente que surgen de las características del estudio en cuestión. Edad, estrato, condición de salud, son ejemplos de criterios de selección.



Por ejemplo, en un estudio de un medicamento para el asma, donde se debe seleccionar la población sobre la cual probar el medicamento, ¿qué sucedería si se efectuase en población sin asma? Obviamente puede suceder que el medicamento muestre resultados excelentes... pero, ¿se puede afirmar que sucedería lo mismo en población con asma? Como se observa, esa población sin asma se convierte en un ERROR DE SELECCIÓN. Otro ejemplo puede ser al estudiar prevalencias de comportamientos de sexualidad en adolescentes... comparar jóvenes de 18 años con los de 15 es bien diferente... ¿cierto? Resulta evidente que las conclusiones de un grupo son distintas que las del otro. Entonces el estudio debe especificar cuál será la edad en cuestión.

Seleccionar apropiadamente la población objeto de una investigación (expresar las condiciones que debe cumplir la población escogida) resulta clave en los diseños cuantitativos ya que ellos buscan inferir al total de la población de similares características. De hecho, una selección inadecuada da pie a lo que se conoce como **SESGO DE SELECCIÓN**, que es uno de los tres errores que la epidemiología estudia con mayor cuidado. La cuidadosa selección de criterios de participación o de exclusión sobre la población objeto de una investigación evita que se incurra en este frecuente sesgo.

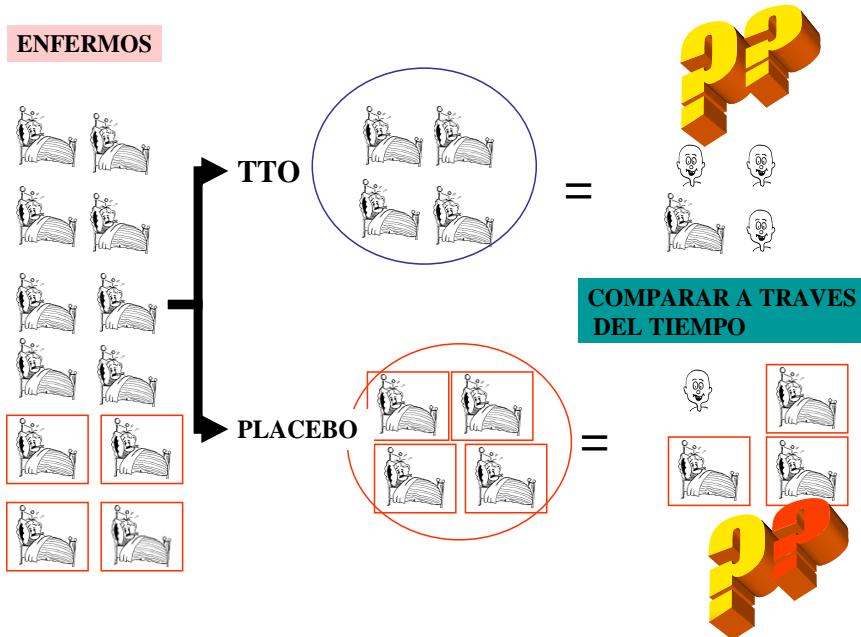
Dado que en muchos estudios se comparan poblaciones, el SELECCIONARLAS para que sean comparables resulta particularmente importante. Por ejemplo, Doll y Hill para poder saber si existía una relación entre el tabaco y el cáncer del pulmón hicieron lo siguiente: Buscaron CASOS de personas con cáncer del pulmón. Luego identificaron características claves de cada uno de estos casos. Posteriormente, y para comparar, buscaron personas de las cuales con un alto grado de seguridad se podía afirmar que no tenían cáncer pulmonar y que en esas características claves fuesen similares a los casos. Se les denominó CONTROLES. Y con esto compararon si el cigarrillo se relacionaba o no con el cáncer. Los sesgos de selección en los que ellos podían haber incurrido (y que, para no hacerlo, estuvieron muy pendientes) serían:

1. Que seleccionasen como CASOS a personas que no tenían cáncer del pulmón.
2. Que seleccionasen como CONTROLES a personas que tenían cáncer.
3. Que las características claves de los casos y de los controles no fuesen similares. Por ejemplo, no se escogió el trabajo como característica clave para seleccionar los controles y resultó que muchos de los casos trabajaban con asbestos y no se buscó en los controles esta característica.

Estas tres condiciones expuestas hacen muy probable que las conclusiones de un estudio que no las haya tenido en cuenta no sean ciertas. Por ello se denominan sesgos o errores que, de estar presentes, invalidan los resultados; es decir, que no puedo asegurar que la realidad sea así y dado que se refiere a la población, se han denominado sesgos de selección (repitiéndolo una vez más) de la población escogida del estudio.

Otro error que se ha de buscar es el expuesto en este otro caso siendo un sesgo (error) más frecuente de lo que parece (tristemente a veces malintencionado). Por ejemplo, si se va a probar un medicamento en una enfermedad, resulta esencial proporcionárselo a los enfermos, pero para saber si sirve o no, se debe tener otro grupo de enfermos para

comparar. Es claro que, en lo posible, no deben existir diferencias entre un grupo de enfermos y otro. Solo así se convierten los dos grupos en comparables.



El dibujo expresa uno de los tipos de sesgo de selección: imagínese que se tiene una población de enfermos de SIDA y como resulta obvio, en esa población existen unos más enfermos (en el dibujo, los que están encerrados por un cuadrado) que los otros (por ejemplo, tienen notoriamente más disminuida la población de linfocitos T, así como más bajo el peso promedio). Entonces, se desea probar la eficacia de un nuevo medicamento (llamado TTO). Resulta lógico que para probar la utilidad se debe dividir a los enfermos en dos grupos y a uno suministrar TTO y al otro no (solo así se puede identificar si sirve o no). Pero se escoge a los menos enfermos en el grupo del TTO (a los que se les dará el medicamento) y a los más enfermos en el grupo que se le dará el placebo. Al cabo del tiempo se mide quiénes se curaron más: si los del grupo TTO o los del grupo Placebo. En este ejemplo, resultó que fueron los del grupo TTO (3 de 4 se mejoraron). Ahora la pregunta es: ¿se puede estar seguro de que este resultado es válido? Ello significaría que se podría asegurar que los que reciben TTO se curarán más que los que reciben placebo. Si bien el resultado pone en evidencia que en TTO se curaron 3 y en placebo solo 1, no se puede estar seguro de que este resultado se debió al TTO. ¿Por qué? Pues porque los que recibieron TTO estaban menos enfermos que los otros (por ende, NO SON COMPARABLES). Algun observador para demostrar que no son comparables puede decir lo siguiente: 'Mire, si uno de los que recibió placebo se curó, ¿cómo descartar que los tres que recibieron TTO no se iban a curar así no recibieran nada?'

Para evitar este sesgo en la selección de la población cuando se vaya a dividir a los enfermos, se debe repartir de manera equitativa a cada grupo (grupo de intervención y grupo de control). Ese equilibrio se alcanza a través de una **repartición aleatoria**. Así, casi con seguridad en cada grupo (brazo) de repartición quedarán una misma proporción de enfermos más y menos graves.

2.4.2. Cumplimiento de principios éticos

No existe duda acerca del inmenso valor que se ha de dar a la ÉTICA en la investigación. La investigación debe siempre ofrecer beneficio a la humanidad y sobre todo a la vida. Para ello se han construido una serie de códigos de ética que todo investigador debe conocer. En todo caso y para cualquier investigación se deben cumplir como mínimo los siguientes principios:

2.4.2.1. PRINCIPIO DE BENEFICIENCIA: La máxima es que nunca se puede dañar a la población objeto. Por ende, se debe dar la garantía de que no existen riesgos. Además, que no se explotará a la población objeto (de ahí la necesidad de la confidencialidad) y, claro que recibirá BENEFICIOS con base al estudio (mínimo conociendo los resultados). Existen ocasiones en donde el investigador ha de tener en cuenta la razón riesgos-beneficios y ahí, por sobre todo, la población objeto ha de conocerla.

2.4.2.2. PRINCIPIO DE RESPETO A LA DIGNIDAD HUMANA. Se refiere al derecho de la población objeto de decidir si participa o no del estudio, además de conocer todo referente a la información sobre el estudio (qué se piensa hacer con la información). Una vez más, la confidencialidad y la exigencia de reserva en los resultados es un derecho de la población objeto.

2.4.2.3. PRINCIPIO DE JUSTICIA: Todo participante en una investigación debe ser tratado justamente. No puede ser discriminado, puede retirarse cuando lo desee, puede tener acceso a cualquier información respecto a la investigación y un trato amable y respetuoso. Se debe restringir la identificación de la población objeto.

2.4.2.4. CONSENTIMIENTO INFORMADO. Todo participante debe ser informado suficientemente acerca de la investigación. Debe entender los riesgos y los beneficios, el uso que se le dará al estudio, cómo se garantizará la confidencialidad y primordialmente dar su consentimiento. Claro, se ha de garantizar el ENTENDIMIENTO por parte de la población objeto.

2.4.2.5. CONSIDERACIONES SUJETOS VULNERABLES: Se ha de tener un particular cuidado con población de niños, jóvenes y personas con discapacidad. Se debe solicitar autorización en la participación de responsables, explicándole a ambos (población objeto y responsables) necesitándose el consentimiento de ambas partes.

2.4.2.6. COMITÉS DE ÉTICA. Es muy conveniente que en todas las entidades que trabajan en el área de salud (prestando o educando) existan comités de ética cuyo papel será proteger a la población objeto cuando surja el interés de hacer una investigación. Los integrantes deben estar al tanto de las leyes que regulan la investigación y deberán verificar el cumplimiento de los principios éticos en el protocolo de la investigación.



En nombre de la ciencia sí que se han cometido atropellos sobre la población. Existen documentos que dejan en evidencia cómo se experimentó sobre población incluso exponiendo a la población a sustancias sin su consentimiento. Algunos prisioneros de guerra en la II Guerra Mundial dan fe de los atropellos a que fueron sometidos en nombre de la ciencia. Por otra parte, la medicina aplicada a los pacientes es muy imperfecta. Con la mejor intención, en muchas ocasiones, las intervenciones (medicación) pueden provocar efectos indeseables incluso más daño que beneficio. De ahí que es esencial que los médicos reconozcamos esta limitación y les expliquemos a los pacientes, de manera sencilla, los riesgos a que se someten con este o aquel tratamiento para que sea el paciente en últimas el que decida si corre o no el riesgo.

2.5. Muestras (de población)

Una muestra es la selección de una porción de la población. Se puede trabajar con muestras que representen el total de población o no. Representar a la población quiere decir que esa porción presenta las mismas características que el total de la población de donde se obtuvo. Claro... con un pequeño margen de error (o diferencia). La importancia está en que al trabajar con una muestra **representativa**, los resultados de la investigación se pueden generalizar al total de la población de donde salió la muestra, evento conocido como **INFERENCIA**.

Lo anterior constituye una de las grandes diferencias entre la investigación cuantitativa y la cualitativa. Estos últimos buscando la comprensión sacrifican las inferencias, pues pretenden entender antes que generalizar lo que estudian. Los diseños cuantitativos buscan la inferencia a través de las muestras, por ende, son muy exigentes en la búsqueda de la representatividad.

Para que no exista confusión en la selección de la muestra, es conveniente siempre definir muy bien ciertos términos. Por ejemplo, **población** (total de casos que cumplen con el principal criterio) y de ella identificar la **accesible** (que es la población con los criterios y que además, es accesible) y la **blanco** (que es la población sobre la cual se quiere hacer una generalización).

2.5.1. Muestras representativas. En torno a la representatividad de una muestra se ha de decir que alcanzarla es todo un esfuerzo donde existen incluso diseños de muestreo (con exigencias específicas). En general, y si se busca representatividad, se debe escoger un **MUESTREO PROBABILÍSTICO** donde cualquiera de la población total tiene la oportunidad de hacer parte de la muestra. Para ello existe todo un proceso metodológico. Un ejemplo puede ser el procedimiento que se sigue en los censos: teniendo en cuenta el total de población, y partiendo que no se puede censa toda, se selecciona un número de población con base a una fórmula. Por supuesto que esta muestra ya tiene un margen de error. Entre más aproximada esté la muestra del total de la población menor será el error. Escogido el número de personas a censar, el siguiente paso es permitir que cualquiera del total de población sean los elegidos. Suponiendo que se sabe dónde viven, una forma es introducir todos los nombres en una bolsa y sacar la cantidad determinada previamente por la fórmula. Así se puede asegurar que todas las personas que forman la población, tuvieron la misma oportunidad.



Claro... lo anterior no siempre es factible, así que se han desarrollado una serie de alternativas.

El muestreo probabilístico puede usar distintos criterios. **La estratificación** se refiere a criterios excluyentes que permiten clasificar a la muestra. Por ejemplo, la edad, donde se clasifica por grupos de 0 a 1 año y de 2 a 5 años. Ambos criterios se excluyen. En este caso si se conoce la repartición por edad del total de la población, la muestra se seleccionará en proporción a las edades del total de población (estratificación por edad).

Ejemplos de muestreo probabilístico (donde cualquier miembro de la población total puede ser seleccionado) son:

- **Muestreo aleatorio simple.** Donde se escoge la muestra a través de un sistema de lotería o de números probabilísticos.
- **Muestreo aleatorio estratificado.** En donde primero se estratifica y luego se selecciona a través de un sistema aleatorio.
- **Muestreo por conglomerados** o de etapas múltiples: Por ejemplo, se quiere una muestra de estudiantes de medicina, entonces se escoge aleatoriamente a las facultades. De las seleccionadas se escoge aleatoriamente cuántos de cada una. Luego de cada una se hace la selección aleatoria de los estudiantes.
- **Muestreo sistemático:** Se usa, por ejemplo, cuando se decide hacer una encuesta de calidad de servicios de salud en distintas entidades y se toma como criterio que de cada 10 entidades, el décimo ingreso será encuestado. Claro... implica colocar los nombres de todas las entidades que se podrían encuestar (que representan la población total) en una bolsa, para así escoger aquellas que salgan luego de contar diez y desplazarse hasta las seleccionadas para encuestar a la persona que ingresa de décimo en tal entidad. Cómo se observa se seleccionó un método que se aplica de manera total (sistemático).

Lo importante a tener de presente es que la muestra representativa debe asegurar en lo posible que el escogido fue producto del azar.

2.5.2. Muestras por conveniencia: Cuando se decide trabajar con una muestra no probabilística, se denomina **MUESTRA POR CONVENIENCIA**. Existen muchas formas de hacerlo, pero lo importante es tener en cuenta que **no se pueden hacer inferencias** de lo hallado.

Hay una gran variedad en el muestreo por conveniencia: **por cuotas**, donde se asigna a la muestra proporciones similares a la población de referencia o blanco. El **intencional**, en donde es el investigador con base en su conocimiento quien selecciona la muestra o el **accidental** cuando se escoge una muestra porque está ahí (accidentalmente).

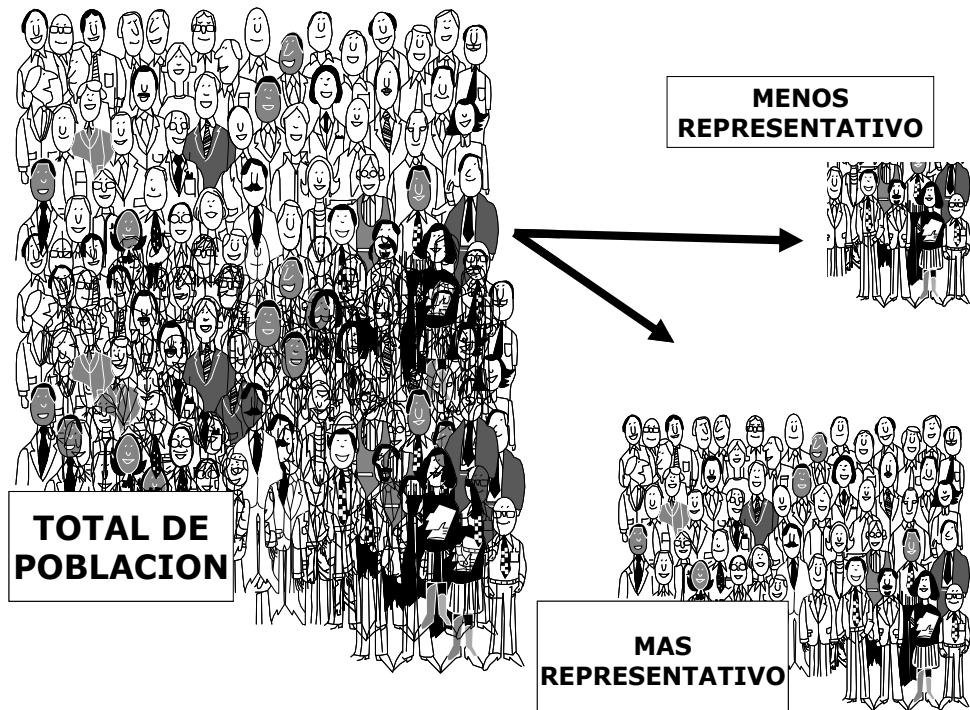
2.5.3. Tamaño de muestra: El cuanto el tamaño de la muestra vale la pena tener en cuenta que entre más grande una muestra, más cerca está de representar el universo. Como se trata de una muestra, existirá algún porcentaje de error. Existen fórmulas estadísticas para calcular el tamaño de la muestra y el error derivado de ello.

Entre los criterios para escoger el tamaño de la muestra está la homogeneidad de la población (qué tan parecidos son entre sí), la magnitud del efecto (referido a la fuerza esperada de las relaciones) o la posibilidad de pérdida de población en estudios prospectivos. Además, y de acuerdo al número de variables, se espera que exista más muestra entre más variables se pregunten.

Claro, entre más pequeña sea la muestra, mayor será el margen de error. Para seleccionar bien el tamaño se deben seguir estos pasos:

1. Identificar muy bien cuál va a ser la variable que se va a muestrear
2. Cuál es la unidad de muestreo
3. Conocer el universo.

Una fórmula para definir el tamaño (n) es usar el término Z (nivel de confianza) elevado al cuadrado (que significa con dos desviaciones estándar y que elevadas al cuadrado serían 4) multiplicado por p (que es 0,5) multiplicado por q (que es 0,5), todo ello dividido por error al cuadrado (error o precisión de estimación que puede ser 2 o 5 o 10 por ciento), lo que sería, $n = (Z^2 * p * q) / E^2$.



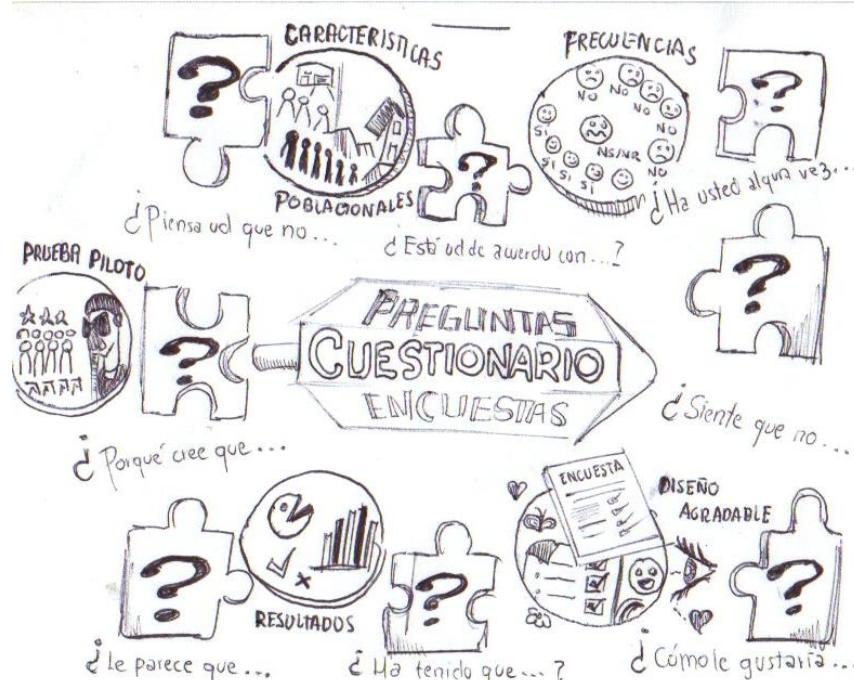
2.6. Paso 9. Selección de las variables (instrumentos a emplear)

Para ser contestada, la pregunta de investigación necesita ser desglosada a través de variables.



Recuerde que los seres humanos para entender fraccionamos la realidad y construimos conceptos. Las **categorías** son fraccionamientos de lo que vemos y que siguen algún criterio. En las investigaciones categorizamos con el fin de descomponer variables que puedan ser analizadas. Por ejemplo, si se requiere indagar acerca del nivel socioeconómico de una muestra poblacional, entonces se busca cuál puede ser la variable que ofrece el mejor dato: los estratos (índicador) parecen ser en este caso y, por ende, la categoría socioeconómica se fracciona como estratos 1, 2 ,3 ,4 ,5 y 6. Así, si se pregunta: '¿Usted a qué estrato pertenece?', la respuesta será el dato.

De las respuestas de las variables se obtienen los datos que serán tabulados. Una vez hecho esto, el investigador deberá buscar la respuesta a la pregunta de investigación analizando los datos de las variables empleadas. Para entender a cabalidad este aspecto piense en esta pregunta de investigación: *¿Será que los lactantes alimentados con leche materna tienen menos episodios infecciosos que los que no reciben?* La respuesta solo puede ser, o que presentan diferencia (en este caso menos episodio de infecciones) o, que no la presentan (no existe diferencia). Pero para poder llegar a una conclusión, se ha de tener en cuenta la respuesta (datos) a una cantidad de preguntas (variables) y es de ahí de donde surgirá la información que se busca. Todas las preguntas se recopilan en un cuestionario, siendo una definición entendible de cuestionario, la recopilación de variables que se emplean para contestar la pregunta de investigación.



Muy seguramente se debe preguntar edad en meses (para así poder comparar por edades a las dos poblaciones, la que recibe lactancia materna y la que no). Indagar estratos socioeconómicos, pues claramente pueden influir en una u otra respuesta. Se deberá preguntar por vacunas, pues ellas pueden ser determinantes en la aparición o no de algunas enfermedades infecciosas; así mismo, preguntar por episodios de diarrea y de infección respiratoria, por ser las infecciones más frecuentes de los lactantes, etc. Lo que se quiere ejemplificar es que para contestar una pregunta tan sencilla, se deben usar una importante cantidad de variables. Sea este el momento de mencionar que es apropiado investigar quiénes se han hecho esta pregunta antes y estudiar los instrumentos (cuestionarios de variables) que existen, ya que puede ser más eficiente usar estos que diseñar uno nuevo (diseñar lo que ya otros diseñaron es una pérdida de tiempo).

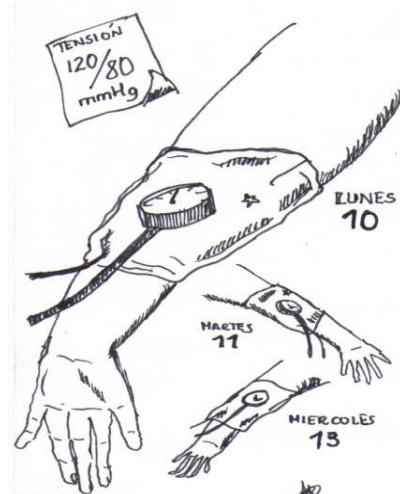
Por lo anterior, se insiste en tener en cuenta que una de las condiciones de una buena pregunta de investigación es que se pueda transformar en variables que la contesten adecuadamente. De hecho, entre menos variables se necesiten... mejor, pues será más fácil interpretar.



Las variables, por supuesto, reciben influencia del diseño. Volviendo al ejemplo en donde la pregunta es: *¿Será que los lactantes alimentados con leche materna tienen menos episodios infecciosos que los que no reciben?* Si se escogen dos cohortes de lactantes, unos alimentados con seno y otros con leches artificiales, la variable puede ser el número de episodios de diarrea o infecciones respiratorias durante el tiempo que dure el estudio y los resultados podrán ser concluyentes en su relación con la realidad. No será así, si se escoge comparar historias (retrospectivo) de niños que recibieron seno con los que no y ver episodios infecciosos en un período. Acá las dificultades de las historias (que esté o no la información) y otras circunstancias, hacen que la posibilidad de dejar pasar elementos importantes sean mayor que en el primer caso.

Una variable apropiada debe cumplir con una serie de requisitos por demás lógicos. Sin duda, el principal es que la respuesta que se obtenga permita ser MEDIDA.

Una medida es una comparación de alguna condición que se hace a la población objeto del estudio y que permite ser entendida. Dicha comparación debe, en lo posible, hacerse con unos instrumentos y, claro, cada vez que se mida a otros sujetos de la investigación, usar ese mismo instrumento, usando en lo posible las mismas condiciones de aplicación del instrumento.



Si se trata de un estudio donde la medida es la tensión arterial entonces es obvio que:

1. El tensiómetro que se usa ha de ser el mismo para ambos grupos de control.
2. La forma y el momento en que se toma la tensión arterial a cada grupo debe ser similar.

Si se trata de una encuesta, entonces el cuestionario es el instrumento.

Dado que al comparar se pueden obtener diferencias, se ha procurado usar números que permitan ver las diferencias. Sin embargo, se pueden usar otros símbolos para ver las diferencias. Por ejemplo, si la variable es la medida de la talla de la población objeto, el dato obtenido en centímetros en cada sujeto objeto de la investigación permite ver y entender de manera muy sencilla las comparaciones entre personas. También existen otras escalas de comparación, como la de Likert que, usando una construcción de las variables (preguntas), las respuestas que se pueden obtener (reactivos) son: excelentes, buenas, normales, malas, pésimas. El ideal en todo caso (y, por supuesto, base necesaria), es que la variable, el instrumento de medida y el dato obtenido, reflejen la realidad. Volviendo al ejemplo de la talla de la población objeto, un tallímetro (instrumento) y la respuesta que se obtendrá, talla en centímetros (dato), sin duda reflejan y comparan el tamaño de la población objeto del estudio. Las encuestas en muchas investigaciones constituyen el instrumento. Asegurarse que el entrevistado entienda las preguntas y que las opciones de respuesta reflejen la realidad son algunas de las características que deben cumplir.

La utilidad inmensa de las medidas es que disminuyen las especulaciones o conjeturas. Pueden hacer del 'yo creo' una certeza, lo que facilita la posibilidad de ponernos de acuerdo a través de la verificación. Además, nos ayudan a entender mejor. Si se dice que fulano es muy alto, cada persona puede calcularle una medida en centímetros distinta. Pero si se dice que mide 188 cm, todos podemos saber con exactitud qué tan alto es. Ese saber se llama **precisión** y esta característica puede ser de mucha utilidad para el avance del conocimiento. ¡Ah...! de una vez vale la pena mencionar que los números son un lenguaje más preciso que las palabras. Decir que el paciente tiene fiebre, da información, pero decir que tiene 38,5 grados centígrados, permite entender con mayor precisión.

Existe una tendencia en la medicina de construir **escalas** (preguntas) que nos acerquen al diagnóstico. Por ejemplo, las escalas para la depresión son preguntas que tienen un valor numérico y que al terminar de contestarlas todas y contar los puntos de las respuestas permiten aproximarnos al diagnóstico de la depresión. Para ello se hace necesario **validar** la escala y ello implica:

1. Diseñar la escala
2. Probarla en una población que tenga y no tenga la enfermedad.
3. Luego se debe confrontar la población y la escala con la prueba reina que es a través de la cual se asegura que tiene o no tiene la enfermedad
4. De esa manera se puede ver la sensibilidad y especificidad de la escala y con ello se obtendrán los **valores de verosimilitud** y así la obtención de las probabilidades de estar o no enfermos.



Diseñar variables donde sea fácil medir no es tan sencillo. Por ejemplo, el test de APGAR de la funcionalidad familiar, pretende a través de la medida otorgada (0 = casi nunca, 1= a veces y 2= casi siempre) identificar la percepción de disfunción familiar. Lo importante es saber si las medidas se corresponden a la realidad (validación). De ahí que cuando se construye un instrumento (llámese test) para medir algo, antes de aplicarlo debe pasar por el proceso de validación o, lo que es lo mismo, demostrar que ciertamente mide la realidad. Ello habitualmente se hace confrontando el instrumento con EL INSTRUMENTO REINA, aquel que nos ofrece la mejor medición de la realidad en una población. Por ejemplo, la mejor manera de detectar trastornos de conducta alimentaria (TCA) es la entrevista psiquiátrica. Si construimos un cuestionario para detectar casos, se debe hacer un estudio tomando una población donde los TCA estén presentes (y, claro, otra donde no estén no presentes). Se les aplicará el cuestionario, pero además la entrevista psiquiátrica; luego se confrontaran los resultados verificando qué tan concordante es el cuestionario con la entrevista psiquiátrica. Así se hace un proceso de validación de instrumento.

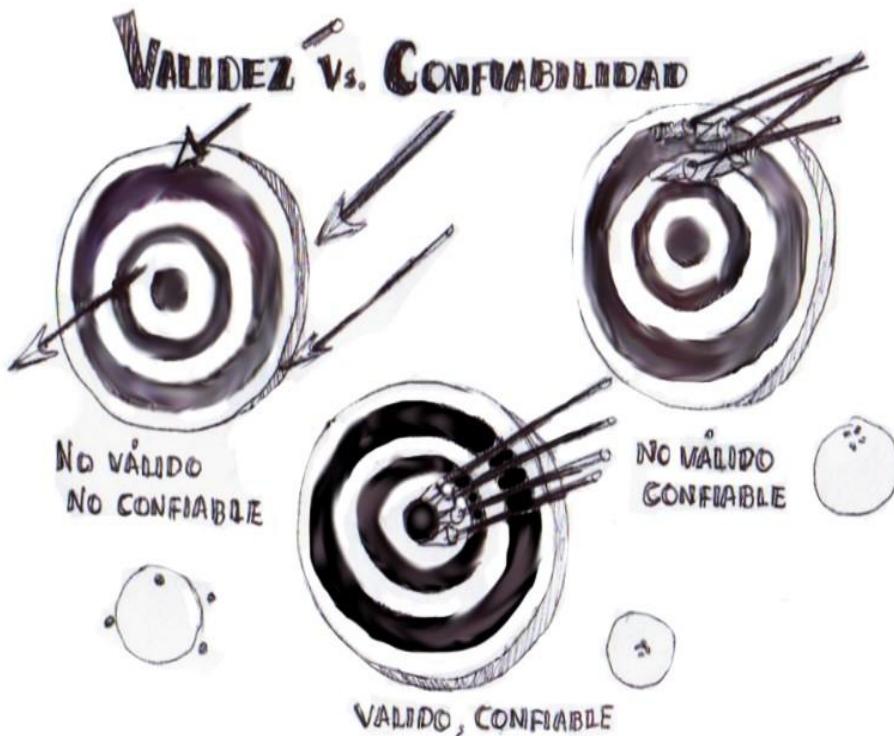
2.7. Sesgos: Para medir se usan instrumentos. La diferencia que existe entre la realidad y el resultado de la medida se denomina **el ERROR**. De ahí surge una fórmula que dice que la medida obtenida es igual a la medida real +/- el error. El análisis del error en las medidas es una tarea en la que se debe insistir porque solo al reconocerlo, se puede combatir. Los errores en general pueden suceder en el proceso de manejo de la **información**, cuando se **selecciona** la población o ante la interpretación de algún resultado (**confusión**). Algunas variedades de errores de información son:

- **Errores derivados de la obtención de la información:** Por ejemplo, si se preguntan asuntos íntimos delante de otros, puede que la persona no conteste con sinceridad. Otro caso es el error ocasionado por cansancio (formularios muy largos) o porque no se entiende, entonces el encuestado llena cualquier respuesta (y no la verdad). Otros ejemplos son los derivados de respuestas inducidas, donde el entrevistado contesta por alguna presión determinada respuesta (como preguntarle a un estudiante al final de la clase si le gustó. La presencia del profesor hace difícil que el estudiante pueda responder con libertad).
- **Errores por el uso del instrumento:** Por ejemplo, usar un tensiómetro descalibrado (información) o una pesa mal calibrada.
- **Errores al tabular** (información). Colocar 'sí' cuando debía ser 'no'.

Sesgos de selección: es cuando se escoge una población que no permite hacer comparaciones (con base a la realidad). Por ejemplo estudiando la relación de cáncer con el café, tomar una población con cáncer que tome café y para comparar tomar una sin cáncer que no tome café. Lo esperado es que la única diferencia entre las dos poblaciones sea tener o no el cáncer. Este fue el caso de un estudio efectuado por el Doctor Mac Mahon, quien para comparar (estudio de casos y controles) pacientes con cáncer de páncreas, seleccionó pacientes con problemas gastroduodenales (a los que se les había prohibido el café) y no se percató de esta situación. Cuando en los resultados se encontró una fuerte relación entre el café y el cáncer, asumiendo una actitud de duda crítica, buscó y encontró el error de selección que al ser corregido (tomando otra población sin esta limitación) hizo que desapareciera la relación (da lo mismo tomar café que no tomar café en el cáncer de páncreas).

Un ejemplo de sesgo de confusión quedó expuesto con la ictericia por salvarsán cuando en el pasado se creyó que era por el salvarsán (medicamento mercurial utilizado para tratar la sífilis) que las personas

desarrollaban un cuadro icterico. Se atribuyó erróneamente al resultado ese concepto. Hoy se sabe que era (y es) el virus de la hepatitis B el responsable.



Si la realidad es igual a dar en el centro, el dibujo expresa gráficamente el sentido de válido y confiable. Las flechas expresan una variable. El centro de la diana, la realidad. El primer dibujo expresa una variable que no da en el centro y que al ser usada nuevamente da en distintos sitios. El segundo una variable que no da en el centro pero al volverse a usar siempre da en el mismo punto. El tercer dibujo muestra una variable que da en el blanco (realidad) y al volverse a usar repite el sitio... Es la variable ideal.

2.7.1. Atributos de una medida: Una buena medida debe tener los siguientes atributos:

- **CONFIABILIDAD:** Es la congruencia que tiene la variable con lo que mide. Si se desea saber el peso de una persona y se pesa en una báscula y se obtiene un peso de 75 kilos, se espera que si se vuelve a pesar se obtendrá nuevamente ese peso. La confiabilidad se obtiene del instrumento de medición, cumpliendo con la condición de que si se vuelve a medir se encuentre lo mismo. Además, la escala derivada de la medida debe permitir entender las diferencias entre una condición y otra. Por ejemplo, la medida de la talla en centímetros usando un tallímetro permite, si se mide a un sujeto en las mismas condiciones, obtener la misma medida y las diferencias entre centímetros de la escala permiten entender quién de la población objeto es más o menos alto.
- **VALIDEZ:** Por validez se entiende que la variable (a través del dato) responda a la pregunta de investigación (mida lo que debe medir). Otra forma de expresarlo es que la variable mide lo más cercano a la realidad. Por supuesto que resulta bastante obvia esta condición; sin embargo, no siempre se cumple, pues la validez exige ser **probada**, es decir, que la variable ciertamente contesta (o sirve para contestar) la pregunta de investigación. Para ello una forma de hacerlo es validando las variables o lo que resulta más pragmático, usando variables ya validadas de otras investigaciones.



Por ejemplo, el test de APGAR familiar que mide la percepción del individuo sobre la funcionalidad de su familia. ¿Es válido? Sin duda... es una percepción que refleja lo que percibe el individuo. Pero ¿será confiable? Es decir, que si se mide hoy y vuelve y se mide en siete días... ¿cambiará? Sería necesario hacer una investigación para contestar, pero lo probable es que cambie de acuerdo a lo que suceda en la relación del individuo con su familia. Si hoy se hacen las preguntas del APGAR FAMILIAR y mañana se repiten (luego de una discusión del encuestado con la mamá, por ejemplo), puede encontrarse una medida diferente. En ambos casos, lo que se mide es válido, pero el test no es confiable: cambia de acuerdo a la situación. Esto se conoce como **variación**.

- **PRECISION:** Es la suma de los dos atributos previos: confiabilidad y validez, siendo por tanto, el atributo ideal.
- **APLICACIÓN DE LA ESTADÍSTICA:** Una buena medida, dado que compara poblaciones o grupos debe permitir que se le puedan aplicar los referentes estadísticos. Algunas de las medidas de frecuencia son ideales para hacer uso de la estadística descriptiva (media, moda, mediana, curva de distribución, desviación estándar). Cuando se va a comparar, el usar medidas que permitan aplicar las pruebas de hipótesis resulta importante para los resultados de la investigación.

2.8. Paso 10. Manejo estadístico de las variables

Las variables arrojan datos. Del buen manejo que se haga de los datos dependerá el entendimiento de los resultados y, por consiguiente, la posibilidad de una buena discusión. Ese es el papel de la estadística: permitir un buen manejo del dato a través del cual se pueda ver más.



Los datos (provenientes de las variables) deben ser manejados usando los referentes estadísticos. El uso de la estadística descriptiva con sus medidas (referentes) de tendencia central (media, moda y mediana) o las de dispersión (desviación estándar, rangos) y con la tabla de frecuencia constituye lo básico en el manejo de datos.

El investigador debe poder predecir cómo va a manejar estadísticamente cada dato proveniente de las variables. A manera de ejemplo, de las variables de identificación de población, deberá saber si expresará la edad media de la población con la desviación estándar o usará una tabla de frecuencia con edades estratificadas. Si usa una escala debe tener en cuenta cómo va a expresar los resultados.

Realmente hacer el ejercicio de identificar los datos de las variables y cuál será el manejo estadístico es importante, pues de no hacerlo la posibilidad de perderse con los resultados es grande. Implica revisar con sumo cuidado el cuestionario donde se va recopilar la información y ver de cada variable cómo espera manejar el resultado. Particular cuidado se ha de tener cuando se usan escalas, pues ellas implican sumar datos y obtener un nuevo resultado, a lo mejor no tan evidente en el cuestionario.

En las investigaciones analíticas se debe tener en cuenta cuáles serán las variables que serán sometidas a la comparación y de una vez precisar cuál será la prueba estadística que se usará para comprobar la presencia o no de diferencias significativas.

2.8.1. Base de datos

La base de datos es el registro donde se traspasa la información obtenida (variables) de la población objeto. Su construcción es hoy más sencilla gracias a la diversidad de programas de computador existentes. La forma más rudimentaria de hacerla es construirla a partir de una hoja de cálculo, pues a través de ella se podrá procesar la información ulteriormente.



Hoy, por la generalización del sistema operativo Windows, la hoja de cálculo más usada es la denominada EXCEL del software de Microsoft.

			¿CÓMO SIENTE A SU FAMILIA?									
CLAVE	1. EDAD	2. SEXO	7. ¿Esta satisfech o con el apoyo que le brinda su familia?	8. ¿Cada satisfech o como la familia le permite participar en las escuelas?	9. ¿Cada satisfech o en la forma como la familia lo impulsa en sus estudios?	10. ¿Esta satisfech o como su familia le brinda el afecto?	11. ¿Cada satisfech o en la forma como comparte el tiempo con su familia?	12. ¿Cómo está con su familia?				
			D. CASI NUNCA	D. CASI NUNCA	D. CASI NUNCA	D. CASI NUNCA	D. CASI NUNCA	D. CASI NUNCA	Más de 10 no aplica			
			1. A VECES	1. A VECES	1. A VECES	1. A VECES	1. A VECES	1. A VECES				
			2. CASI SIEMPRE	2. CASI SIEMPRE	2. CASI SIEMPRE	2. CASI SIEMPRE	2. CASI SIEMPRE	2. CASI SIEMPRE				
			3. NO ENTIENDE LA PREGUNTA	3. NO ENTIENDE LA PREGUNTA	3. NO ENTIENDE LA PREGUNTA	3. NO ENTIENDE LA PREGUNTA	3. NO ENTIENDE LA PREGUNTA	3. NO ENTIENDE LA PREGUNTA				
			444. NO APLICA	444. NO APLICA	444. NO APLICA	444. NO APLICA	444. NO APLICA	NO APLICA				
aclave			edad	asexo	faga	fagimp	fagaf	fagap				
1	10	2	444	444	444	444	444	444	2220			
2	10	1	2	2	2	2	2	2	10			
3	10	1	1	1	1	2	2	2	8			
4	10	1	1	1	1	2	2	2	9			
5	10	1	1	1	444	0	2	1	448			
6	10	2	2	2	2	1	2	1	8			
7	10	1	2	2	2	2	2	2	10			
8	10	2	0	2	2	2	2	1	7			
9	10	1	1	1	2	1	2	1	7			
10	10	2	2	2	2	2	2	2	10			

En el ejemplo se está observando la base de datos alimentada con una encuesta donde se aplicó un cuestionario para medir la percepción de funcionalidad familiar en un grupo de jóvenes. Cada encuesta debe ser enumerada (aparece como clave en la imagen), cada columna es para una variable, las filas para los encuestados y las preguntas deben tener las probables respuestas (datos), en este caso codificadas a través de un número. En el proceso, con el fin de abrir la base de datos con el programa EPIINFO, se insertó una fila con siglas codificadas que identifican las preguntas. La variable destinada en la última columna muestra los resultados de la suma de los puntajes obtenidos en las pregunta 7 a la 11, para así, obtener el resultado de esta escala de Apgar familiar (véase en la figura anterior el encabezado: 'Cómo siente a su familia').

2.8.2. Paso 11. Construcción del formulario

La pregunta de investigación requiere construir un instrumento donde se recopilan las variables a través de las cuales se obtendrá la respuesta. Dicho instrumento se denomina cuestionario. Como tal, cada pregunta (variable) ha de procurar cumplir con la condición de que sea sencilla y entendible. Preguntas cortas, y con respuestas cerradas son útiles en este sentido. Ideal es utilizar preguntas ya probadas por otras investigaciones. No usar muchas preguntas (para evitar la posibilidad de cometer errores originados por el cansancio), así como crear diseños agradables a los ojos para los formularios, resulta particularmente importante. En las preguntas, las posibilidades de respuesta a través de los reactivos han de ser excluyentes (la respuesta de uno descarta el resto) y, por supuesto, evitar que sugieran las respuestas (ofrecer todas las posibilidades).



El instrumento de una investigación es el medio seleccionado para indagar sobre las variables que contestan las preguntas de investigación. Pueden ser, entre otros, cuestionarios a través de los cuales se obtienen los datos. Un formulario puede ser preguntado o auto diligenciado. Pero, así mismo, un instrumento puede ser el equipo o aparato que se utiliza para recoger el dato. Por ejemplo, si se trata de una investigación acerca de la tensión arterial, el tensiómetro será el instrumento.

En torno a los cuestionarios (instrumentos con preguntas) resulta evidente que tienen menos sesgos los autos diligenciados (es más probable que lo que escriban sea la verdad en comparación con un cuestionario preguntado por el encuestador).

Todo formulario debe permitir ver las respuestas que se obtendrán de la investigación de la siguiente manera:

- Con variables que describan las características de la población objeto
- Con variables que expongan las frecuencias
- Con variables que expongan los resultados.

Los resultados en los artículos de investigación siempre exponen estas tres condiciones.

Claro... el formulario resulta clave para la tabulada de la información. De su buen diseño y de las escalas adecuadas de las respuestas, depende en buena medida, el grado de error probable al tabular. Además, el diseño de un cuestionario es igualmente importante. El que sea (**en su diseño**) afable al ojo del encuestado, contribuye a que lo llene con mayor veracidad.

Hacer un cuestionario es clave en una investigación que escoge este instrumento como el principal mecanismo. ¿Cómo saber si está bien hecho? Pues probándolo en una población similar a la que se va a estudiar. Esto se conoce como **prueba piloto**. Por ejemplo: se hizo un cuestionario de conocimientos actitudes y prácticas de sexualidad para aplicar en un colegio. Entonces se recomienda que se aplique en un grupo de jóvenes de características similares a los del colegio escogido. Las dudas que surjan, las interpretaciones que ofrezcan han de hacer que el cuestionario se corrija. Sólo así se podrá tener la certeza de que el cuestionario, cuando se aplique en la población objeto, será entendido.

2.9. Paso 12. Plan para el trabajo de campo

Hasta este momento, el investigador sabe qué es lo que quiere. Ha definido una pregunta, ha revisado, sabe al respecto. Debe haber escogido un diseño, unas variables las cuales sabe cómo va a manejar y, claro, ha logrado identificar a la población objeto. Ahora debe efectuar un plan de actividades encaminado a realizar la investigación, con fechas (cronograma) y debe obtener la información precisa de los recursos que va a necesitar. Pensar y definir cómo se hará todo esto es lo que se denomina un **plan**. Si bien existen muchas matrices de planeación, la propuesta que hacemos es la siguiente:

- Mención de la actividad. Se la nombra.
- Definición de la actividad: Se dice en qué consiste
- Descripción de la misma. Se dice qué se va hacer y cómo se hace.
- Tiempo: Se dice cuánto tiempo se toma hacer la acción.
- Funciones: Se menciona quién la debe hacer.
- Infraestructura: Se dice qué sitio se necesita para hacerla.
- Equipos: Se menciona el equipo que se necesita para hacer la actividad.
- Materiales: Se describen los materiales que se necesitan.

Respecto a las actividades por hacer, una probable clasificación sería:

I. ACTIVIDADES DE CONSECUCIÓN DE LA POBLACIÓN

II. ACTIVIDADES DE TRABAJO DE CAMPO

III. ACTIVIDADES TENDIENTES AL ANÁLISIS DE LA INFORMACIÓN

IV. ACTIVIDADES PARA LA PRESENTACIÓN DE LA INVESTIGACIÓN.

2.9.1. Descripción de las actividades

Consecución de la población

Se expuso que uno de los criterios de factibilidad más importante es el acceso a la población. En salud, la población objeto puede ser enfermos, o personas con características particulares. En general, están agrupados en algún sitio, tales como instituciones de salud (enfermos), colegios (adolescentes), hogares comunitarios o clubes de la tercera edad.



- Para acceder a la población, lo primero que se ha de hacer es el **análisis ético de la investigación** y presentárselo a los comités respectivos, habitualmente constituidos en los lugares donde están los investigadores. De dicha reunión ha de quedar un acta y el objetivo es garantizar la beneficencia para la población objeto.



De hecho una buena parte de los comités editoriales de las revistas biomédicas no aceptan artículos donde no demuestren con nitidez el cumplimiento de los aspectos éticos.

- **Cartas de presentación:** El siguiente paso es elaborar las cartas de presentación para los sitios donde está la población objeto. Deben tener en su constitución:
 - a) Presentación de los investigadores. Quiénes son y de dónde vienen.
 - b) Presentación de la investigación que se quiere hacer. Brevemente explicar lo que quieren indagar y cómo eso beneficia a la población.
 - c) Explicar lo que se le hará a la población objeto. Explicar las actividades del trabajo de campo que se plantean (encuestas y cuándo).
 - d) Presentar las consideraciones éticas. Hacer notas del cumplimiento de los principios éticos.
 - e) Presentar formato de consentimiento informado cuando sea necesario.
 - f) Mencionar qué entidad respalda la investigación y a cargo de quién están.
 - g) Mencionar qué se le entregará a la entidad colaboradora. Explicar lo que se comprometen entregar.
- **Explicación a la población objeto:** El investigador debe programarse para explicar a cada participante acerca de la investigación y estar dispuesto a resolver las dudas que surjan, así como garantizarle el derecho a no participar.

Trabajo de campo

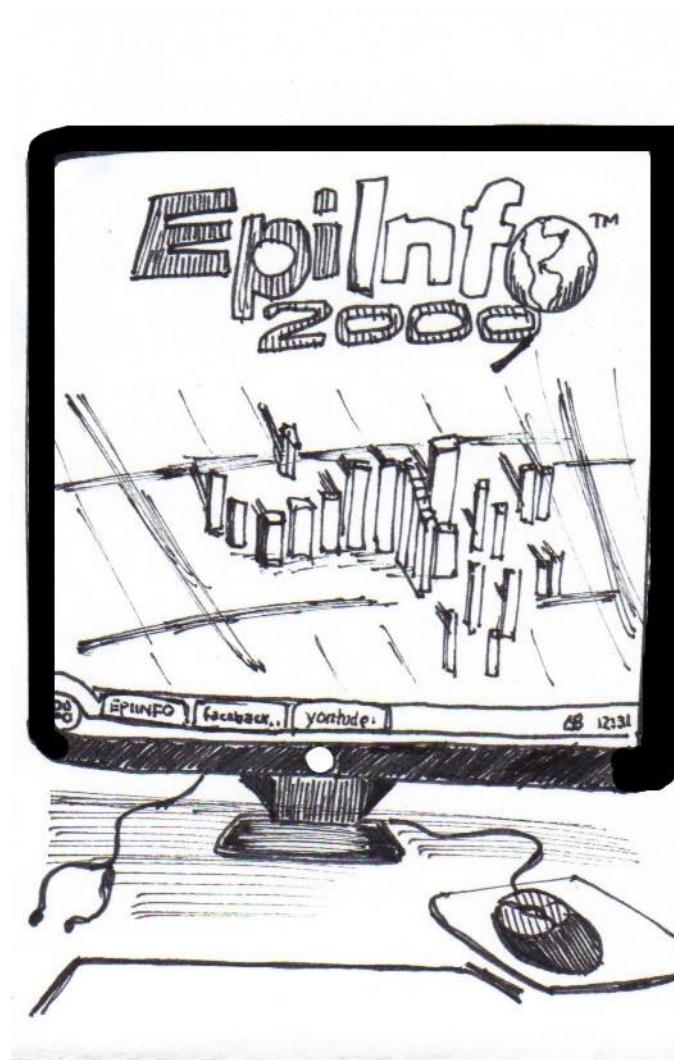


- **Preparación de los instrumentos:** Conseguida la población, el investigador debe replicar los instrumentos que empleará. Para ello debe haber efectuado la prueba del instrumento luego de la cual definirá, de acuerdo a la población objeto, el proceso de replicación del instrumento.
- **Ejecución de la investigación:** Se refiere a la programación de las actividades atinentes a aplicar los instrumentos en la población objeto. Se debe tener en cuenta:
 1. El sitio donde aplicará el formulario. Debe tener unas condiciones adecuadas para tal tarea. La posibilidad de privacidad si es una

encuesta o cómo disponer de los espacios, son aspectos a considerar por parte del investigador.

2. Asegurar el anonimato de la información: referida a que el investigador cree condiciones que aseguren que en la información entregada, no se buscará identificar a los sujetos objeto del estudio.
3. Recolección de la información: El investigador debe pensar en todas las actividades necesarias para que una vez recogida la información, pueda ubicarla en el sitio donde será procesada. Implica entre otras, guardar los formularios debidamente marcados y asegurar el transporte de los mismos. A este respecto se deberá tener en cuenta:
 - Preparación de encuestadores: los encuestadores deben saber exactamente cómo harán la toma de información.
 - Preparación de la población objeto: Se debe saber qué se le debe decir o cómo organizar a la población objeto para lograr cumplir con la recolección de información.

Análisis de la información



Una vez se tiene la información, se debe diseñar todo el proceso de tabulación y obtención de la respuesta. El primer paso es:

- a. **Construcción de la base de datos:** Se refiere al proceso a través del cual se trascibe toda la información obtenida (datos) a algún programa de computador que permita trabajar dichos datos. Para poder cumplir con esta actividad se necesita pensar en cómo se tabulará la información, quién lo hará, cuándo, qué necesita para hacerlo.

- b. **Obtención de los resultados:** Si las variables empleadas han sido adecuadas, el proceso de obtención de resultados ha de ser sencillo. Implica, por supuesto, el saber trabajar el programa de computador. Existen programas que facilitan esta actividad. El investigador debe, en lo posible, conocer y manejar alguno de ellos.
- c. **Manejo estadístico de la información:** Es importante elaborar un análisis estadístico de los resultados. La obtención de intervalos de confianza en las frecuencias (ver numerales 4.4.2.), así como las pruebas de hipótesis, se convierte en un importante aporte a la investigación. Decidir cómo se va a hacer resulta importante.

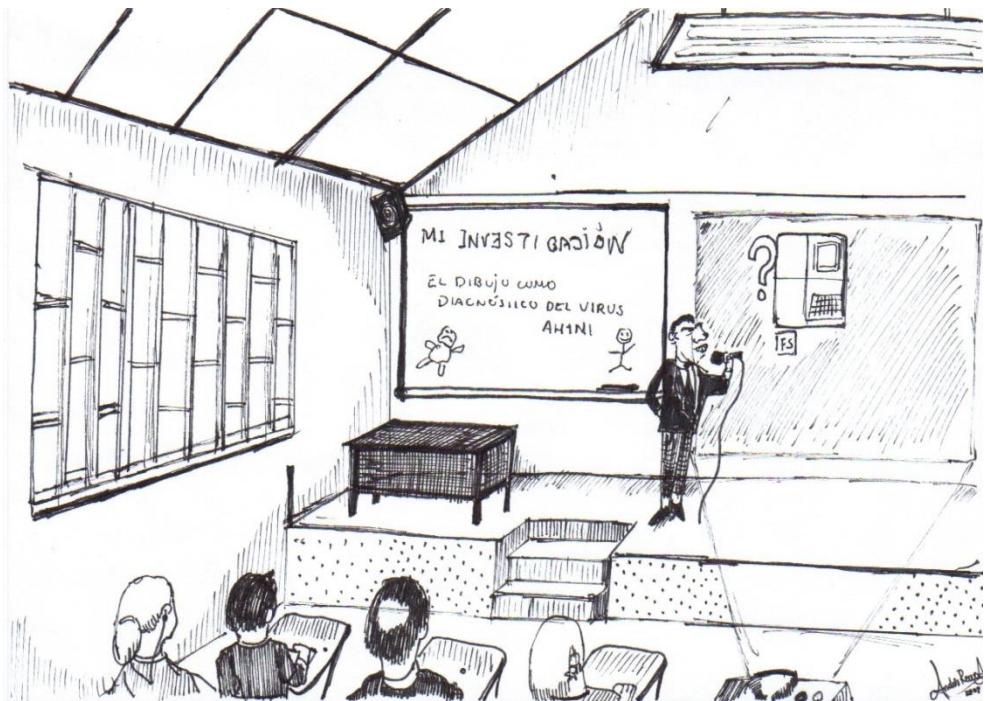


Existe una enorme cantidad de programas que ayudan a construir y manejar bases de datos para obtener las respuestas de la investigación. El programa EPI INFO del CDC es uno muy usado y que está al alcance de todo el mundo ya que su distribución es gratuita.

Presentación de la investigación

Las últimas actividades que se deben programar son las referidas a la divulgación de la investigación. Implican:

- 1. **Redacción del informe:** Contempla todas las actividades destinadas a redactar el informe de la investigación, el cual habitualmente se rige bajo las normas de la revista (si es un artículo) o de presentación (si es tesis).
- 2. **Elaboración de la presentación.** Usa criterios previamente establecidos.



La presentación de las investigaciones ante auditorios es clave en la divulgación. Por lo general, no debe tomar más de 20 minutos; consta de una introducción donde se presenta el contexto y la pregunta de investigación con la importancia de contestarla. El desarrollo, donde se expone la metodología y los resultados. Y las conclusiones con el análisis donde se presentan la interpretación y la aplicación de lo investigado.



2.9.2. Recursos que se necesitan para cumplir con las actividades

Una vez el investigador ha definido todas las actividades, debe explicar qué necesita para poder ejecutar cada una de ellas. Esto se conoce como el análisis de recursos. Repitiendo lo expuesto previamente se debe identificar **de cada actividad**:

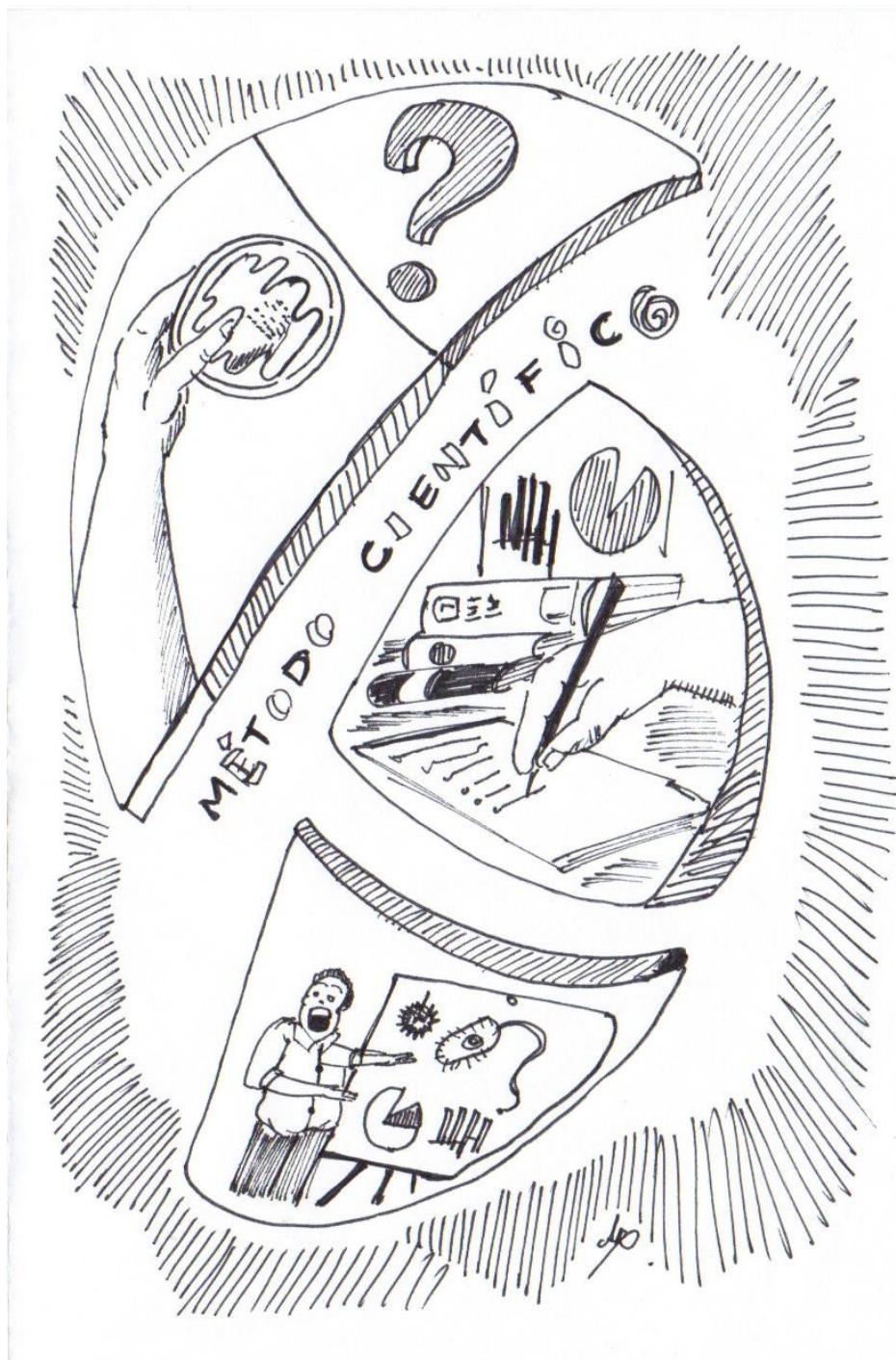
- Mención de la actividad. Se la nombra.
- Definición de la actividad: Se dice en qué consiste.
- Descripción de la misma. Se dice qué se va hacer y cómo se hace.
- Tiempo: Se dice cuánto tiempo se toma hacer la acción.
- Concentración: Cuántas veces se debe hacer la actividad para que quede completa.
- Funciones: Se menciona quién la debe hacer.
- Infraestructura: Se dice qué sitio se necesita para hacerla.
- Equipos: Se menciona el equipo que se necesita para hacer la actividad.
- Materiales: Se describen los materiales que se necesitan.

Contestado todo lo anterior es difícil que se haya olvidado algo y, por ende, se garantiza la posibilidad real de ejecución.



Más de una vez los investigadores creen que con solo la revisión del marco teórico la investigación se podrá efectuar. Sin embargo, podemos asegurar que sólo los que planeen bien, tendrán éxito.

2.10. Paso 12. El protocolo de investigación



Estará de acuerdo que una vez concluidos estos pasos, el investigador está listo para realizar la investigación. Las otras fases son las que se dan durante la ejecución de la investigación. Si ello es cierto, nada mejor que dejar por escrito dicha preparación. Esto es lo que se conoce como **protocolo de investigación**. A continuación se intentará construir una guía de preguntas para que el investigador las responda, esperando que al hacerlo, constate que ha pensado en todo. Se va a intentar plantearlas en orden (usando números), dejando en claro que existen otros modelos de protocolo.

Paso 1

- 1.1. ¿Explique cuál es el tema escogido? Acá debe usar referencias.
- 1.2. Dé tres explicaciones de por qué escogió este tema.

- ¿Tiene alguna experiencia en el tema?
- ¿Existe alguna facilidad para usted sobre este tema?
- ¿Sabe si existe bibliografía al respecto?

Paso 2

- 2.1. Redacte el problema a manera de pregunta.
- 2.2. Defina el significado de los componentes del problema o pregunta (ponga referentes).
- 2.3. Mencione cuáles son las probables respuestas que se intuyen (sólo si es un trabajo analítico, pues si es descriptivo debe poner lo que espera describir).
- 2.4. Mencione cuál es la importancia de resolver esta pregunta (con referencias).

Paso 3. Revisión bibliográfica

- 3.1. Qué disciplina(s) se dedica(n) a la pregunta.
- 3.2. Mencionen cuáles son las palabras claves para la búsqueda en bases de datos (términos Mesch).
- 3.3. Anexe los resúmenes de los artículos que haya revisado (sigue la guía de cómo se lee un artículo).

Paso 4. Definición del marco teórico

- 4.1. Mencione debidamente referenciadas las teorías o modelos conceptuales que revisó acerca del problema y/o sus componentes (qué se sabe al respecto).
- 4.2. Mencione qué variables le son útiles en la investigación, junto con las fuentes (de dónde las obtuvo).

- 4.3 Mencione qué diseños se han empleado para contestar la pregunta o, en su defecto, para abordar los componentes de la pregunta.



Por ejemplo, si el problema es *¿Cuál es la relación entre A y B en la población C?*:

Los componentes son: A, B y C. Entonces el marco teórico no debe olvidar la integración de estos tres aspectos. Una propuesta de su desarrollo puede ser la siguiente:

1. Describir y explicar qué es A.
2. Describir y explicar qué es B.
3. Describir qué es C.
4. Explicar cómo se relaciona A, con B y con C.

La idea es que el investigador se vuelva un experto en estos cuatro aspectos. Claro... el marco teórico surge de la revisión de artículos, por lo tanto, aprender a revisarlos resulta particularmente importante.

Paso 5. Formulación de la hipótesis

- 5.1. Enuncie las respuestas probables a la pregunta de investigación.
Para ello:
- 5.1.1. Describa dónde y qué encontró para que sea una u otra respuesta (referenciado).
- 5.1.2. Si el diseño se lo permite: Plantee su hipótesis alterna y la hipótesis nula.

Paso 6. Selección del tipo de diseño

- 6.1. Enuncie el tipo de diseño que va a emplear.
- 6.2. Mencione qué actividad va a desarrollar: Descripción, comparación o intervención.
- 6.3. Mencione cómo va a manejar el tiempo: transversal, retrospectivo o prospectivo.
- 6.4. Mencione si piensa trabajar con población total o muestra. Si es esto último, explique si será aleatoria o por conveniencia.

Paso 7. Identificación de la población

- 7.1. Mencione los criterios de elegibilidad de la población.
- 7.2. Si va a emplear muestra explique:
- 7.2.1. Cómo va escoger el número.
- 7.2.2. Cómo va escogerla: aleatoria o por conveniencia.
- 7.2.3. Explique cómo
- 7.3. Mencione cómo cumplirá los principios éticos:
- 7.3.1. ¿Puede garantizar que no existe riesgo de daño?
- 7.3.2. ¿Qué beneficios se desprenden para la población objeto?

- 7.3.3. ¿Cómo asegura que puedan o no participar?
- 7.3.4. ¿Cómo garantiza la confidencialidad?
- 7.3.5. ¿Cómo asegura el buen trato de los participantes?
- 7.4. Si necesita consentimiento informado: descríbalo.

Paso 8. Selección de las variables

- 8.1. Describa las variables que va a emplear.
 - 8.1.1. Describa los datos que se pueden obtener de cada variable.
- 8.2. Explique cómo, a través de las variables, puede contestar la pregunta de la investigación.
- 8.2.1. Mencione de dónde sustenta (bibliografía) las variables que va a emplear.

Paso 9. Construcción del formulario

- 9.1. Presente el formulario que va a emplear.
- 9.2. Deje constancia de las referencias de donde obtuvo cada variable.
- 9.3. Exponga el diseño del cuestionario.

Paso 10. Plan de la investigación

- 10.1. Explique cómo se va a conseguir la población.
- 10.2. Diseñe una carta de presentación para entidad donde está la población.
- 10.3. Mencione qué y cuántos instrumentos va a necesitar para hacer la investigación.
- 10.4. Mencione dónde efectuará el trabajo de campo.
- 10.5. Explique cómo garantizará los principios éticos en el sitio donde obtendrá la información.
- 10.6. Mencione cómo va a tabular la información.
 - 10.6.1. Presente la matriz de la base de datos.
- 10.7. Ahora coloque fechas a cada una de estas actividades.
- 10.8. Calcule qué va a necesitar para cada actividad: en tiempo, en personas, en equipos, en infraestructura, en materiales.

Paso 11. Estudio piloto

- 11.1. Mencione qué población le puede servir de prueba para el cuestionario.
 - 11.1.1. Mencione cómo puede acceder a la misma.
- 11.2. Mencione cuándo puede hacer una prueba de su investigación.
- 11.3. Presente el cuestionario corregido luego de la prueba piloto.
- 11.4. Presente la base de datos ya probada.
- 11.5. Explique cómo se obtendrá la respuesta a partir de la base de datos.

FASE TRES: TRABAJO DE CAMPO

La fase tres, cuatro y cinco forman parte de la ejecución de la investigación. Como tales, han sido planeadas en párrafos previos (de ahí que el lector tendrá la sensación de que es una repetición). Acá de lo que se trata es de explicar las ejecuciones, las cuales son el resultado diferencial del plan con la capacidad de adaptación, pues al enfrentarse a la realización, los planes se pueden modificar. Un buen investigador debe tener una gran capacidad de adaptación. Sus componentes están destinados a comentar cómo se efectuó el trabajo de campo, queriendo con ello decir que cualquiera que los lea tenga la posibilidad de repetir la investigación. Desde el punto de vista de sus componentes, una forma de subclasificar esta fase es:

3. Paso 14. Recolección de la información

Describe lo que se hizo para obtener la información de donde se obtendrán los resultados. Se debe mencionar:

3.1. Cómo se accedió a la población: Explicar cómo se hizo el proceso.

3.2. Cómo se recogió la información. Explicar cómo se manejaron los formularios, explicando su ordenación para ser procesados.

FASE CUATRO: ANÁLISIS DE DATOS

Constituida por los pasos que van desde que se obtuvo la información de la investigación, hasta obtener la respuesta a la pregunta. Consta de:

4. Paso 15. Registro de información recogida en la base de datos

Obtenida la información de la población objeto a través de los cuestionarios se debe trasladar a la base de datos (tabular la información).

The diagram illustrates the process of data entry from a questionnaire into a database. It shows a questionnaire table with variables, responses, and their codification into data and codes.

PREGUNTAS (Variables)

RESPUESTA

PROBABLES RESPUESTAS CODIFICADAS (Datos)

Siglas para Programa EPINNFO

CODIGO

		¿CÓMO SIENTE A SU FAMILIA?						
CLAVE	1. EDAD	2. SEXO	7. ¿Está satisfech o con el apoyo que le brinda su familia?	8. ¿Está satisfech o como la familia le permite participa r en las decisiones?	9. ¿Está satisfech o en forma como la familia lo impulsa en sus intereses?	10. ¿Está satisfech o como su familia le brinda el afecto?	11. ¿Está satisfech o en la forma como compart e el tiempo con su familia?	12. ¿Cómo está con su familia?
		1. MASCULIN O	0. CASI NUNCA	0. CASI NUNCA	0. CASI NUNCA	0. CASI NUNCA	0. CASI NUNCA	Más de 10 no aplica
2. FEMENIN O	1. A VECES	1. A VECES	1. A VECES	1. A VECES	1. A VECES			
NO APlica	2. CASI SIEMPRE	2. CASI SIEMPRE	2. CASI SIEMPRE	2. CASI SIEMPRE	2. CASI SIEMPRE			
	3. NO ENTIENDE LA PREGUNTA	3. NO ENTIENDE LA PREGUNTA	3. NO ENTIENDE LA PREGUNTA	3. NO ENTIENDE LA PREGUNTA	3. NO ENTIENDE LA PREGUNTA			
	444. NO APlica	444. NO APlica	444. NO APlica	444. NO APlica	NO APlica			
aclave	edad	basexo	faga	faap	fagimp	fagaF	fagtie	fappar
1	10	2	444	444	444	444	444	2220
2	10	1	2	2	2	2	2	10
3	10	1	1	1	2	2	2	8
4	10	1	1	2	2	2	2	9
5	10	1	1	444	0	2	1	448
6	10	2	2	2	1	2	1	8
7	10	1	2	2	2	2	2	10
8	10	2	0	2	2	2	1	7
9	10	1	1	1	2	1	2	7
10	10	2	2	2	2	2	2	10

Idealmente el digitador (quien traslade la información) debe estar bien familiarizado con la base de datos y los códigos a emplear. Entre menos digitadores menos margen de error al trasladar la información. Cada encuesta debe ser numerada, facilitando incluso el que se hagan pruebas de la veracidad con que se traslada la información.

4.1. Paso 16. Procesamiento la información

Procesar la información es poder obtener las respuestas de la investigación a partir del manejo de la base de datos. La facilidad depende de la habilidad para manejar la hoja de cálculo o el programa que se esté usando (Epi Info, por ejemplo). Prácticamente todos los resultados se restringen a:

4.1.1. La obtención de las características de la población, la que se expone de manera habitual en una tabla de frecuencia.

4.1.2. Obtención de las frecuencias de las variables estudiadas. También se expresan en una tabla de frecuencias.

Tabla 1: Frecuencias de las variables independientes y dependientes de los adolescentes de 12 colegios públicos de la localidad de Suba en el 2006.

	Nº	%
EDADES		
De 10 A 12 años	3 965	32
De 13 A 16 años	7 200	59
De 17 A 19 años	997	8
GENERO		
Mujeres	6 274	51
ESTRATO		
Estrato 1	347	3
Estrato 2	8 958	73
Estrato 3	2 475	20
ESTRUCTURA FAMILIAR		
Nuclear completa	6 862	56
Nuclear incompleta	3 774	31
Recompuestas	1 112	9
Hogares donde faltan ambos padres	227	2
FUNCIONALIDAD FAMILIAR		
Buena función familiar	7 722	63
Disfunción moderada	2 731	22
Disfunción severa	920	7
RIESGOS EN TORNO A LA SEXUALIDAD		
Presencia de relaciones sexuales	2 078	17
Ausencia de protección en la primera relación sexual	1 791	66
Relaciones sexuales con más de una pareja en el último año	511	4
RIESGOS EN TORNO A LAS ADICCIONES		
Tomó una vez (o más) la última semana	3 669	30
Actualmente fuma	2 023	16
Bebe alcohol y le gusta	2 024	16
RIESGOS EN TORNO A LA DEPRESIÓN		
Haber pensado en el suicidio	4 182	34
Haber pensado en el suicidio y querer hacerlo	446	4
RIESGOS EN TORNO A LA VIOLENCIA		
Haber sido golpeado físicamente durante el último año	3 715	30
Haber golpeado físicamente a alguien durante el último año	3 816	31
Llevar un arma para defenderse	1 063	9
RIESGOS EN TORNO A LOS TCA		
Mujeres con TCA	2 821	23

El diseño de la tabla de frecuencia está sujeto a las especificaciones de las revistas, las que con pequeñas variaciones entre sí, informan en sus instructivos para publicar las exigencias pertinentes. Una tabla de frecuencia debe ser titulada con: número de tabla (orden), expresando qué está mostrando, de quién y cuándo se obtuvo la información.

Debe aparecer antes de empezar a mencionarla. La tabla usada de ejemplo une las características de la población con las frecuencias estudiadas en una investigación sobre adolescentes escolarizados y riesgos psicosociales en la localidad de Suba en el año 2006.

4.1.3. Obtención de la respuesta a la pregunta de investigación, usualmente mediante una tabla comparativa.

TABLE 2. Unconditional logistical regression of high-risk psychosocial behaviors and the family of adolescents from 12 public schools in the locality of Suba in the year 2006.

HIHG RISK BEHAVIORS	YES	NO	OR	IC 95%	
				LL	UL
Having Being Assaulted in the Past Year					
Gender					
Male	2 062 (37)	3 574 (63)	1.7	1.51	1.8
Female	1 639 (28)	4 224 (72)			
Family					
Severe Dysfunction	520 (59)	362 (41)	4.1	3.46	4.75
Functional	1 836 (25)	5 418 (75)			
Carrying a Weapon					
Gender					
Male	752 (13)	4 864 (87)	3.1	2.63	3.57
Female	305 (5)	5 584 (93)			
Family					
Moderately dysfunctional	299 (11)	2 311 (89)	1.5	1.28	1.77
Functional	524 (7)	6 737 (93)			
Severe Disfunction	152 (17)	732 (83)	2.4	1.9	3.0
Functional	524 (7)	6 737 (93)			
Eating Disorders					
Family					
Reconstituted	232 (42)	320 (58)	1.5	1.2	1.8
Blood relationship	1 571 (31)	3 451 (69)			
Moderately dysfunctional	741 (30)	1 717 (70)	1.7	1.4	1.9
Functional	915 (27)	2 531 (73)			
Severe Disfunction	257 (50)	255 (50)	2.3	1.9	2.8
Functional	915 (27)	2 531 (73)			
Sexual Relations					
Gender					
Male	1 266 (22)	4 607 (78)	1.9	1.72	2.15
Female	801 (13)	5 312 (87)			
Family					
Incomplete	875 (19)	3802 (81)	1.1	1	1.27
Complete	1060 (16)	5 690 (84)			
Moderately dysfunctional	581 (22)	2106 (78)	1.4	1.21	1.56
Functional	1 108 (15)	6 485 (85)			
Severe Disfunction	282 (31)	622 (69)	2.0	1.69	2.43
Functional	1 108 (15)	6 485 (85)			

Para llegar a presentar la respuesta a la pregunta de investigación, se deben haber sometido los datos al manejo estadístico correspondiente (que debe haber sido planeado previamente) y con base en ello presentarlo en una tabla. En la tabla usada para este ejemplo, la pregunta de investigación es qué relación existe entre comportamientos de riesgos y las diferencias de género y la función y estructura familiar en una población escolar de la localidad de Suba. Como se observa en el título, la tabla de contingencia fue sometida a un análisis de regresión (prueba estadística). Igualmente la tabla permite ver todas las frecuencias con sus respectivos porcentajes (en paréntesis), la medida obtenida (Odds Ratio) con sus respectivos intervalos de confianza, haciendo notar que en las tablas solo se exponen los que han sido significativos.

FASE CINCO: DIVULGACIÓN DE LA INVESTIGACIÓN

"Lo que no publica... no existe", sentencia definitiva y que expresa la importancia de divulgar las investigaciones. El asunto es que este proceso se ha venido protocolizando exigiendo de los investigadores la capacidad de seguir instrucciones, componente para el cual, algunos, no hemos sido bien entrenados.

Toda investigación debe ser divulgada. La mención del material y métodos del proceso ayuda enormemente a muchos otros que usarán esa experiencia. Y en cuanto a los resultados, siempre son valiosos, ya que servirán a otros investigadores interesados en el tema. Si bien las investigaciones donde se pueden inferir son muy importantes, aquellas que no lo hacen por su diseño, son igualmente interesantes y deben ser divulgadas. A continuación se mencionarán algunos aspectos a tener en cuenta para la divulgación.

5. Paso 17. Publicación de la investigación

Existen dos modalidades de publicación de investigaciones; las tesis y los artículos. Las primeras son usadas en procesos pedagógicos donde, y a través de la descripción minuciosa y exhaustiva, los investigadores reciben correcciones que hace el tutor. Son, por principio, extensos documentos donde cada paso queda debidamente descrito. Los artículos en cambio resaltan por la búsqueda de brevedad, queriendo con ello facilitar a los interesados el que lean exactamente lo que necesitan. Claro... el artículo va dirigido a una comunidad muy específica de lectores y de hecho existen revistas especializadas en temas y áreas, que los investigadores deben buscar para divulgar sus hallazgos.

5.1. Cómo escribir un artículo

"Lo breve y bueno dos veces bueno, lo largo y bueno, no es tan bueno y lo malo y breve no es tan malo".

Los artículos deben ser presentados con el siguiente formato (normas de Vancouver):

1. **TITULO:** Compuesto por el título de la investigación y los autores.
 - 1.1. Título de la investigación: Debe decir QUÉ se hizo, CÓMO se hizo y en QUIÉN se hizo. Se puede agregar DÓNDE y CUÁNDO.

1.2. Autores: Apellidos y nombres con el último título académico y declarando si existen o no conflictos de interés.

2. **RESUMEN:** Máximo de 250 palabras. Debe tener:

- 2.1. Objetivo del estudio: mención del objetivo general del estudio.
- 2.2. Material y métodos con:
 - 2.2.1. Tipo de estudio (dos o tres palabras).
 - 2.2.2. Población: mencionar las características principales de la población (un párrafo).
 - 2.2.3. Instrumentos: Mencionar las principales variables empleadas (uno, máximo dos párrafos).
 - 2.2.4. Manejo estadístico: Mencionar qué medidas se obtuvieron, cuál fue el manejo estadístico y, si se usó algún programa, mencionar cuál.
- 2.3. Resultados: mencionar los principales resultados con su correspondiente significancia.
- 2.3.1. Conclusiones: Mencionar la principal conclusión (o respuesta de la pregunta de investigación) Un párrafo.
- 2.4. Palabras claves: Mencionar cinco palabras claves, necesariamente términos MeSH.

El resumen debe escribirse en español y en inglés.

3. **INTRODUCCIÓN:** (Máximo media página a un espacio). Se debe mencionar:

- 3.1. El contexto de la investigación (con referencias)
- 3.2. La pregunta o los objetivos
- 3.3. La importancia de contestarla.

4. **MATERIAL Y MÉTODOS:** Ha de ser lo más explicativo posible. Debe tener:

- 4.1. **Tipo de estudio:** mencionar las características del diseño (tres palabras máximo).
- 4.2. **Población:** Mencionar cómo se consiguió la población. Decir cuáles fueron los criterios de inclusión y exclusión. Cómo se preservaron los criterios éticos.
- 4.3. **Variables:** mencionar las variables empleadas con referencias de dónde las obtuvieron, explicar cómo ellas contestan la pregunta de investigación, mencionar las medidas empleadas y el manejo estadístico.
- 4.4. **Recolección de la información:** explicar si hubo o no prueba piloto, luego explicar cómo se recogió la información, teniendo el cuidado de no mencionar nombres propios (confidencialidad). Luego, qué se hizo y cómo se procesó.

5. RESULTADOS: Debe tener:

- 5.1. Características de la población, idealmente en una tabla.
- 5.2. Frecuencias encontradas: mencionar las frecuencias idealmente en una tabla.
- 5.3. Respuesta a la pregunta de investigación: idealmente en una tabla. Este último aspecto es propio de investigaciones analíticas e intervenciones.

6. DISCUSIÓN: Tiene tres partes:

- 6.1. Una explicación de los sesgos probables y cómo se evitaron.
- 6.2. Una comparación de las frecuencias encontradas con las referenciadas.
 - 6.2.1. Acá dar una explicación del por qué de las diferencias.
- 6.3. Comparación de la "respuesta" encontrada con las referenciadas.
 - 6.3.1. Explicar el por qué de las diferencias.

7. BIBLIOGRAFÍA: Seguir las normas para citarla.

8. AGRADECIMIENTOS: Mencionar a quién se le dan (aquí sí se puede hacer mención por nombres propios).



Un artículo ideal, sin tablas y sin bibliografía debe tener entre 3000 y 3500 palabras (de acuerdo a la revista). Las revistas biomédicas imprimen en cada número las instrucciones que deben seguir los que estén interesados en publicar.

5.2. Paso 18. Presentación de la investigación

Las investigaciones pueden ser escogidas para ser presentadas en conferencias. Al respecto, recordar que el tiempo que habitualmente se da es de unos 20 minutos. La presentación debe tener tres partes:

- Introducción: donde se expone el contexto, la pregunta de investigación y la importancia de contestarla.
- Desarrollo: donde se presentan los resultados y se discuten.
- Conclusiones: donde el expositor termina mostrando las implicaciones de los resultados.

Otra modalidad de presentación es la de carteleras (una denominada póster) donde, y con unas imágenes sugestivas, se expone el resumen haciendo énfasis en las conclusiones.

De esta manera se han descrito de manera ordenada las fases de una investigación. Recuerde que lo que se ha expuesto es una forma de clasificar, aceptando que existen otras. Esta será útil si le permite a los futuros investigadores (usted) asesorar su proceso y concluirlo de manera exitosa.



Capítulo **4**



El papel de la Estadística: Construyendo un idioma entendible

<u>4.1. Los datos</u>	219
<u>4.2. Las medidas</u>	223
4.2.1. Medidas de frecuencia	224
4.2.1.1. Concepto de indicadores: Palabras para una definición.....	226
4.2.2. Medidas de asociación	231
4.2.3. Medidas de impacto	232
<u>4.3. La estadística descriptiva</u>	232
4.3.1. La tabla de frecuencia.....	233
4.3.2. Las gráficas.....	236
4.3.3. Las medidas de tendencia central.....	241
4.3.3.1.La media o promedio	243
4.3.3.2. La moda.....	246
4.3.3.3.La mediana	247
4.3.4. Las medidas de dispersión.....	248
4.3.4.1. El recorrido	249
4.3.4.2. Los recorridos intercuartiles	251
4.3.4.3. La desviación estándar (con los pasos para obtenerla)	252
4.3.5. La curva de distribución.....	256
4.3.5.1. Explicación de la curva de distribución y su utilidad a través de las tablas de peso y talla.....	263
4.3.5.2. La curva de distribución y la similitud y diferencias estadísticas	268
4.3.5.3. La curva de distribución y las hipótesis estadísticas.....	271
4.4. La estadística analítica	278
1.4.1.La herramienta esencial de la comparación (Tabla de contingencia o de 2X2) y su aplicación con IC del 95% y prueba de la p	281
1.4.1.1.Ejemplo a través de un brote del uso de la tabla	282
1.4.1.1.1.El concepto de significancia a través del ejemplo.....	285
4.4.1.1.1.2. Ampliando la prueba p	289
4.1.1.1.3. La prueba Ji (Chi cuadrado)	292
4.1.1.1.4.Entendiendo la prueba T.....	294
4.4.2. El error tipo I y tipo II	297
4.4.3.Las correlaciones	299
4.4.4.La inferencia estadística	301



CAPÍTULO 4. EL PAPEL DE LA ESTADÍSTICA: CONSTRUYENDO UN IDIOMA ENTENDIBLE

La realidad es que el estudio de la salud y la enfermedad en la población (epidemiología) se hace utilizando la investigación y el lenguaje con que se expresan los resultados es la **estadística**. La estadística es un conocimiento que se ocupa de describir y analizar los datos. **Los datos son las probables respuestas a las variables** y éstas, son las preguntas a través de las cuales se descompone y se busca la contestación del problema de investigación.



Cuando se investigan grupos (por ejemplo, poblaciones) se hace necesario tener unos referentes para describir las características de tales grupos. Eso es lo que hace la **estadística descriptiva**; ofrece los referentes descriptivos (ejemplo: media, moda mediana, desviación estándar). Luego, si se desea entender algún aspecto determinado, es preciso comparar ese grupo con otro grupo; ahí interviene la **estadística analítica o comparativa**. Finalmente, quizás sea apropiado saber si un resultado en un “pequeño” grupo podrá ser extrapolarse al resto de la población; para ello surge la utilidad de la **estadística inferencial**.

A continuación se hará una somera explicación del concepto de datos y ulteriormente acerca de la estadística bajo el contexto de la epidemiología.

4.1. Los datos...

Un **dato** es la probable respuesta a una variable. A manera de ejemplo, si se pregunta cuántos casos de SIDA existen en tal país, el número obtenido es el dato. Sin embargo, existen otras formas de contestar esta pregunta (variable). Por ejemplo, en vez de decir el número, el cual no tiene un referente que permita compararlo, se puede usar la proporción (en la cual dicho número es el numerador y el denominador puede ser el total de la población y el resultado de esta relación se multiplica por 100). Así es más probable que el ser humano entienda.

Las variables pueden ser contestadas con datos de distintos tipos. Una forma de clasificar a los datos puede ser la siguiente:

1. **Datos cuantitativos**
2. **Datos cualitativos**

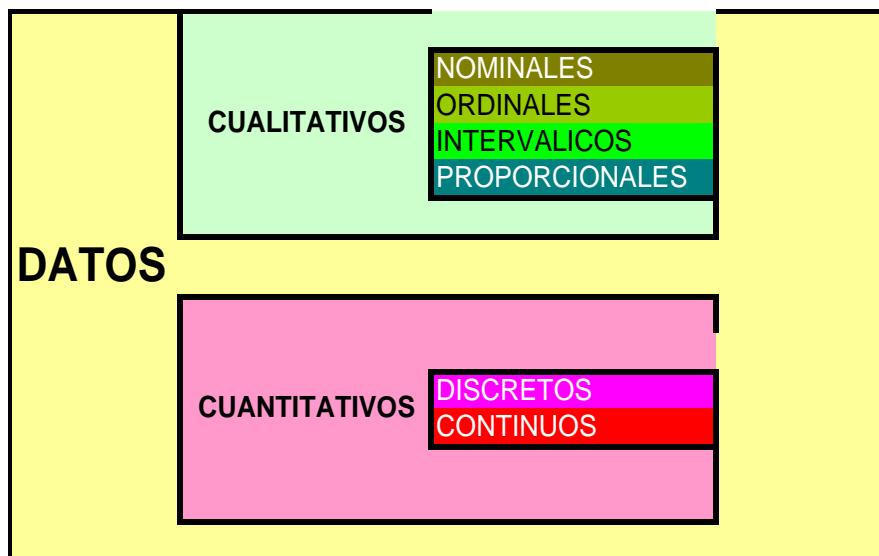
1. Los datos cuantitativos son aquellos en que la variable se contesta a través de un número a diferencia de los cualitativos que usan palabras. Los datos cuantitativos pueden subdividirse en dos:

- 1.1. Discretos
- 1.2. Continuos

Los datos discretos son aquellos cuya respuesta solo puede emplear un número entero (no permite más subdivisiones). Por ejemplo, la variable casos (personas enfermas) de tal enfermedad solo puede ser expresada con un dato como un número entero: no puede decirse que existen 7 y media personas enfermas... Otros ejemplos pueden ser la respuesta a la

variable de número de caries en un examen bucal o la cantidad de hijos...

Los datos continuos son el resultado de respuestas que pueden variar (y ofrecer más medidas, por ende, continuar) de acuerdo al instrumento de medición. Por ejemplo, la altura de una persona. El dato puede ser 175 cm, si se usa un instrumento que mide centímetros. Pero puede ser 1754 mm si se usa una escala de referencia en milímetros. Como se observa, existe la posibilidad de continuar con otras medidas.



Los datos cuantitativos tienen una enorme ventaja sobre los cualitativos y es precisamente que la estadística puede trabajarlos mejor. Ya se hablará de esto más adelante.

2. Los datos cualitativos se subdividen a su vez en:

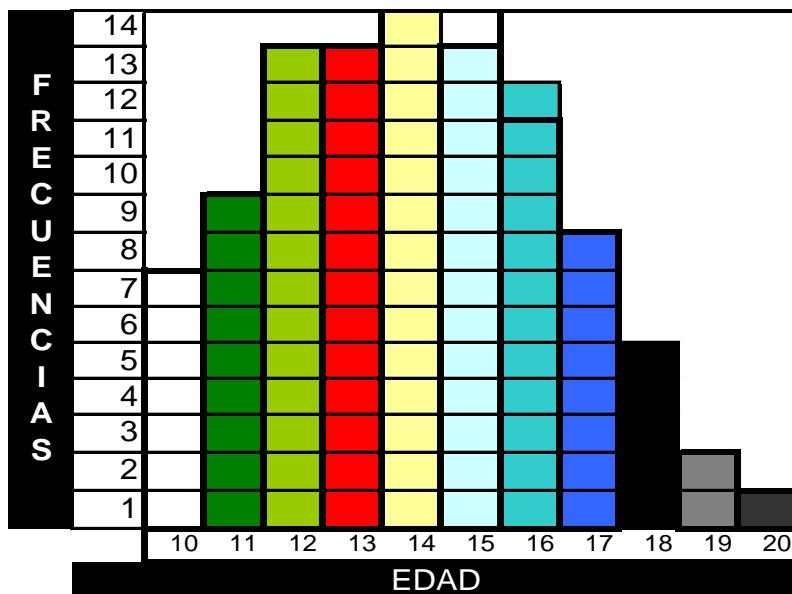
- 2.1. Nominales
- 2.2. Ordinales
- 2.3. Interválicos
- 2.4. Proporcionales

Los **datos nominales** son aquellos que usan respuestas que nombran. Por ejemplo, el color de tal cosa... o cuáles son las enfermedades más frecuentes. Algunos autores crean una subdivisión importante por la frecuencia, denominada **existencial** referida a que lo tiene o no lo tiene (existe o no). Por ejemplo, ante la variable tiene o no cáncer. El dato es

sí o no. Dado que la respuesta probable es una de dos se denominan también **dato dicotómico**, donde lo usual es que se mencionen las características del dato a través de la variable. Por ejemplo, la variable tiene o no tiene cáncer del pulmón, es una variable nominal dicotómica ya que el dato resultante es SI o NO. Por supuesto, que existen variables que producen datos nominales donde las categorías son más de dos.

Los **datos ordinales** aparte de nombrar generan un orden. Por ejemplo, ante la variable de cómo le pareció el servicio, la respuesta puede ser excelente, bueno, regular, malo o pésimo (escala de Liker). Como se observa estos datos van en orden.

Los **datos interválicos** son también ordinales, pero como complemento, permiten identificar la diferencia entre uno y otro. Por ejemplo, la variable de coeficiente intelectual, provoca datos de 100 o de 110 o de 90, donde el lector puede ubicar **orden, distancia y diferencias**. Otro ejemplo puede ser la medición de la temperatura con base a la escala Celsius. Un dato de 10 grados establece una diferencia con el de 40 grados pero exactamente no implica que el que tiene 40 grados tiene 4 veces más calor que el que tiene 10 grados. La graduación del termómetro es efectuada con un criterio diferente. Esto a diferencia de los **datos proporcionales** que sí ofrecen referentes en donde se puede medir la magnitud de la diferencia. Por ejemplo, si se menciona un peso de 50 kilos y otro de 100 kilos, se sabe cuál es el mayor, pero además, se identifica de cuánto es la diferencia (en este caso del doble del peso en kilos). En tanto, con el coeficiente intelectual se sabe que 90 es diferente de 110, pero no se sabe qué existe entre estas diferencias. Un mejor ejemplo puede ser el siguiente: Supóngase que la variable es la edad. El dato en un caso se coloca como edad en años y en otro se hace a través de la ubicación de unos rangos. Por ejemplo, por cada 10 años. Entonces en el primer caso se coloca el espacio para expresar la edad en años cumplidos. En el segundo, los datos están para que se conteste así: De 0 a 9 años. De 10 a 19 años. De 20 a 29 años. Entonces, si la edad es 13 años, así se expresará en el primer caso. En cambio en el segundo caso se colocará de 10 a 19 años. Acá no se puede saber de cuánto es la diferencia, por ejemplo, entre 21 años y 13 años; en cambio, en el caso en que el dato se exprese como años cumplidos, se sabe qué distancia existe entre uno y otro dato. Claro, una u otra información será útil de acuerdo a un contexto, pero lo que importa es que se aprecie la diferencia entre el dato interválico y el proporcional.



Los datos cuantitativos son especialmente manejables a través de la estadística. Por ejemplo, el obtener en una base de datos la edad de la población puede permitirle al investigador una serie de descripciones que pueden ser de utilidad, como la obtención de la edad media de la población o a qué edad se sitúa el 25 o el 50% de la población. Por supuesto, que puede verse con sencillez las edades rango (extremos), elementos todos que permiten que los investigadores vean más. Obviamente con una variable cualitativa nominal (datos SÍ o NO) no se pueden hacer estas descripciones. La gráfica expresa las edades de una población de adolescentes. El eje horizontal muestra las edades y el vertical las frecuencias. Cada casilla representa un adolescente con su edad.

Para resaltar lo anterior, observe la gráfica: la estadística usa para describir unos referentes (**medidas de tendencia central y medidas de dispersión**) y también las gráficas. En el ejemplo está el denominado **histograma** (gráfica esencial de la estadística) donde se expresan las edades de un grupo de adolescentes que participaron en una investigación. Ciertamente es fácil ver en ella: las edades extremos (10 y 20 años), la edad media se puede aproximar visualmente (entre 13 y 14 años), la edad que más se repite (14 años). Incluso se puede saber cuál es el total de población (suma de todas las casillas = 109).



Un aspecto clave cuando se efectúan preguntas de investigaciones es seleccionar variables que sean contestadas con datos que expresen la realidad, se entiendan y sean manejables estadísticamente. Un ejemplo donde prácticamente se combinan todos los datos es la historia clínica. Desde las preguntas abiertas de la enfermedad actual, hasta las variables cuantitativas de los signos vitales (TA, FC, FR, TA) o las interválicas de la temperatura y las cualitativas nominales dicotómicas de si tiene o no tal manifestación....

4.2. Las medidas

Una medida (o escala) es la comparación de un dato que se hace con un referente y que busca ser entendida. Lo que se mide son variables y de éstas se obtienen los datos. Recordando, las variables son preguntas a través de cuyas respuestas se puede contestar la pregunta de investigación. Por supuesto, que una variable puede ser medida con distintos referentes y será tarea del investigador el emplear como dato aquel que mejor expresa la comparación. Por ejemplo, con la variable altura de las personas es mejor usar variables que empleen datos cuantitativos (mucho mejor decir cuánto mide que decir alto o bajito) e incluso de ellos son mejores los centímetros que los milímetros.

La epidemiología usa tres tipos de medida. Ellos son:

- 4.2.1. Medidas de frecuencia
- 4.2.2. Medidas de asociación
- 4.2.3. Medidas de impacto



4.2.1. Medidas de frecuencia

Las medidas de frecuencia pueden usar distintas escalas, casi todas matemáticas y sin duda son las más empleadas en la epidemiología. Prácticamente no existe investigación donde no se expongan medidas de frecuencia y son las empleadas para la labor descriptiva. Usan:

- Los números absolutos
- Las proporciones
- Las tasas
- Las razones

A manera de ejemplo y tomando como caso la infección por VIH/SIDA, según el Boletín de Situación de Salud en Colombia 2006 se puede afirmar que:

1. El **número absoluto** de casos de VIH/ SIDA informado en Colombia en el 2006 fue de 4.824 casos.
2. Otra manera de decirlo es que la **tasa** por 100.000 habitantes de VIH/SIDA es de 11,5 (se obtiene producto de la aplicación de la regla de tres en donde se plantea que si por 42.090.502 de habitantes de Colombia existen 4.824 casos de VIH/SIDA, en 100.000 habitantes cuántos = 11,5).

Usando ahora la población: de los 42.090.502 colombianos, los hombres son 20.668.157. Esta cifra se puede presentar así:

- Por cada 100 colombianos, el 49,1 por 100 son hombres (**proporción = expresa usando el 100 como referente**). Además, el 30,7% de la población es menor de 15 años. El número absoluto es el resultado de la regla de tres que plantea que si el 100% son 42.090.502, el 30,7% serán 12.921.784.
- Otra manera de presentar esta información es la **razón** de población menor de 15 años / población mayor de 15 años = 0.4 que es lo mismo que decir que en Colombia por cada persona mayor de 15 años existen 0.4 personas menores de 15 años (resultado obtenido de la división de 12.921.784 = población menor de 15 años entre 29.168.718 = resto de población o lo que es lo mismo el resultado de restarle a 42.090.502 los 12.921.784 de los menores de 15 años).

Con la misma operación se puede sacar la razón de cuántos colombianos mayores de 15 años existen. La cifra es el resultado de dividir 29.168.718 entre 12.921.784 y es igual a 2,3. Se expresa así: Por cada menor de 15 años, existen 2,3 mayores de 15 años. Recuérdese que en las razones el resultado siempre se compara como referente con el 1.

Lo cierto es que los informes de salud presentan los resultados de distintas maneras. Existen algunos acuerdos. Por ejemplo, las estructuras de poblaciones se presentan con números absolutos o porcentajes (proporciones por 100). **Las tasas** (pregunta sobre el total de casos de la pregunta multiplicado por un múltiplo de 10) son usadas para expresar morbilidad. En recursos humanos se usan con frecuencia las razones. Por ejemplo, la razón de médicos por cada 1000 habitantes (que es de 1,3). Se obtiene dividiendo el número total de médicos (54.718 en el 2006) entre el total de colombianos. El resultado es 0.0013. Es lo mismo que decir que por cada colombiano, hay 0.0013 médicos (esa es la razón). Al colocar de referente 1000, la razón queda de 1,3 o lo que es lo mismo que en cada 1000 colombianos se encuentra 1,3 médicos. A propósito, un número absoluto utilizado es el promedio de médicos que egresan por año que en el 2006 fue de 2.393.

Muchas medidas de frecuencia son utilizadas como referentes de comparación y, por consiguiente, utilizadas como indicadores.

4.2.1.1. Concepto de indicadores: Palabras para una definición

Buscando las palabras que han de participar en una definición de indicador está el término **referente** ya que los indicadores sirven como puntos de reparo para comparar situaciones. Adicionalmente ha de estar el término **estimación** pues es evidente que a través de los indicadores se hacen aproximaciones a la realidad. Y, por supuesto, que cualquier indicador surge de las **medidas**. Con estas palabras, una propuesta de definición de indicador puede ser: **Medida que permite ser comparada como un referente que estima la realidad de una situación determinada.**

Muchas medidas de frecuencia se han convertido en **indicadoras**. Por ejemplo, las tasas de mortalidad o las prevalencias e incidencias (expresadas como proporciones) de algunas enfermedades, son ampliamente empleadas. La **INCIDENCIA** y la **PREVALENCIA** resultan ser datos que, de tener referencias previas, permiten analizar información.

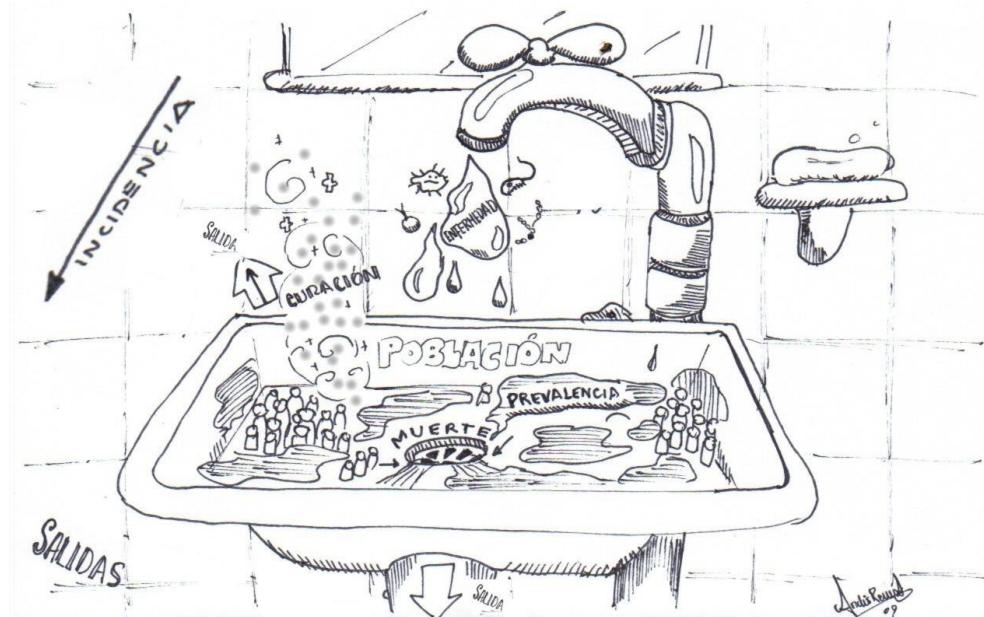


Como modelo, al medir la desnutrición crónica a través de la talla baja en Colombia en el año 2005 se encontró que el porcentaje de niños de 5 a 9 años en zona urbana fue de 9.7% (de cada 100 niños de 5 a 9 años evaluados, se encontró que 9.7 tenían una talla más baja de la esperada). En el campo la prevalencia fue de 18.5 %. De ello se puede deducir que el estado nutricional de la ciudad es mejor que el rural (apreciación que surge al comparar los dos datos). Además, y si se compara con la prevalencia en ese mismo período (en los meses y en el sitio que se midió) pero en 1995, cuando la prevalencia urbana de menores de 5 a 9 años fue de 12.5 % y la rural fue de 19,1%, se puede afirmar que el país ha mejorado. ¿Por qué? Sencillamente porque la disminución de prevalencia de niños con talla baja para su edad solo es posible porque tengan un mejor acceso a los alimentos. De hecho, y como indicador de situación de un país, la nutrición sin duda siempre es y será un buen parámetro.



Incidencia: Se refiere a los casos nuevos de una enfermedad que se presentan en una comunidad y en un período determinado. **Prevalencia:** Se refiere al total de casos de una enfermedad en un momento y en un lugar determinado.

Colimon explica que la enfermedad en una comunidad se comporta como una llave que da la entrada a las enfermedades en el lavamanos (comunidad) y que existe una salida (desagüe).



La entrada se manifiesta por la velocidad de producción de la enfermedad (incidencia). La salida por la curación o la muerte. Lo que permanece en el 'lavamanos' es lo que se llama prevalencia. La relación entre la entrada y la salida condiciona la prevalencia de la enfermedad en la comunidad. Por ejemplo, la incidencia de hipertensión no es muy alta (entrada) pero la salida sí es escasa (muy pocos se curan de hipertensión y la muerte sobreviene luego de mucho tiempo de padecerla), por ende su prevalencia será alta.

Cuando se estudia la salud y la población, las situaciones de referencia de donde se pueden obtener indicadores son:

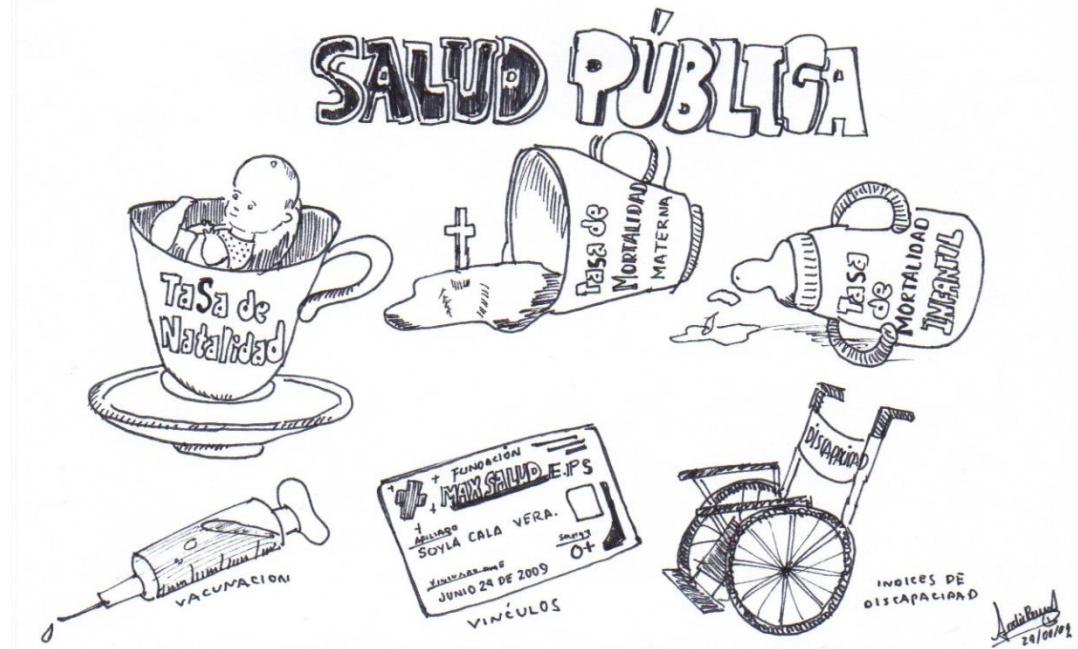
- La muerte
- La enfermedad
- La discapacidad de la población
- Los factores de riesgo de la población
- Aspectos interviniéntes en la salud poblacional tales como prevalencia de felicidad, calidad de vida, desarrollo humano.



No existe duda que la muerte o la enfermedad son un excelente referente de comparación. Por ejemplo, si en un país se mueren (o padecen) más personas de enfermedades tales como el SIDA (en África, por ejemplo) y en otro de enfermedades coronarias (USA), la comparación nos permite asegurar que las condiciones de uno u otro son diferentes. La discapacidad y los factores de riesgo sirven exactamente igual. Estos son considerados indicadores negativos (pues miden lo "malo"). Hoy están surgiendo nuevos indicadores que miden más la salud que la enfermedad. Aspectos como calidad de vida, percepción de felicidad y satisfacción social están siendo elaborados y utilizados como referentes.

Ejemplos de utilidades de los indicadores pueden ser:

1. **Identificar** necesidades en salud. Por ejemplo, la mortalidad infantil, donde resulta obvio que esta población es la que más se resiente cuando las condiciones son inadecuadas y, por ende, la que mejor responde cuando empieza a mejorar. A manera de ejemplo en el Chocó la tasa de mortalidad en el 2005 fue de 89 por 1000, en tanto que en Bogotá fue de 25,8 por 1000. ¿Qué análisis (resultado de la comparación) se puede hacer?
2. Evaluar recursos para la salud. Índice de pobreza humana.
3. Evaluar impactos: Tasa específica de fecundidad en mujeres de 15 a 19 años (por mil).
4. Monitorear: Número de personas desplazadas por violencia por año.
5. Atributos de políticas: Tasa de desempleo.
6. Negociación: Gasto nacional en salud por año en proporción del PIB.
7. Racionalización: Índice de condiciones de vida en el año.
8. **Evaluar investigaciones.**
9. **Tamizar hipótesis.**



Como ejemplos en salud pública están:

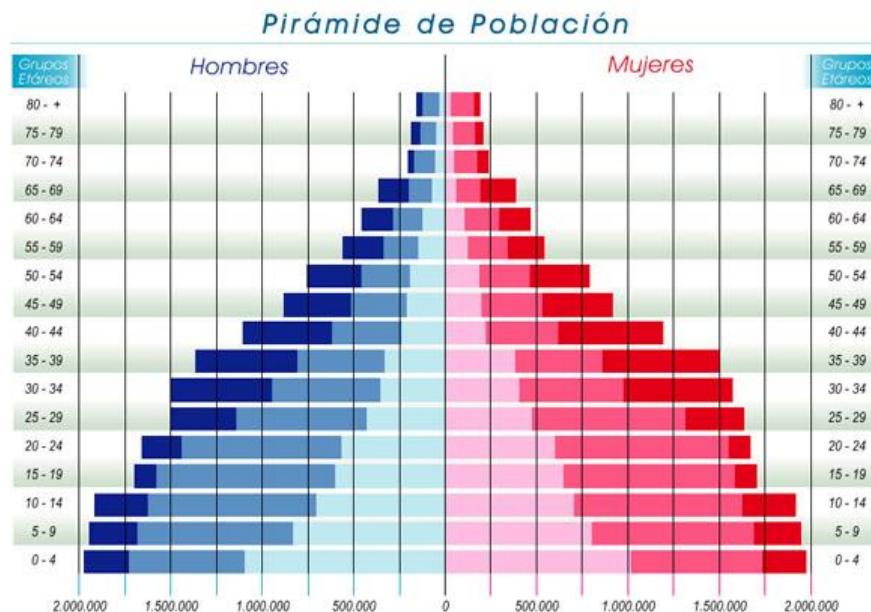
1. La mortalidad infantil que refleja la inequidad.
2. La mortalidad materna refleja el acceso a los servicios de salud.
3. Las tablas de vida (expectativa de vida y los años de vida perdidos).
4. En morbilidad se usa la letalidad o relación entre los que se mueren y los que padecen de una enfermedad obviamente en un período de tiempo.
5. El bajo peso al nacer.
6. Los análisis de discapacidad o de calidad de vida.
7. Indicadores de acceso a servicios de salud (pueden ser: cobertura de vacunación, la atención prenatal, la atención profesional, la planificación o el número de citologías).

Las fuentes de estos indicadores son diversas. El DANE, Planeación Nacional, Fedesarrollo o el Ministerio de Protección Social, son apenas unos ejemplos nacionales. La OMS, el Banco Mundial o la OPS, son entidades que aportan indicadores de distintas partes del mundo.

De todo lo anterior ha de quedar en claro el papel esencial de los indicadores en los análisis. Un buen indicador debe cumplir con unos requisitos mínimos:

- Factibilidad: que se pueda obtener.
- Oportunidad: que se tenga cuando se necesita.
- Continuidad: que tenga referentes previos que permitan la comparación.
- Costo/eficiencia: que el beneficio que da sea mayor que la inversión para obtenerlo.
- Validez: que refleje la realidad.
- Precisión: que si se vuelve a medir dé el mismo resultado.
- Sensibilidad: que detecta lo que se busca.
- Especificidad: que permite descartar.
- Relevancia: que permite ver la importancia de una situación.

Precisamente, y a manera de ejemplo, la cartilla de situación de salud en Colombia 2006 ofrece entre otros los siguientes indicadores (situaciones):



1. La pirámide poblacional de Colombia, comparada en tres momentos: 1985, 1993 y 2005.
2. Indicadores demográficos tales como población total y reparticiones por edades específicas.
3. Dinámica de población tales como tasas de fecundidad, media anual de defunciones.
4. Indicadores socioeconómicos, tales como niveles educativos o niveles de ingresos, gasto nacional de salud.

5. Vinculación al Sistema General de Seguridad Social.
6. De ofertas de servicios de salud (coberturas de servicios o recursos humanos en salud).
7. Accesos de servicios públicos domiciliarios.
8. Indicadores de mortalidad.
9. Indicadores de morbilidad.
10. De salud sexual y reproductiva.
11. De riesgos profesionales.
12. De vacunación.
13. De enfermedades de notificación obligatoria.
14. De estado nutricional.
15. De riesgo de enfermedades crónicas.
16. De salud mental y consumo de sustancias psicoactivas.

4.2.2. Medidas de asociación

	Cáncer	No cáncer	
Si fuma	80	20	100
No fuma	40	160	200
120		180	
ASOCIACION			OR = 16

Las medidas de asociación buscan las relaciones que pueden existir entre la causa y el efecto. Todas son el resultado del uso de la tabla de comparación o de tabla de 2X2. Son dos. Unas son las medidas de diferencia que comparan las frecuencias de los expuestos con los no expuestos. Otras que obtienen razones entre los expuestos y los no expuestos. Las dos escalas empleadas son el RR o riesgo relativo ($a/(a+b) \div c/(c+d)$) y el OR o razones de ventaja (ad/bc).

4.2.3. Medidas de impacto

Son prácticamente específicas de la práctica de la epidemiología. Una se llama el riesgo atribuible (o fracción etiológica) y expresa la responsabilidad de la exposición en la aparición del evento adverso. Por ejemplo, cuál es el riesgo atribuible del cigarrillo en el cáncer del pulmón.

La segunda medida es la denominada fracción prevenible donde se mide el impacto que tiene una intervención protectora, Por ejemplo, cuáles serán los casos que se presentarán (de ahí surgen cuántos casos se previenen) en presencia de una acción protectora.

Si bien no se va ampliar cómo se obtiene, vale la pena ampliar el concepto para que se sepa cómo se usan estas medidas. Hoy sabemos que la exposición del cigarrillo es el factor causal más importante en la génesis del cáncer del pulmón. A través del **RIESGO ATRIBUIBLE** se puede calcular el porcentaje de responsabilidad del cigarrillo, el que por cierto está en cerca de un 92%. Esta cifra se obtuvo de la aplicación del RIESGO ATRIBUIBLE también denominado **FRACCION ETIOLOGICA**. Pero surge la pregunta: ¿y si las personas dejaran de fumar?... ¿Cuántos casos de cáncer se podrían prevenir? Ese resultado es lo que se conoce como **FRACCION PREVENIBLE** que es la otra medida de impacto.

4.3. La estadística descriptiva

La estadística, en una simplista interpretación, no es otra cosa que un proceso de análisis de información, que busca hacernos entender los datos.



La Estadística se ocupa de los métodos y procedimientos para recoger, clasificar, resumir, hallar regularidades y analizar los datos, siempre y cuando la variabilidad e incertidumbre sean una causa intrínseca de los mismos; así como de realizar inferencias a partir de ellos, con la finalidad de ayudar a la toma de decisiones y, en su caso, formular predicciones.

Buscando ser didácticos, el ser humano aprende a través de la descripción y de la comparación. Con base en estos dos procedimientos anhela predecir. La investigación es la forma como el ser humano describe y compara. La estadística ofrece referentes para aprender a describir y comparar en las investigaciones. Para describir, la estadística utiliza una serie de presentaciones y de aspectos referenciales (medidas de tendencia central y de dispersión). La principal forma de presentar la información es la tabla de frecuencia que, con mucho, constituye la principal forma de presentar información en la estadística descriptiva (no la única y, de hecho, las gráficas le hacen algo de competencia).

4.3.1. La tabla de frecuencia

Para describir, la estadística toma los datos obtenidos y los agrupa siguiendo algún criterio. De esa manera el investigador observa más. A manera de ejemplo, obsérvese la siguiente **base de datos** que muestra las notas de un curso con identificación y sexo:

Nº	IDENT	SEXO	NOTA	Nº	IDENT	SEXO	NOTA	Nº	IDENT	SEXO	NOTA	Nº	IDENT	SEXO	NOTA
1	ACE	m	2,4	51	DIAZ	h	3,8	101	MAR	m	3,8	151	ROJ	h	3
2	ACO	1	3,2	52	DIAZ	m	3,6	102	MAR	2	2	152	ROJ	h	2,8
3	AMA	m	3	53	DIAZ	h	3,4	103	MAR	3	153	ROJ	m	2,4	
4	ARE	m	3,4	54	DIAZ	m	3,4	104	MAR	h	3,2	154	ROM	m	3,4
5	ARIZ	m	3,8	55	DUA	h	3	105	MED	m	2,8	155	ROM	h	2,2
6	ARIZ	h	4,6	56	DUG	m	3,2	106	MED	h	3,2	156	ROM	h	2,8
7	ARIZ	m	2,8	57	ECH	m	2,6	107	MED	m	3,4	157	ROM	m	2,2
8	ARR	h	3,6	58	ECH	m	3,4	108	MEN	m	3,2	158	RON	h	2,8
9	ART	h	2,8	59	ESL	m	3,4	109	MEN	m	3	159	RON	m	3
10	BALI	m	3,8	60	ESP	h	3	110	MOL	h	3,6	160	ROZ	m	3,4
11	BAR	m	2,8	61	EST	m	3,6	111	MOR	m	3,4	161	SAN	m	3,6
12	BAR	m	2,8	62	FALL	h	3,4	112	MOR	h	4,2	162	SAN	m	3,2
13	BAR	h	2,8	63	FAR	m	3,2	113	MOY	m	1,8	163	SAN	m	3,8
14	BEL	m	3,6	64	FLO	m	3,2	114	NEG	h	2,6	164	SAN	h	3,6
15	BEL	h	3	65	FLO	m	2,4	115	NEV	m	2,8	165	SAN	h	2,4
16	BEL	m	2,8	66	FON	h	3,4	116	NOV	m	2,4	166	SOS	m	3
17	BEL	m	3	67	FUE	h	3,2	117	OLA	m	3	167	SOT	h	3
18	BEN	h	2,8	68	FUE	m	3	118	OLA	m	3,4	168	TOB	h	2,6
19	BEN	m	3	69	GAL	h	3,2	119	ORT	h	3,4	169	TOR	m	2,8
20	BON	m	3,2	70	GAR	h	3,6	120	PAN	m	3,4	170	TOU	m	3,2
21	BRU	h	3	71	GAR	m	3,6	121	PAR	m	2,8	171	TRU	m	3,2
22	BUIT	m	3,2	72	GAR	h	3,2	122	PAR	m	3,4	172	URIE	h	2,8
23	BUIT	h	2,6	73	GAR	h	3,2	123	PED	h	2,4	173	URR	h	3,2
24	CAIC	h	2,8	74	GAR	m	2,4	124	PEN	m	3,4	174	VAL	h	3,4
25	CAJ	m	2,8	75	GAR	m	4,2	125	PER	h	3	175	VAL	m	3,6
26	CAL	h	3,4	76	GAR	h	4,2	126	PER	m	3,4	176	VAL	m	3
27	CAL	h	3,8	77	GOD	m	3,4	127	PER	h	2,4	177	VAR	m	2,6
28	CAN	m	4	78	GON	m	3	128	PER	m	3,4	178	VAR	m	2,6
29	CAR	h	2,6	79	GON	h	2,6	129	PINE	m	3,2	179	VEL	m	3
30	CAS	h	2,6	80	GON	h	3,4	130	PIN	m	3,6	180	VICI	h	3,2
31	CAS	h	3,6	81	GON	m	3	131	POR	h	3,8	181	VILL	m	3,2
32	CAS	h	2,8	82	GON	m	2,2	132	PUE	m	3,4	182	ZAM	m	3,6
33	CAS	m	3,6	83	GON	m	3,6	133	PUL	m	2,2				
34	CAS	h	3,2	84	GON	m	3,2	134	QUE	m	3,6				
35	CAS	h	3,6	85	GON	m	3,2	135	QUIN	m	3				
36	CAS	h	2,6	86	GUT	h	3,8	136	QUIN	m	3,6				
37	CIFU	m	2,8	87	HER	m	3,6	137	RAM	h	3,2				
38	COR	m	3,4	88	HER	h	3,4	138	RAM	m	3,2				
39	COR	m	2,8	89	HOY	h	3,8	139	RAM	m	3,8				
40	COR	m	3,4	90	JAIM	m	3	140	RAM	h	2,6				
41	COR	m	3,4	91	KUA	h	3	141	REIN	m	2,4				
42	CUA	m	2,8	92	LON	m	3,6	142	REY	m	2,2				
43	CUB	m	2,4	93	LOP	m	3	143	RING	m	3				
44	CUE	m	2,8	94	LUG	h	3,6	144	RING	h	3,4				
45	CUE	m	3,2	95	LUG	h	3,2	145	RIVE	m	2,8				
46	CUIN	m	3	96	LUN	m	3,8	146	ROG	m	2,2				
47	DAN	m	3	97	MAR	m	3,8	147	ROD	m	3,4				
48	DEL	m	3,2	98	MAR	h	2,6	148	ROD	m	3,2				
49	DEL	h	2,4	99	MAR	m	3,4	149	ROD	h	3,6				
50	DIAZ	h	3,4	100	MAR	h	2,8	150	ROD	m	3,8				

En una base de datos se trasciben los datos de las variables. En este caso la base está constituida por la columna de enumeración de los sujetos, por la identificación, por el género y la calificación que obtuvieron. Si la base se ordenara, una buena cantidad de información presente se pondría al alcance de los ojos.

Como se ve, se trata de una lista de calificaciones cuya escala va del 0 al 5, de determinada materia. Al exponerla sin orden, la capacidad de análisis se disminuye. Ahora observe:

Clases	Frecuencias	Frecuencia acumulada	% de Frecuencia	% de Frecuencia acumulada	Marca de clase	Marca con frecuencia
De 1,5 a 1,9	1	1	1	1	1,7	1,7
De 2,0 a 2,4	18	19	10	10	2,2	39,6
De 2,5 a 2,9	36	55	20	30	2,7	97,2
De 3,0 3,4	87	142	48	78	3,2	278,4
De 3,5 a 3,9	35	177	19	97	3,7	129,5
De 4,0 a 4,4	4	181	2	99	4,2	16,8
De 4,5 a 4,9	1	182	1	100	4,7	4,7
Totales	182		100			

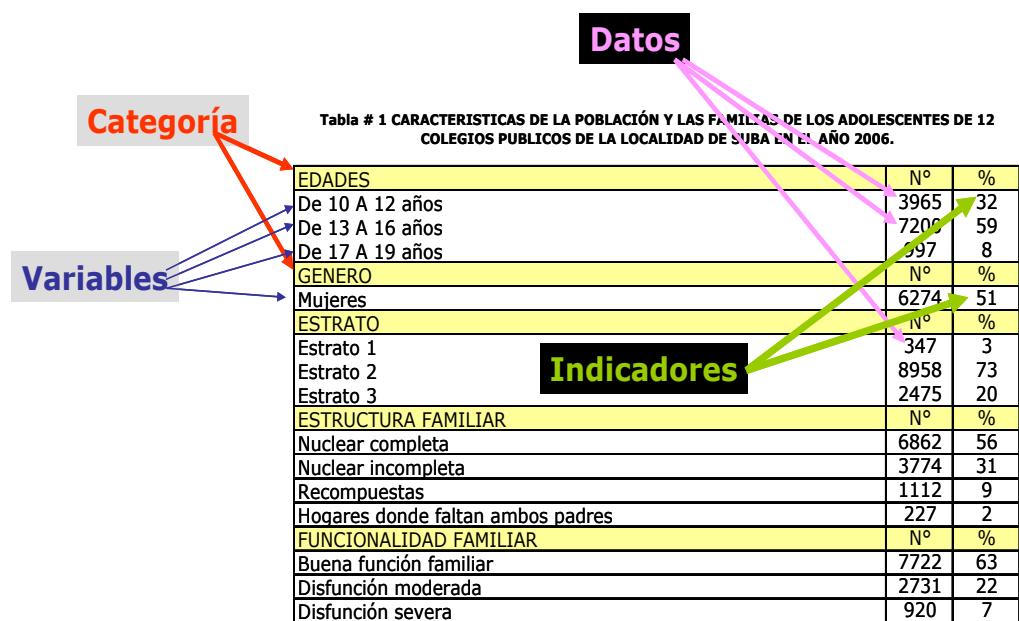
La tabla de frecuencias que se presenta en los textos tiene tan solo tres columnas: los criterios (clases), las frecuencias y el porcentaje de frecuencias. En este ejemplo se muestra una tabla de frecuencias completa.

La tabla de frecuencia precisamente **ordena** los datos (cuando se ordena se ve más). En primer lugar, los agrupa siguiendo un criterio. Esto se denomina **clases**. En esta caso van de menor a mayor (claro... el criterio de ordenamiento debe ser entendible para el organizador). Cada clase tiene límites y como ley, los límites son excluyentes de los de las clases que están al lado. La razón es obvia: un dato no puede estar situado en dos clases diferentes. Las **filas** en la tabla expresan las clases en que se ordenaron los datos. La **columna** de frecuencia muestra el número de veces que se repiten los datos de cada clase. En el ejemplo, los datos de la clase con calificaciones (datos) de 3,0 a 3,9 son los que se repiten un mayor número de veces (15). Sin embargo, ese número de veces puede ser expresado usando un referente que permita compararlo con los otros datos. Eso se consigue usando el **porcentaje de frecuencia** (tercera columna) donde la frecuencia de 15 corresponde al 50% del total de datos. Ahora se sabe que la mitad de los datos de esta tabla obtuvieron una calificación entre 3,0 y 3,9. Realmente estas tres columnas constituyen la base o esencia de la tabla descriptiva. No obstante, a la tabla se le pueden incorporar otras columnas que permiten ver otras características. Por ejemplo, la **frecuencia acumulada** en donde al número de datos de la clase se le agrega (se suma) el número de datos de la clase siguiente. A manera de ejemplo, así como la clase de 2,0 a 2,9 tiene 13 datos, para la clase siguiente (de 3,0 a 3,9) que tiene 15, la frecuencia acumulada es el resultado de la $13 + 15$, es decir 28. La columna **de porcentaje de frecuencia acumulada** expresa esto mismo pero usando los porcentajes. La columna de **marca de clase** muestra el dato punto medio de la clase.

Tabla N° 1 Frecuencias de calificaciones en un curso de investigación en una facultad de medicina en el año 2003 en la ciudad de Bogotá

Clases	Frecuencias	Frecuencia acumulada	% de Frecuencia	% de Frecuencia acumulada	Marca de clase	Marca con frecuencia
De 1,5 a 1,9	1	1	1	1	1,7	1,7
De 2,0 a 2,4	18	19	10	10	2,2	39,6
De 2,5 a 2,9	36	55	20	30	2,7	97,2
De 3,0 3,4	87	142	48	78	3,2	278,4
De 3,5 a 3,9	35	177	19	97	3,7	129,5
De 4,0 a 4,4	4	181	2	99	4,2	16,8
De 4,5 a 4,9	1	182	1	100	4,7	4,7
Totales	182		100			

No existe duda de la importancia de desarrollar la competencia de construir tablas de frecuencia. Es probablemente, la mejor forma de describir algo. Cuando se haga, recuerde que cualquiera que la vea debe entenderla. Por tanto, debe estar muy bien titulada con número, mención de lo que describe y la fuente de donde se obtuvo. Sin embargo, la tabla de frecuencia que se presenta no utiliza todas las columnas. Usa únicamente las clases, las frecuencias y el porcentaje de frecuencia.



Recuerde toda la información que le ofrece la tabla de frecuencia convencional. En el ejemplo están descritas las características demográficas y familiares de los adolescentes de 12 colegios públicos de la localidad de Suba.

4.3.2. Las gráficas

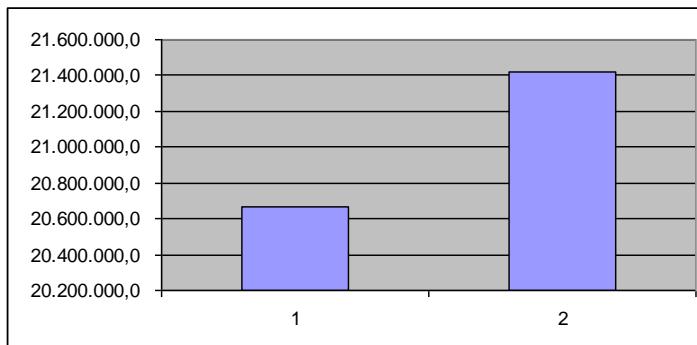
'Una imagen vale más que mil palabras'. Este adagio tiene una profunda implicación. El lenguaje se ha construido para que las personas se entiendan. Pues así como se ha construido el lenguaje de las palabras, se ha construido el lenguaje de los números y también se ha construido el lenguaje de imágenes. A veces, las imágenes transmiten mejor una idea que las palabras. Precisamente eso es lo que significan las gráficas: imágenes que muestran frecuencias de eventos y son otro de los lenguajes que se pueden usar para hacer entender eventos.

Por ejemplo, para expresar de cuánto y de qué género es la población colombiana se puede usar:

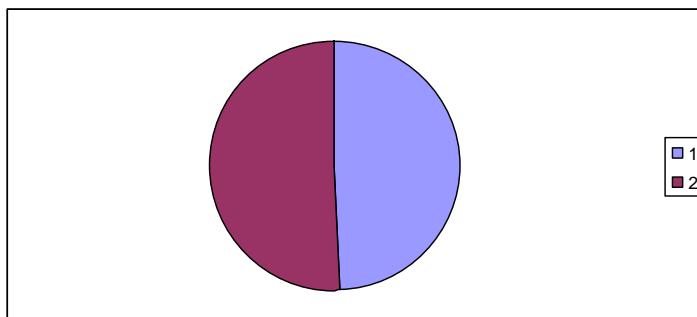
Una tabla:

Población	Número	%
Hombres	20.668.157,0	49,1
Mujeres	21.422.345,0	50,9
Total	42.090.502,0	100,0

Una gráfica de barras expresaría lo siguiente:



Y un pastel (o sector o porción) expresa lo siguiente:

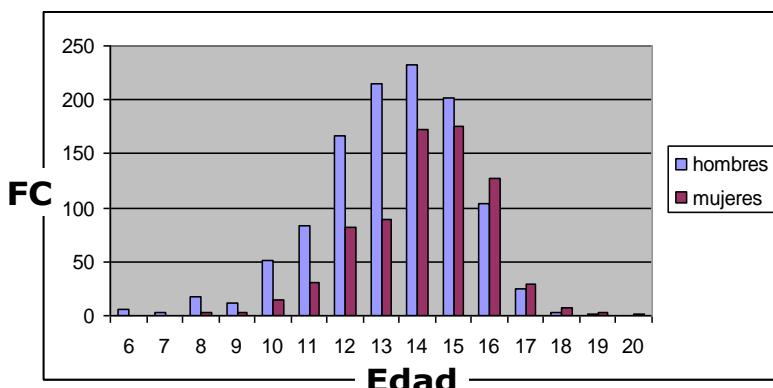


La tabla sin duda presenta la información más completa (por ello, es la principal forma como se describe la información en las investigaciones). El diagrama de barras (en este caso) tuvo que ser referido a la diferencia entre los hombres y las mujeres para poder mostrar algo. Y el pastel (o círculo) muestra bastante bien la totalidad (círculo completo) junto con la porción que cada dato ocupa. Estará de acuerdo que en este caso, la tabla y el círculo son la mejor manera de expresar esta información.

EDAD	hombres	mujeres
6	6	0
7	3	0
8	18	3
9	11	3
10	51	14
11	84	31
12	166	82
13	215	89
14	233	172
15	202	176
16	104	127
17	25	29
18	3	7
19	2	3
20	0	2
Total	1123	738

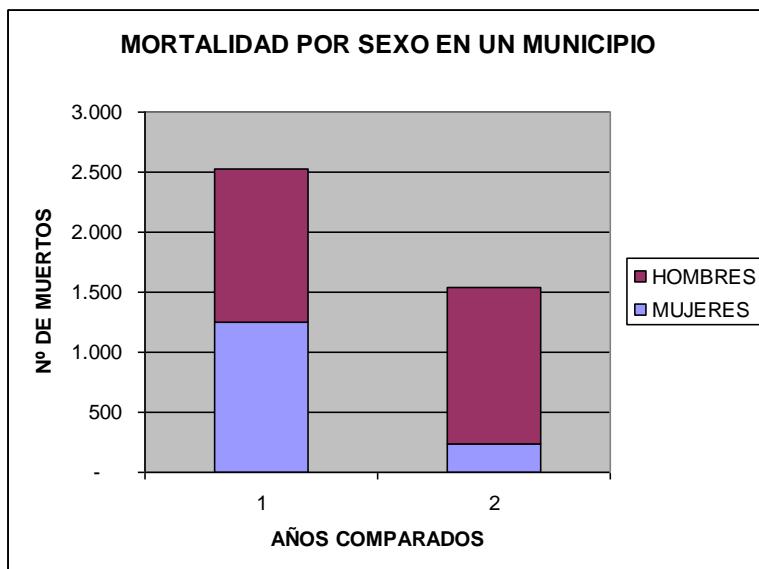
Para no dar la idea de que la gráfica de barras no es útil para mostrar información, observe este ejemplo: En un estudio sobre 12 302 adolescentes (entre 10 y 19 años) de colegios públicos se les preguntó a qué edad tuvieron la primera relación sexual. La pregunta fue contestada por 1 871 jóvenes (que afirmaron haber tenido relaciones sexuales) y se estratificó por género. La tabla y la gráfica de barras se presentan a continuación:

Y la gráfica de barras:



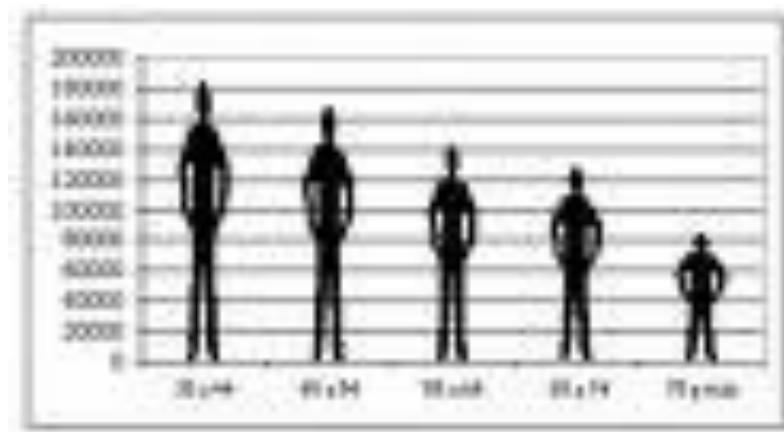
En el eje de la Y está el número de casos que afirman haber tenido relaciones sexuales (50, 100 etc.). En el eje de las X están las edades y en azul los hombres comparados con el rojo que son las mujeres. Por cierto... ¡alarmante lo expresado!... Según esta tabla el inicio de la actividad sexual se hace a los 6 años en el hombre y 8 en la mujer: ¿será que han sido abusados sexualmente?

Otra gráfica muy empleada son las barras compuestas:

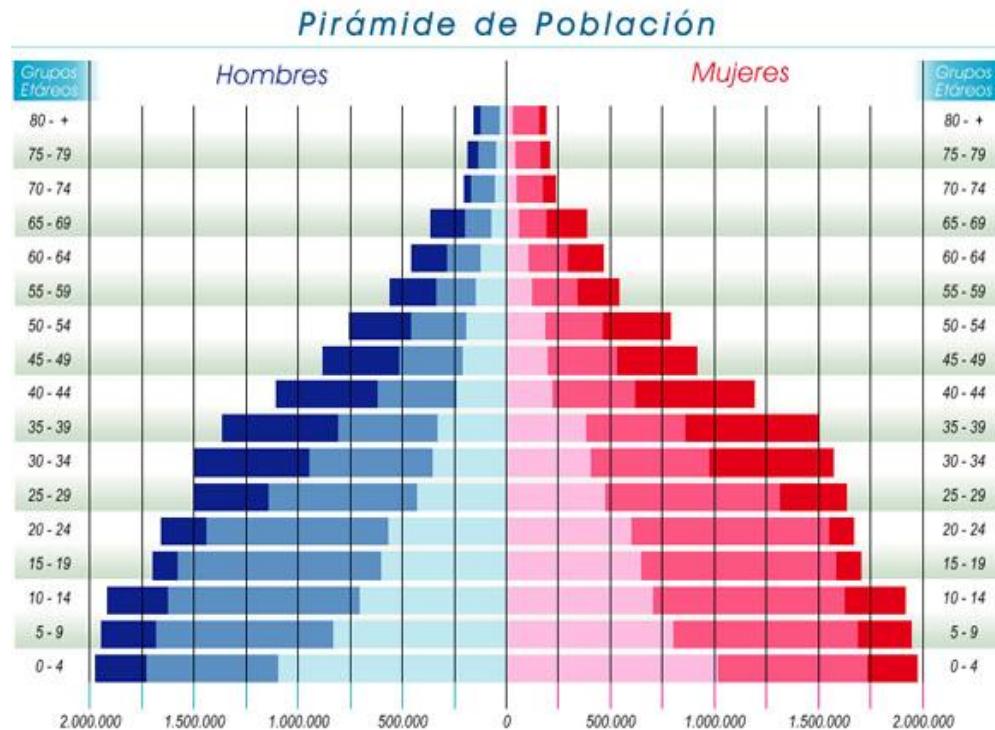


Permite comparar un mismo evento (mortalidad) con una misma variable (género). El ejemplo muestra cómo al comparar un año con otro, si bien se disminuyeron los casos de muerte en el año 2, la proporción de hombres se incrementó (o la de mujeres se disminuyó).

Otra gráfica empleada son los pictogramas. En ella se usa un dibujo representativo.



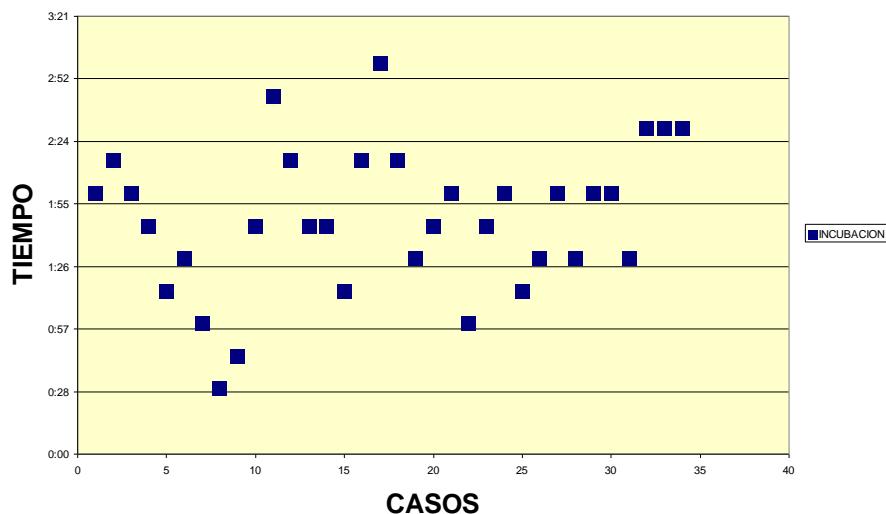
Las pirámides de población se usan mucho en epidemiología.



El eje Y expresa las edades agrupadas y el X la cantidad de población. Con los colores de las barras se quiere comparar años de censos previos para poder ver el crecimiento de la población. Así, por ejemplo, el azul representa los hombres y el azul tenue fue del censo de 1993, el tono siguiente, y más fuerte, del año 2000 y el azul oscuro del 2006 (por ejemplo). Lo propio con los tonos rojos que representan a la mujer.

Otra gráfica son los diagramas de dispersión. Por ejemplo, observe la gráfica que sigue a continuación producto del análisis en torno a la aparición de la enfermedad producida por la ingesta de alimentos dañados en donde se relaciona el número de casos con el factor tiempo, donde la hora 00.00 es luego de comer.

PERIODO DE INCUBACION



Entre el tiempo 00:00 y 0:28 minutos se presentó un caso. Entre 00:28 y 00:57 minutos se presentaron tres casos. Como se observa entre la hora 1:26 y la hora 1:55 se presentaron once casos (también entre la hora 1:55 y 2:24 se presentaron 11 casos). ¿Qué implica? Lo más probable es que permita plantear los tiempos de incubación de esta intoxicación.

Si bien en este texto no se detendrá en la forma como se hacen los diagramas, es importante insistir que los diagramas en muchas ocasiones permiten que los lectores entiendan mejor un informe, convirtiéndose en la mejor forma de presentarla. (Más adelante se explicarán dos diagramas de mayor uso: el histograma y el polígono de frecuencias).

Los croquis (o mapas) se emplean con frecuencia:



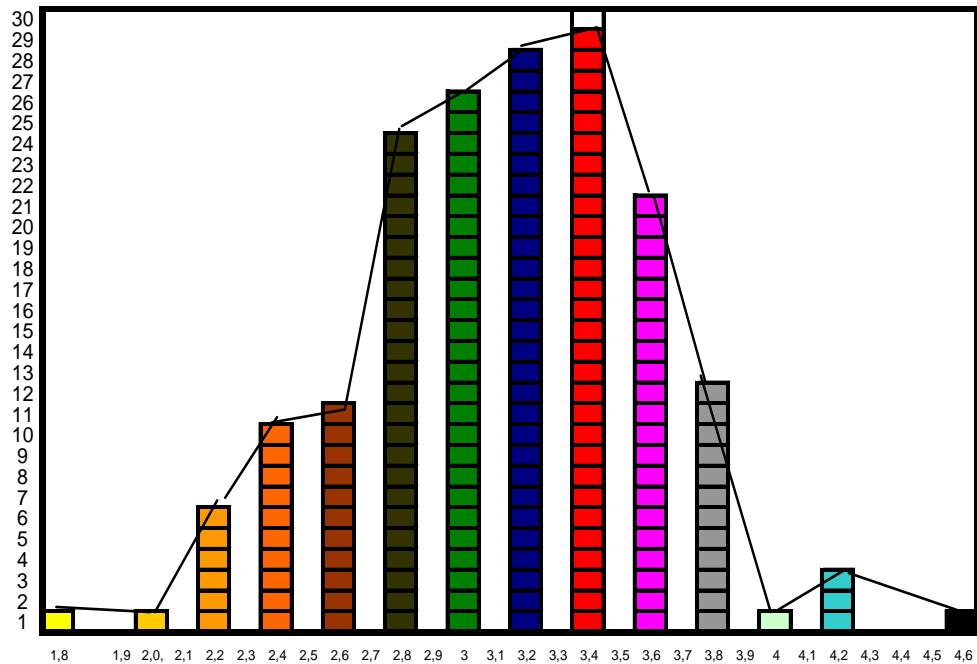
Este croquis muestra cómo Snow expuso a través de puntos los lugares donde aparecían los casos de cólera en la epidemia de 1849. Sin duda, no existe una mejor forma de exponer que existe una relación entre los casos y la ubicación geográfica.



Lo cierto es que hoy, a través de los programas de los computadores, la realización de las gráficas es un asunto menor. Prácticamente el computador alimentado con los datos las construye.

4.3.3. Las medidas de tendencia central

El ser humano para entender ha de usar la comparación. A través de ella, comprende. Para comparar, se han de tener **referentes** que permitan hacer las apreciaciones pertinentes. La estadística descriptiva propone que los resultados de las investigaciones usen unos referentes en su presentación. Al hacer que todas las investigaciones usen los mismos referentes, las comparaciones serán más fáciles. A continuación, y a través de un ejemplo, se van a explicar los dos grupos de referentes usados por la estadística descriptiva, a saber: medidas de tendencia central y de dispersión.

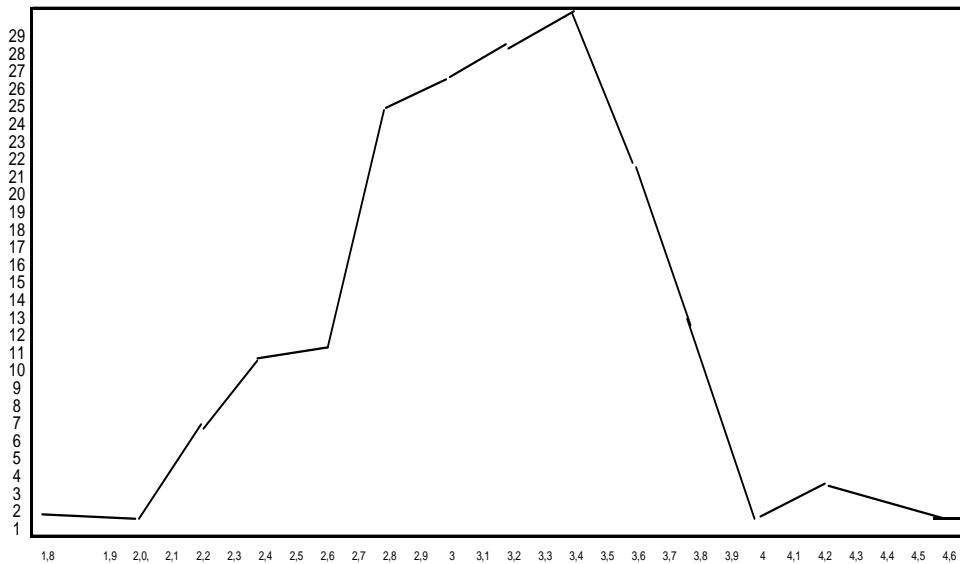


El **histograma** permite ver muchas cosas... las notas extremo (alta y baja), la que más se repite y, casi permite identificar, la nota promedio... Por supuesto que es el histograma la base de la gráfica de barras. Realmente es una maravilla de gráfica... ¿cierto?

Observe el **histograma**: Se trata de una presentación de calificaciones de un curso de estudiantes de medicina. En el cuadro, en sentido vertical (Y) se describen las frecuencias de 1 a 30. Recuerde que la frecuencia corresponde al número de veces que se repite algo. En el sentido horizontal (X), se describen las calificaciones. En este caso la primera calificación que se registra será la menor que obtuvieron los estudiantes (1.8) y se registran hasta la mayor obtenida (4.6). Luego se coloca cada nota obtenida por los estudiantes en las casillas (cuadritos) correspondientes a las calificaciones. A manera de ejemplo: un estudiante obtuvo 1.8, por ende será solo una casilla, ninguno obtuvo 1.9, uno obtuvo 2.0, seis estudiantes obtuvieron 2.2 y así sucesivamente. Evidentemente esta gráfica permite ver:

1. Cuáles son las notas límites (la más baja = 1.8 y la más alta = 4.6). Eso se conoce como **rangos** (y es una medida de dispersión que expresa qué tan extendidos están los datos).
2. Muestra cuál es la nota que más se repite (3.4 que se repite 29 veces).
3. Al sumar todas las casillas permite ver el total de datos. En este caso son 182 calificaciones o datos.

Además, si se unen las frecuencias de los datos a través de una línea en el último dato de cada frecuencia, se forma otra gráfica denominada **POLÍGONO DE FRECUENCIA**.

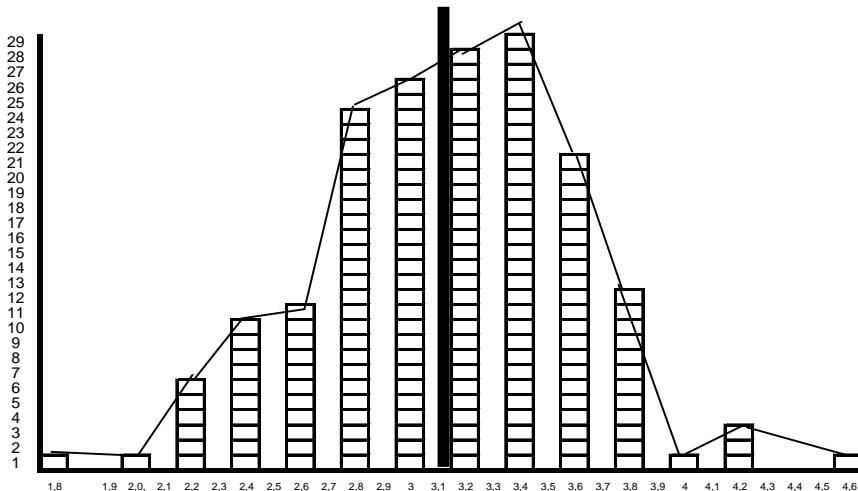


El polígono de frecuencias se construye a partir de los histogramas.

Tanto el histograma como su consecuencia, que es el polígono de frecuencia, sirven para hacer entender cómo se comportan las frecuencias de las variables estudiadas. Pero, falta más. Lo que se buscan son referentes en torno a la distribución de datos, referentes que sirvan para comparar el comportamiento de este grupo con otros. Para ello, la estadística descriptiva ha desarrollado una serie de medidas que van a comparar tomando el centro como referente. Se denominan las medidas de tendencia central. A continuación se expondrán algunas.

4.3.3.1. La media o promedio

La media aritmética es el dato que se ubica justo en el centro entre el rango (o, lo que es lo mismo, el centro de los dos límites del rango). En el caso colocado de ejemplo (notas o calificaciones), se trata de averiguar cuál es la nota situada en el medio, referido el término medio a la nota en la distancia en donde los límites de los rangos es similar. Para ello, se suman todos los datos (notas) y se dividen por el número de datos. La línea gruesa del centro pone en evidencia la calificación media (3,1).



Realmente es importante que se entienda a la media pues con mucho va a ser la medida que más se utiliza como referente. Se debe recordar que la estadística y sus referentes sólo sirven para poder comparar el comportamiento de un dato referido al grupo o de un grupo referido a otro. En un sentido eminentemente práctico y como ejemplo, el uso de la media surge cuando se quiere comparar el comportamiento de este curso con otro. El comparar las medias de los dos grupos ofrece un buen referente que permite entender. Otra posibilidad es que se quiera comparar una calificación con respecto al grupo. Entonces comparar la calificación con la media del grupo permite tener un buen referente para opinar. Claro... en ese caso, se debe sacar la nota que se pretende comparar y sacar la media del grupo. Ahí ciertamente quedan los dos datos para comparar.



Son muchísimos los ejemplos de la utilidad de la media en epidemiología e investigación. Por ejemplo:

1. Cuando quiero saber si puedo comparar dos poblaciones, entonces veo si la edad media de un grupo y otro son similares. Para ello, y en cada grupo, sumo las edades de cada participante, y lo divido por el número de participantes, obteniendo la edad media de cada grupo.
2. Otro ejemplo puede ser el cálculo de sobrevida de una enfermedad mortal, por ejemplo el cáncer. Para ello se toman historias de personas que han padecido dicho cáncer, se observa cuándo se diagnosticaron y se determina el tiempo (en años, por ejemplo) de sobrevida a partir de dicho diagnóstico de cada caso. Luego se suman todos esos años y se dividen por el número de pacientes (o casos) estudiados y se obtiene una sobrevida media de 5 años (por ejemplo). Esto permitirá ante un caso nuevo, emitir un pronóstico de sobrevida.
3. Otro más, cuando quiero ver el efecto de un medicamento, por ejemplo para controlar la tensión arterial. Se suministra en hipertensos y se observa en cuánto tiempo controla las cifras tensionales. Se suman todos los tiempos transcurridos y se dividen por el número de casos, así se obtiene el tiempo medio de cuándo empieza la acción.

Con mucho, la media aritmética (porque existen otras medias tales como la geométrica) es el referente de medidas centrales más empleado. Por ejemplo, cuando se estudiaron los antirretrovirales en pacientes con SIDA. Ya se sabía (a partir de varios estudios) que la mortalidad de las personas con SIDA a un año tenía una media de un 80% (de cada 100, 80 estaban muertos al año, obteniéndose este 80% como dato medio de varios estudios). Para ver la efectividad de los antirretrovirales se tomaron dos grupos de pacientes con SIDA (cada grupo lo más parecido al otro). A los casos de uno de dichos grupos se le administró antirretrovirales y al otro placebos y luego se compararon las medias de mortalidad (cuántos se mueren en promedio) de ambos grupos. Lo que se espera (y que fue lo que se encontró) es que en los que recibieron los antirretrovirales el número de muertes al año se disminuyó dramáticamente, por ejemplo, descendió a un 15% (nueva media de mortalidad de este grupo), en tanto que los que recibieron placebos se mantuvo en el 80%. Claro... la efectividad de los antirretrovirales quedó demostrada. Muchas veces en epidemiología se usará la media de cualquier evento o situación como referente. ¡No lo olvide!



De hecho, la media puede ser usada para probar hipótesis (y de hecho es una medida ampliamente utilizada para ello). Siguiendo con el ejemplo, la mortalidad media de los estudios revisados de pacientes con SIDA a un año era del 80%. Entonces se dice que A es igual a 80%. Luego, se aplicaron antirretrovirales a los pacientes. Con ello la mortalidad a un año bajo en su media a 15%. Esto representa B. Resulta obvia la diferencia clínica, pero falta indagar si esta diferencia se sostiene estadísticamente. Entonces para expresar que los antirretrovirales sirven en torno a la mortalidad, la media de mortalidad usando antirretrovirales debe ser diferente a la A (es decir, A es diferente a B). Si surgieran unas medias parecidas se diría que A es igual a B y, por ende, lo que se está usando no sirve (no modifica). Por supuesto que en este ejemplo se ha de usar una prueba estadística de hipótesis, asunto que se revisará más adelante.

4.3.3.2. La moda

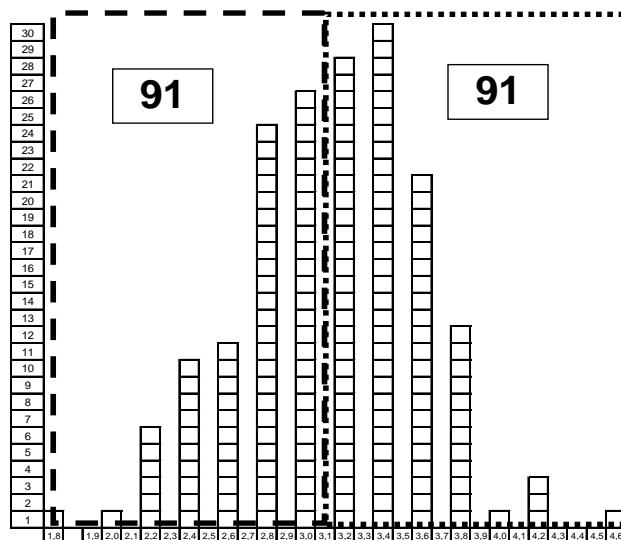
Cuando se estudia un comportamiento en una población, puede haber alguno (casi siempre lo hay) que se repita más que los demás. Eso es lo que se denomina la moda. En el ejemplo de las calificaciones, el 3,4 es la nota que más se repite.



En la salud, una de las utilidades de la moda está en los perfiles epidemiológicos. La enfermedad que más se repite (moda) se identifica con facilidad en los sistemas de vigilancia o en los informes de atención (RIPS, etc.). Conocer la moda puede ser útil. Por ejemplo, para definir capacitaciones del recurso humano que trabaja en salud (se preparara más al recurso humano en lo más frecuente). Así mismo, se podría obtener la moda de la manifestación más frecuente de determinada enfermedad. Para ello se toman varios casos de una misma enfermedad y se obtienen las frecuencias de las manifestaciones, se comparan y se mira cuál es la más frecuente.

4.3.3.3. La mediana

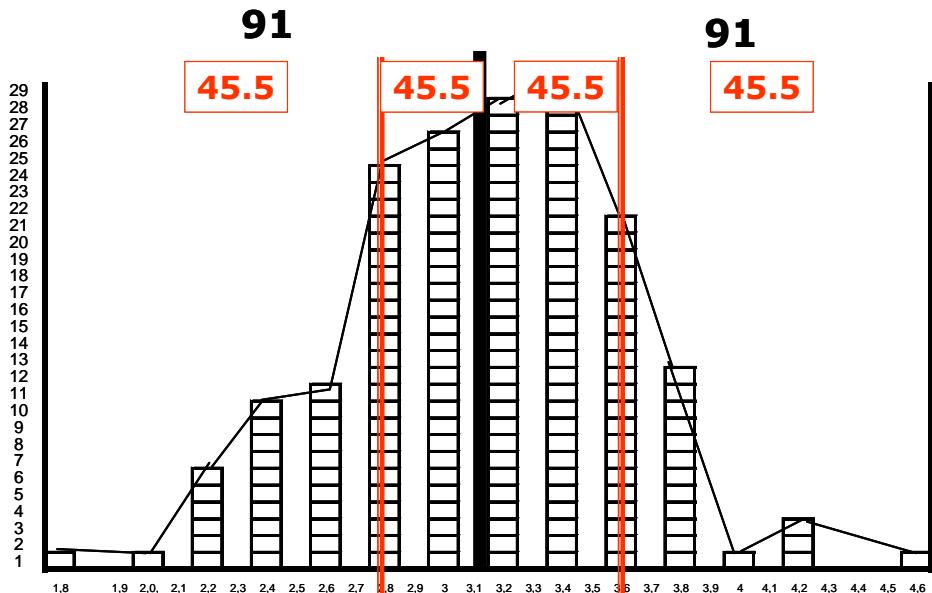
La mediana es una medida de referenciación de la población que usa el dato que divide a la población en dos mitades. Para poder verlo gráficamente y volviendo al ejemplo de las calificaciones, en un total de 182 datos (población total = 182), la mediana viene a ser la nota en donde la población se dividen en dos mitades iguales (por el número de datos o lo que es lo mismo, 92 datos a cada lado), siendo en este caso el dato (calificación) de 3,2.



Recuerde, la mediana divide la población en dos mitades a diferencia de la media que divide los datos en dos mitades. Ambas se expresan a través del dato, pero reflejan dos aspectos diferentes. A pesar de ello, muchas veces coincide con el dato, no es de sorprenderse cuando esto no sucede.



Retomando el ejemplo de la sobrevida del cáncer, en donde la media fue de cinco años, otra pregunta puede ser en qué momento la mitad de la población está muerta (y, por ende, la otra mitad viva). Para ello, supongamos que el total de población (casos) fue de 50. En el año 4,5 estaban vivos 25 y ya habían fallecido los otros 25, entonces ese año (4,5) fue el año en que dividió a la sobrevida en dos).

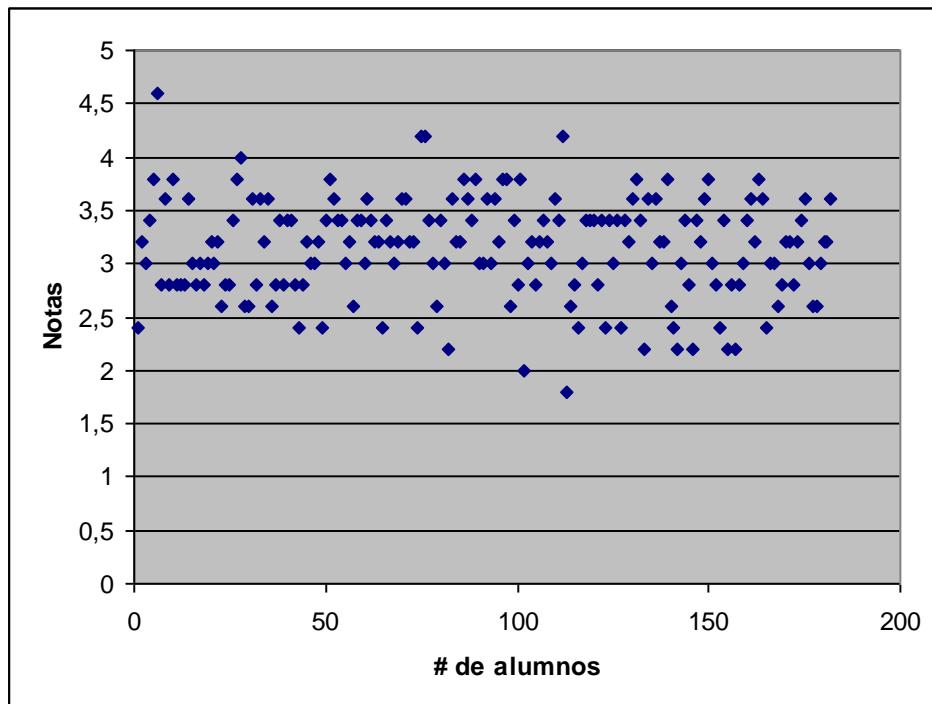


La gráfica expresa un histograma donde aparecen los 182 datos. La nota (dato) que divide a la población en dos mitades fue 3,1. Quiere decir que hacia un lado hay 91 alumnos con sus calificaciones y lo mismo hacia el otro. La calificación 2,8 divide a las poblaciones otra vez en dos mitades. La calificación 3,6 hace lo propio con la otra mitad. Esto se conoce como CUARTILES que es la división de la mediana en dos mitades. Sirve para tener referentes de ubicación. Por ejemplo, entre qué calificaciones estuvieron el 25% de los alumnos que sacaron las calificaciones más bajas y el 25% de los obtuvieron las calificaciones más altas. A propósito, esta es una medida de dispersión que será explicada un poco más en las páginas que siguen.

4.3.4. Las medidas de dispersión

Otro referente utilizado en los estudios de grupos puede ser cómo se comporta el nuevo dato (que es el que da pie al análisis) con respecto al previo, tomando de referente el centro, pero viendo cómo se distancia del mismo (dispersa). Esto es lo que se denominan las medidas o referentes de dispersión. La diferencia con las medidas de tendencia central es que en ellas los referentes son la media, la moda y el mediana del grupo control o de referenciación. Las medidas de dispersión hacen referencia a cómo quedan los datos agrupados alrededor de las medidas de centralización. A continuación se explicarán algunas de ellas.

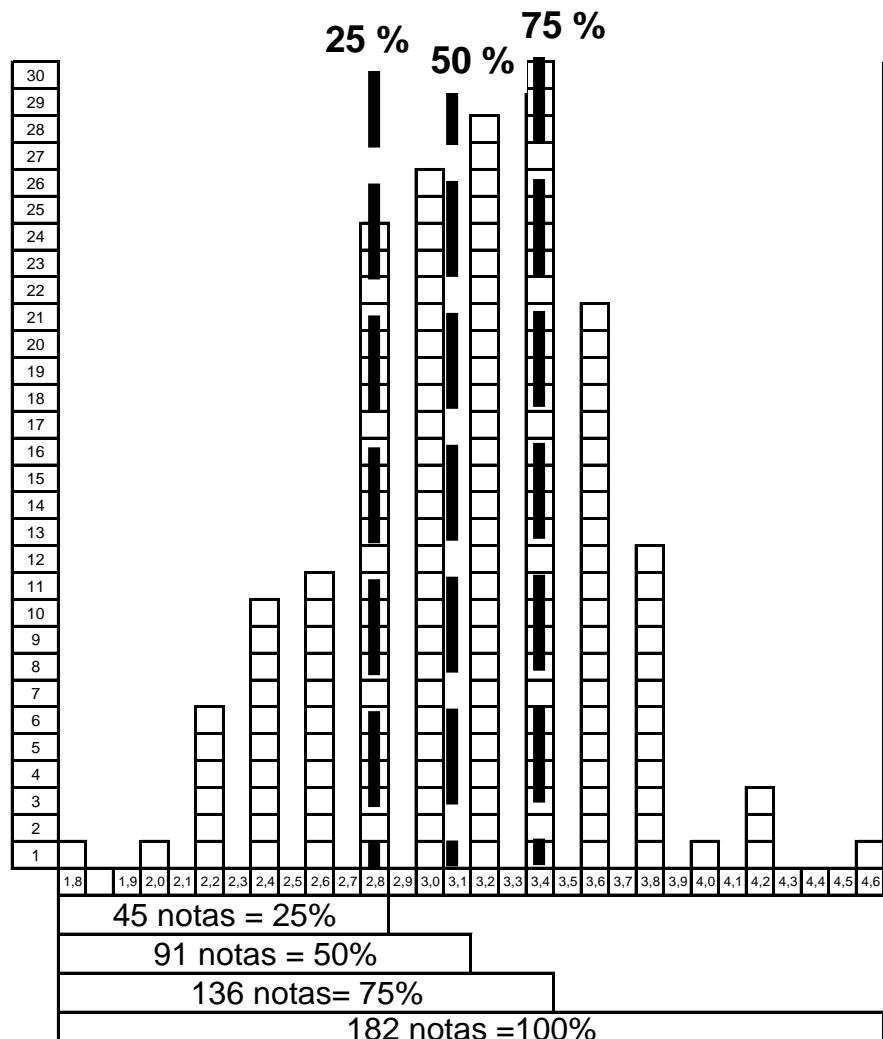
Partiendo nuevamente de las calificaciones del curso de medicina, observe el siguiente diagrama:



Este diagrama (denominado de dispersión) expresa la ubicación de las notas de los 182 alumnos. El eje X muestra el número de alumnos. Con el eje Y se muestran las calificaciones. Los puntos reflejan la calificación alcanzada por cada estudiante. El total de puntos expresa el total de estudiantes. Las notas extremos muestran (4.6 y 1.8), que la mayoría pasaron e incluso que la media (permite sospechar) se ubica entre 3,5 y 3. Estará de acuerdo que viendo esta figura se puede afirmar cómo es la dispersión de los datos... ¿cierto? Partiendo de este diagrama se harán las siguientes explicaciones, por ende, es muy importante que entienda lo mejor posible esta gráfica.

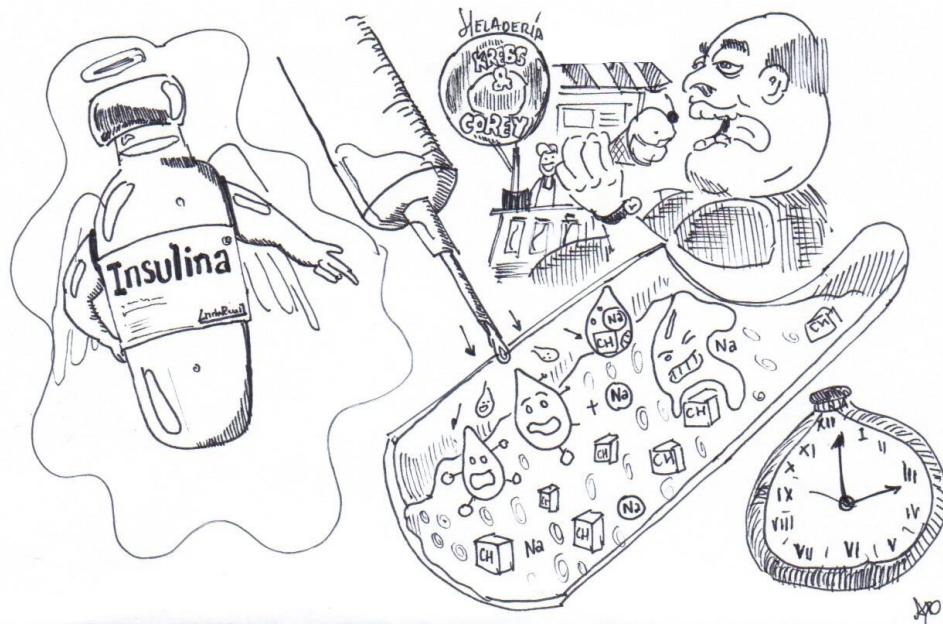
4.3.4.1. El recorrido

En este referente, se usa la distancia entre los dos límite de los rangos o lo que es lo mismo, la diferencia entre los valores extremos.



No se confunda con el cambio de aspecto de la gráfica en este capítulo. A veces es más ancha otras más angosta sólo por el efecto de la forma como se está pegando la imagen al texto.

Retomando el ejemplo de las notas, se observa que los rangos son 1,8 y 4,6. Por ende, el recorrido es de 2,8 o, lo que es lo mismo, las calificaciones en este grupo tuvieron una dispersión comprendida entre 1,8 y 4,6 con un recorrido de 2,8 (esa es la diferencia entre los dos extremos). Imagínense que los extremos estuvieran entre 0,5 (en vez de 1,8) y 4,9 (en vez de 4,6). Acá el recorrido sería de 4,4. Entonces se puede afirmar que la dispersión de las notas del grupo que obtuvo 2,8 es menos que la del grupo que obtuvo un recorrido de 4,4. Claro, entre menor sea el recorrido, el comportamiento de los datos tiende a ser más parecido (más homogéneo).



 Un ejemplo clínico puede ser el análisis de definir cuándo aparece un efecto de un medicamento. Por ejemplo, cuál es el recorrido del efecto de un tipo de insulina (para bajar la glicemia) en una población de diabéticos. Pregúntese en cuál existe una mayor ventaja para prescribirla: 1) Si tiene un recorrido muy amplio. Por ejemplo, entre 1 hora y 6 horas lo que da un recorrido de 5 horas. 2) O si el efecto se presenta en la mayoría de la población en un recorrido estrecho. Por ejemplo, entre 1 y 2 horas (rangos) con un recorrido de 1 hora. Probablemente la respuesta es que es mucho mejor tener recorridos estrechos pues así será más fácil construir la prescripción (o forma como se debe suministrar).

4.3.4.2. Los recorridos intercuartiles

En esta medida el referente grande es la población. Busca identificar en qué dato se ubica el 25% del total de la población, en qué dato el 50% en qué dato el 75%. Volviendo al ejemplo de las notas, las líneas dividen dónde se ubica el 25% de la población, el 50% y el 75%. Entonces el dato de 2,8 es donde se ubica el 25% de los datos (más o menos corresponde a 45 datos). El dato de 3,2 es la mediana o sea donde se ubica la mitad de los datos, a saber 91. Y en la calificación de 3,6 se agrupan el 75% de los datos o, lo que es lo mismo, 136 datos.

Usando un ejemplo clínico se puede retomar el del efecto insulínico donde el recorrido fue de una hora y en donde se completa la visión de referencia cuando se determina que la dosis suministrada tuvo su efecto en el 75% de las personas a los 20 minutos. Una vez más, esta información sirve para precisar cómo se deberá prescribir.



Así como se puede indagar en qué datos se ubican el 25%, el 50% el 75% (**cuartiles**) de la población, también se puede ubicar los **deciles**, es decir, en qué datos se ubica el 10%, el 20% el 30% etc. de la población. Tomando el ejemplo de las notas, el 10% de la población son 18. Entonces el primer decil estará ubicado entre la nota 1,8 y 2,4 porque entre esas notas existen 18 datos (decil 10%). Este concepto se ampliará más adelante a través de la explicación de las tablas de crecimiento. Lo cierto es que todas estas divisiones y subdivisiones solo buscan crear referentes para ayudar a entender ó, lo que es lo mismo, icrear referentes para comparar!

4.3.4.3. La desviación estándar (con los pasos para obtenerla)

Se parte de acá en adelante que se tiene un grupo de datos y las referencias que se van a obtener pertenecen a dicho grupo. Tales referencias servirán ulteriormente para ser confrontados con un nuevo dato (comparadas) y la pregunta que se hace constantemente es si tal dato (el nuevo) se comporta de manera similar o diferente a los datos del grupo.



Por ejemplo, se tiene un estudio de un colegio donde la edad media de la población es de 13 años y se ha estudiado una cantidad de variables sobre vida sexual. La pregunta es: Si se compara con otro colegio... ¿se puede hacer? La respuesta depende de crear diferentes variables a través de las cuales se pueda ver que la población de un colegio y otro son **COMPARABLES**, pues si son diferentes no se podrán hacer comparaciones; una de las variables que hace comparable o no a dos grupos es la edad media: claro, si voy a preguntar sobre inicio de vida sexual en ambos colegios para luego comparar, resulta clave que las edades de ambos sean similares; ahí ciertamente si existe diferencia me permitirá hacer análisis que expliquen por qué, pero si en uno de los colegios la edad media es de 13 años y en el otro de 17 años, las diferencias que encuentre se pueden explicar por la edad (no vale la pena dar otras explicaciones, pues estos grupos por edad no son comparables).

Una medida de referencia es la distancia de los datos de la media grupal. Se denomina **desviación media** en donde sencillamente se compara cada dato con la media. La manera de obtenerla es así:

1. Se obtienen las diferencias entre cada dato con la media grupal.
2. Se eleva al cuadrado.
3. Se suman los resultados y el resultado se denomina desviación media.

A manera de ejemplo, observe la siguiente base de datos:

Nº	IDENT	SEXO	NOTA	Nº	IDENT	SEXO	NOTA
1	ACERO DUA	m	2,4	19	BENITEZ RUI	m	3
2	ACOSTA BEL	h	3,2	20	BONILLA MO	m	3,2
3	AMADO AMA	m	3	21	BRUGES CA	h	3
4	AREVALO NC	m	3,4	22	BUITRAGO M	m	3,2
5	ARIZA ALTA	m	3,8	23	BUITRAGO M	h	2,6
6	ARIZA AGUIL	h	4,6	24	CAICEDO PE	h	2,8
7	ARIZA SAAVE	m	2,8	25	CAJICA ALVA	m	2,8
8	ARRIETA NIN	h	3,6	26	CALDERON E	h	3,4
9	ARTUNDUAG	h	2,8	27	CALVO PERE	h	3,8
10	BALLESTERC	m	3,8	28	CANON GOM	m	4
11	BARRERA BA	m	2,8	29	CARDENAS F	h	2,6
12	BARRERA CA	m	2,8	30	CASTANEDA	h	2,6
13	BARRERA CO	h	2,8	31	CASTANO LO	h	3,6
14	BELLO GONZ	m	3,6	32	CASTIBLANCO	h	2,8
15	BELTRA AVIL	h	3	33	CASTILLO GU	m	3,6
16	BELTRAN OC	m	2,8	34	CASTILLO W	h	3,2
17	BELTRAN SA	m	3	35	CASTRELLON	h	3,6
18	BENAVIDES	h	2,8	36	CASTRO VEL	h	2,6

Se trata de las calificaciones de 36 alumnos de un curso de Bioestadística. La media fue 3,15. La moda 2,8 y la mediana 2,9. Estas son las medidas de tendencia central que referencian este grupo.

Para sacar la desviación media se hace lo siguiente:

Nº	IDENT	SEXO	NOTA = x	media=x	(x-x) al 2
1	ACERO DUA	m	2,4	3,15	0,5625
2	ACOSTA BEI	h	3,2	3,15	0,0025
3	AMADO AMA	m	3	3,15	0,0225
4	AREVALO NI	m	3,4	3,15	0,0625
5	ARIZA ALTA	m	3,8	3,15	0,4225
6	ARIZA AGUIL	h	4,6	3,15	2,1025
7	ARIZA SAAV	m	2,8	3,15	0,1225
8	ARRIETA NIN	h	3,6	3,15	0,2025
9	ARTUNDUAG	h	2,8	3,15	0,1225
10	BALLESTERC	m	3,8	3,15	0,4225
11	BARRERA BA	m	2,8	3,15	0,1225
12	BARRERA CA	m	2,8	3,15	0,1225
13	BARRERA C	h	2,8	3,15	0,1225
14	BELLO GONZ	m	3,6	3,15	0,2025
15	BELTRA AVIL	h	3	3,15	0,0225
16	BELTRAN OC	m	2,8	3,15	0,1225
17	BELTRAN SA	m	3	3,15	0,0225
18	BENAVIDES	h	2,8	3,15	0,1225
19	BENITEZ RUI	m	3	3,15	0,0225
20	BONILLA MC	m	3,2	3,15	0,0025
21	BRUGES CA	h	3	3,15	0,0225
22	BUITRAGO M	m	3,2	3,15	0,0025
23	BUITRAGO M	h	2,6	3,15	0,3025
24	CAICEDO PE	h	2,8	3,15	0,1225
25	CAJICA ALVA	m	2,8	3,15	0,1225
26	CALDERON E	h	3,4	3,15	0,0625
27	CALVO PERE	h	3,8	3,15	0,4225
28	CANON GOM	m	4	3,15	0,7225
29	CARDENAS F	h	2,6	3,15	0,3025
30	CASTANEDA	h	2,6	3,15	0,3025
31	CASTANO LO	h	3,6	3,15	0,2025
32	CASTIBLANCO	h	2,8	3,15	0,1225
33	CASTILLO GU	m	3,6	3,15	0,2025
34	CASTILLO W	h	3,2	3,15	0,0025
35	CASTRELLON	h	3,6	3,15	0,2025
36	CASTRO VEL	h	2,6	3,15	0,3025
				sumatoria = Des media	8,35

Esta base de datos muestra las calificaciones de un curso de Bioestadística en el 2007. La primera columna es la clave de cada alumno. La segunda el apellido. La tercera el género. La cuarta columna la calificación. En la siguiente columna se colocó la media para poder hacer la operación matemática que expresa la última columna, que es la diferencia entre cada calificación y la media elevada al cuadrado. La última casilla muestra el resultado de la sumatoria que es la DESVIACION MEDIA.

La verdad es que esta desviación media se usa poco. De hecho solo se obtiene para poder obtener otra medida de dispersión denominada **VARIANZA**. Se obtiene de dividir la desviación media entre el número de datos menos 1: en el ejemplo: $8.35/36-1= 0.2385$.

Esta medida de referenciación tampoco tiene tanto uso (en particular en el caso nuestro que somos investigadores y no estadísticos), pero sirve para obtener la que verdaderamente se usa como el gran referente de las medidas de dispersión que es la **Desviación estándar (DS)**. Recuérdese que lo que estamos buscando es obtener referentes de cómo se dispersan los datos del centro. Se obtienen sacando la raíz cuadrada de la varianza. En el ejemplo:

Raíz cuadrada de 0.2385 =0.48

En resumen la formula sería así:

$$\sqrt{\frac{\text{Sumatoria } (x-\bar{x})^2}{N-1}}$$

De esta manera se ha obtenido el perfecto referente de dispersión, que permite ver cuánto de la media se desvían los datos que conforman el grupo. En la situación de este curso de Bioestadística, la calificación media ha sido 3.15. Al preguntar cuánto se dispersan o desvían los datos de esta media se puede afirmar que:

menos 2 ds	menos1 ds	media	más 1 ds	mas 2 ds
2,17312451	2,66156226	3,15	3,63843774	4,12687549

1. Las calificaciones entre 2.6 y 3.6 están incluidas al aplicar una desviación estándar (suma y resta de la DS de la media).
2. Las calificaciones entre 2.17 y 4.1 están incluidas al aplicar dos desviaciones estándar. En sentido práctico, siempre se usa de referente la aplicación de **dos desviaciones estándar**.

En cuanto a su interpretación, lo único que nos está diciendo la desviación estándar es que en este grupo de calificaciones, las calificaciones entre 2.1 y 4.1 de seguro forman parte de los datos esperados, pues eso se puede afirmar de los datos que se encuentran entre los límites al aplicar las dos DS.

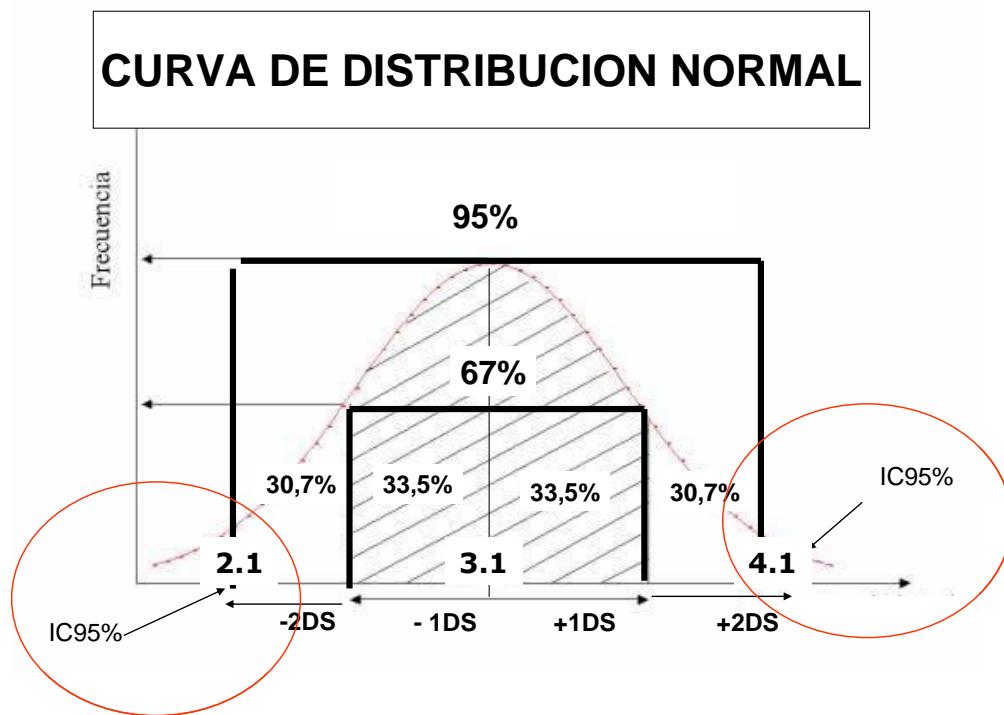


Como quien dice... la DS nos hace saber que todos los datos comprendidos entre los datos límites productos de aplicar dos DS, pertenecen a **un mismo grupo**. Piense por un momento en lo siguiente: imagínese que al examinar la respuesta de un medicamento contra la fiebre en una población se obtuvo que la controló con una media de tiempo de 15 minutos después de administrado. Al aplicar dos DS se obtuvo unos límites entre 7 min y 22 min. De ello se puede afirmar que el comportamiento de este medicamento contra la fiebre permita hacer esperar su efecto entre 7 minutos y 22 minutos. Esto es lo esperado y cualquier tiempo entre 7 y 22 minutos **forman parte de ese esperado....**

Así se calcula la DS, pero es muy importante entender su concepto. Para poder entender mejor su uso se tiene que entender el concepto de la curva de distribución normal.

4.3.5. La curva de distribución

Recopilando lo expresado en el párrafo previo, al estudiar un grupo de eventos o poblaciones y agruparlos alrededor de la media, los datos se van distribuyendo a los lados de esa media y se forma una curva, que se ha denominado **curva de distribución normal** (si cumple con ciertos criterios). Su análisis en torno a forma, ancho o altura se ha convertido en una referencia importante para ser comparada cuando se estudian frecuencias (eventos que se repiten) en distintas poblaciones.



Cuando se afirma que se forma una curva de distribución de datos es por completo cierto. Tal curva es el resultado de distribuir cada dato a lado y lado del dato media. La gráfica expresa cómo se ha formado la curva de distribución usando el ejemplo de las notas. Se observa cómo de la MEDIA (3,1 en el ejemplo de las notas o calificaciones) se dispersan los datos. Los porcentajes ponen de manifiesto cuántos datos (notas) quedan cobijados en una curva de distribución normal en torno a las desviaciones estándar (dos desviaciones = 95% de los datos). Así mismo, la gráfica expresa el concepto de INTERVALOS DE CONFIANZA del 95% como las dos notas rango que se obtienen al aplicar dos desviaciones estándar (límite superior 4,1 y límite inferior 2,1).



¡Ojo...! al aplicar una desviación estándar (a lado y lado de la media) se agrupan un 33,5% de datos a cada lado o lo que es lo mismo, el 67% de los datos del grupo están agrupados al aplicar una desviación estándar. Claro... si se aplica dos desviaciones a ese 33,5% de una lado se le agrega un 30,7% más. Como resultado de sumar esa aplicación, los porcentajes de datos (o población o lo que se esté describiendo) que quedan incluidos al tomar dos desviaciones estándar son del 95%.

Retomando el ejemplo y volviendo a las notas (calificaciones) que han servido de ejemplo, se puede afirmar que el 95% de los alumnos quedaron comprendidos en dos desviaciones estándar o lo que es lo mismo, entre 4,1 y 2,1. Como tales, calificaciones de 2,50 o de 3,90 son desde el punto de vista de la curva de distribución, **ESTADÍSTICAMENTE SIMILARES**. Sea este el momento de hacer la diferencia clínica de la estadística. Clínicamente 2,50 es diferente a 3,9, pero estadísticamente son similares (porque ambos datos están en la misma curva).



¡La diferencia clínica con la estadística sí que es importante! Clínicamente es una diferencia que presenta un sujeto en comparación con otro analizando un referente. Volviendo al ejemplo del medicamento contra la fiebre: imagínese que a una persona (de la investigación) la fiebre le disminuyó a los 11 minutos y a otro a los 20 minutos. Entre ambas personas existe una diferencia clínica en torno al momento del efecto. Pero estadísticamente ambos tiempos son similares pues están cobijados entre los límites de aplicar dos desviaciones estándar.

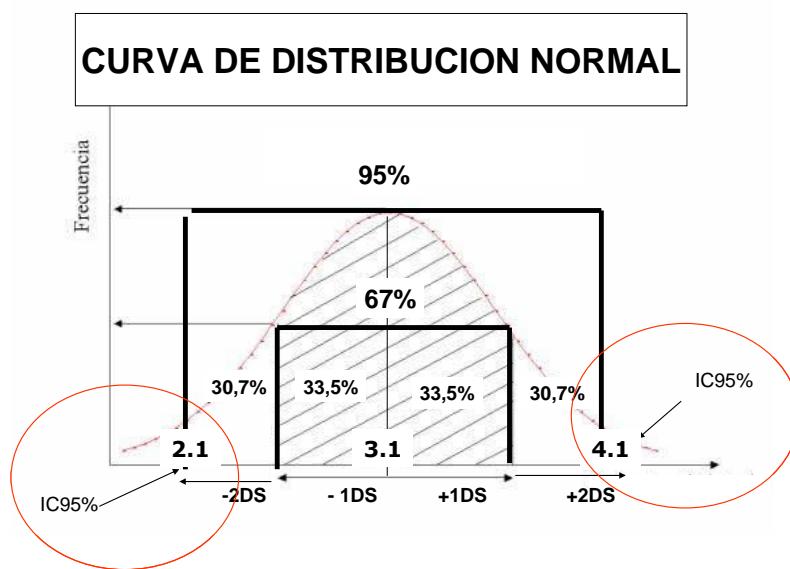
Ahora imagínese que un alumno sacó una nota de 4,6. Comparado con el referente de la curva de distribución normal se afirma que dado que está por fuera de la curva (configurada con dos desviaciones estándar) esta nota presenta DIFERENCIA SIGNIFICATIVA desde el punto de vista de la estadística con las del grupo. Por ende, su **comportamiento es diferente**. ¿Por qué? Ahí se plantean explicaciones. Muy seguramente (y planteando una hipotética explicación) la forma de estudiar del sujeto que la obtuvo es diferente a la de los demás, por ello su calificación fue diferente.



El comportamiento esperado de una frecuencia (o dato) es que esté incluida entre los datos obtenidos con la aplicación de dos desviaciones estándar. Si sale de los rangos (datos límites establecidos al aplicar las dos DS ó, lo que es lo mismo, Intervalos de Confianza del 95%) se considera que dicho dato (frecuencia) se comporta de manera diferente. Se dice que existe diferencia significativa de ese dato con respecto a los demás con que se compara). Ahí surgen preguntas que tratan de explicar dicho comportamiento diferente...

Lo que nos dice la DS y la curva de distribución, es que los grupos de población que se emplean en los estudios tienen una **tendencia a comportarse**. Que una forma de medir esa tendencia a comportarse es a través de ésta medida de dispersión. Por ejemplo, si se quisiera saber cómo es el comportamiento de la hemoglobina en la población de La Paz se deberá obtener la DS de la hemoglobina en una muestra representativa de ciudadanos de La Paz. En los límites de la aplicación de dos DS se obtiene cómo se comporta el 95% de la hemoglobina en dicha población y se dice que estadísticamente todos los resultados que se encuentren entre esos dos límites (Intervalos de Confianza) derivados de la aplicación de dos DS constituyen la **NORMALIDAD** de ese comportamiento en el caso de la hemoglobina en esta población. Pero, sea este el momento de decir que dicho comportamiento de la hemoglobina en los habitantes de La paz no es igual al comportamiento de habitantes de la zona Caribe (a nivel del mar). Ellos de seguro tienen un comportamiento de la curva de distribución de la hemoglobina diferente y esa es su NORMALIDAD. Por ello **la normalidad estadística es sencillamente el resultado de la curva de distribución de una población específica**, criterio muy diferente a la normalidad clínica o la normalidad social. Lo más importante de la curva de distribución es que ella se puede usar, entre otras, para saber si dos poblaciones son comparables; por ejemplo: en estudio de anemias... ¿son comparables las poblaciones de la Paz (Bolivia) con las de Cartagena (Colombia)? (Por si las moscas... la respuesta es no).

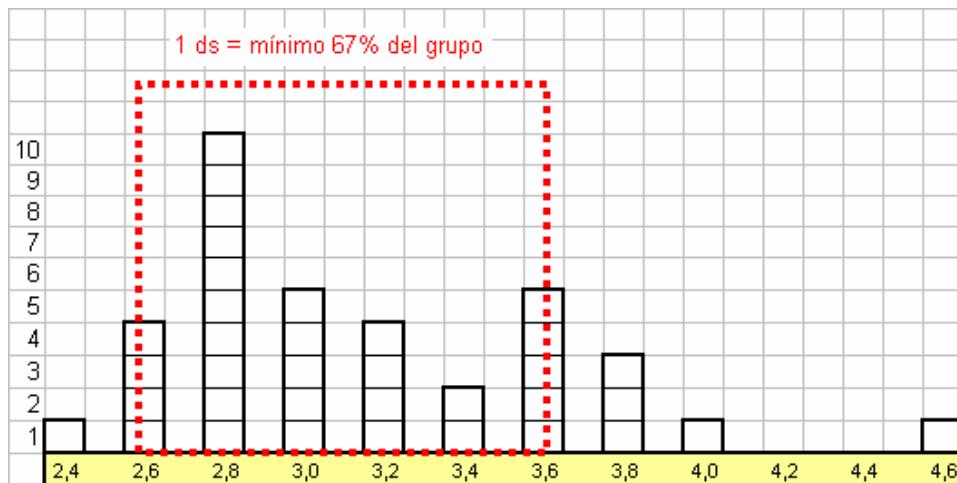
Para completar el concepto de la curva de distribución normal, a continuación se explicará un poco más el de distribución de población (o de datos):



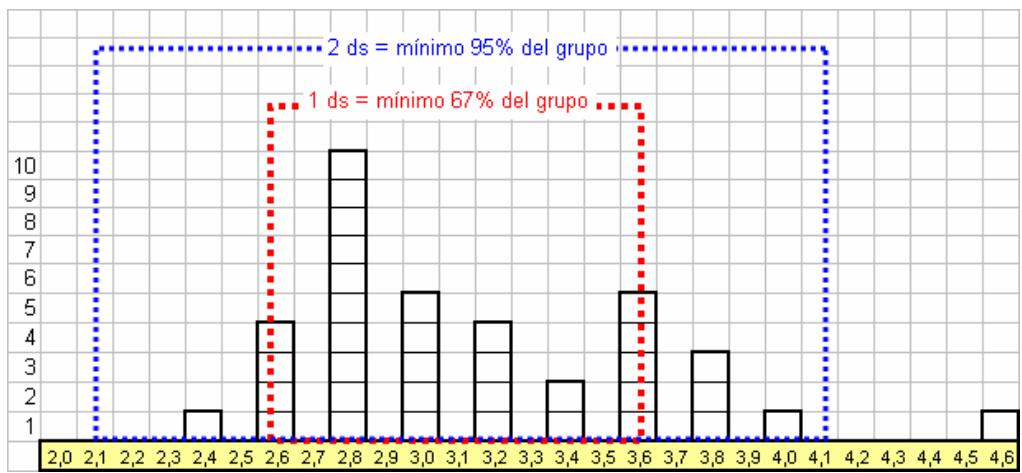
Como se observa la desviación típica o estándar fue de 0,48. Como es una medida de referenciación (para comparar) su uso puede ser como sigue. Se tiene una nota de 4,0. ¿Cómo está referenciada con respecto a la desviación estándar? Se puede decir todo lo siguiente:

1. Al tomar la media (3,1) más una desviación estándar quedará $3,1 + 0,48 = 3,6$. Con ello se establece que las notas comprendidas entre 3,1 y 3,6 están ubicadas a más una desviación estándar del centro. Lo propio se puede decir al restar la media - una desviación estándar, siendo en el ejemplo el resultado 2,6 (entre 2,6 y 3,6 se ubican las notas entre la media y menos una desviación estándar).

Entre estos dos rangos (2,6 y 3,6) se distribuyen **al menos el 67%** (pueden ser más, pero de seguro no menos) de las notas (una DS). Si quiere confírmelo observando el histograma y contando el número de datos que existe entre 2,6 y 3,6, recordando que 24 datos son el 67% del total (36 notas). En el ejemplo del histograma los datos comprendidos entre 2,6 y 3,6 son 30, es decir más del 67%.



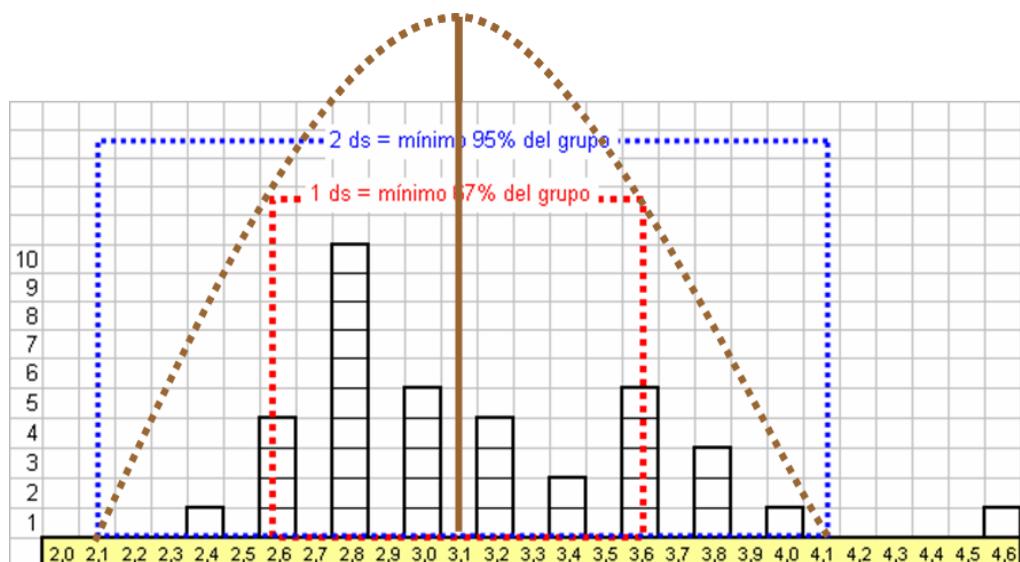
2. Si se toman dos desviaciones estándar, la calificación se extiende hasta 4,1 (al sumar) y hasta 2,1 (al restar). Estos límites deben cobijar al menos al 95% de la población (es decir, 34 datos o calificaciones). El histograma sería así:



Como se observa, las DOS DESVIACIONES ESTÁNDAR cobijan 35 calificaciones (entre 2.1 y 4.1), algo más del 95% del total de calificaciones.

En el ejemplo, la nota de 4,0 queda ubicada a dos desviaciones estándar del centro, o lo que es lo mismo: estadísticamente pertenece a este grupo o es similar estadísticamente al resto de calificaciones.

Al aplicarse el programa estadístico de gráficas se dibujará una curva de distribución que quedaría aproximadamente así:



Entonces y para resumir, se puede afirmar lo siguiente:

1. Cuando se estudia un grupo, un referente importante es la curva de distribución. La interpretación que se ha de hacer es que mínimo el 95% de los datos estarán cobijados al aplicarle a la media dos desviaciones estándar (más y menos).
2. Todos los datos que estén cobijados pertenecen a este grupo y, por tanto, **estadísticamente son similares**. Los que estén por fuera de la curva son **DIFERENTES SIGNIFICATIVAMENTE**.
3. Este referente sirve para comparar datos. Por ejemplo, si se pregunta: Un alumno sacó una calificación de 4.0: ¿pertenece a este grupo en torno al comportamiento de la calificación? La respuesta es que SÍ, pues esta calificación está dentro de la curva que se forma al aplicar dos desviaciones estándar.

Aprovechando la gráfica hágase esta pregunta: 'En ese grupo, la calificación de 4.6... ¿pertenece o no al mismo grupo?' Obviamente se sabe que sí, pues de hecho un alumno la sacó, pero vean que está por fuera de la curva de distribución normal. Por consiguiente, estadísticamente se debería contestar que no. Dado que esto no es cierto (no REFLEJA la realidad) podemos afirmar que estamos ante un ERROR, que es decir que algo no pertenece cuando realmente pertenece. Sea el momento de decir que existe otro error, que es decir que sí pertenece cuando la realidad es que no. Se llaman los **errores alfa y beta** que más adelante se volverán a revisar...

Las aplicaciones de este concepto son inmensas en el mundo de la investigación. La curva de distribución es uno de los referentes claves de las comparaciones. A manera de ejemplo se mencionan:

1. Cuando se van hacer investigaciones en dos grupos de población y se necesita saber si existen diferencias entre los grupos. Se escogen las MEDIAS claves y se compararan entre los dos grupos. Lo que se espera es que no exista diferencia entre ambos grupos.
2. Cuando entre dos grupos se obtiene un resultado y se necesita saber si existe diferencia entre uno y otro.



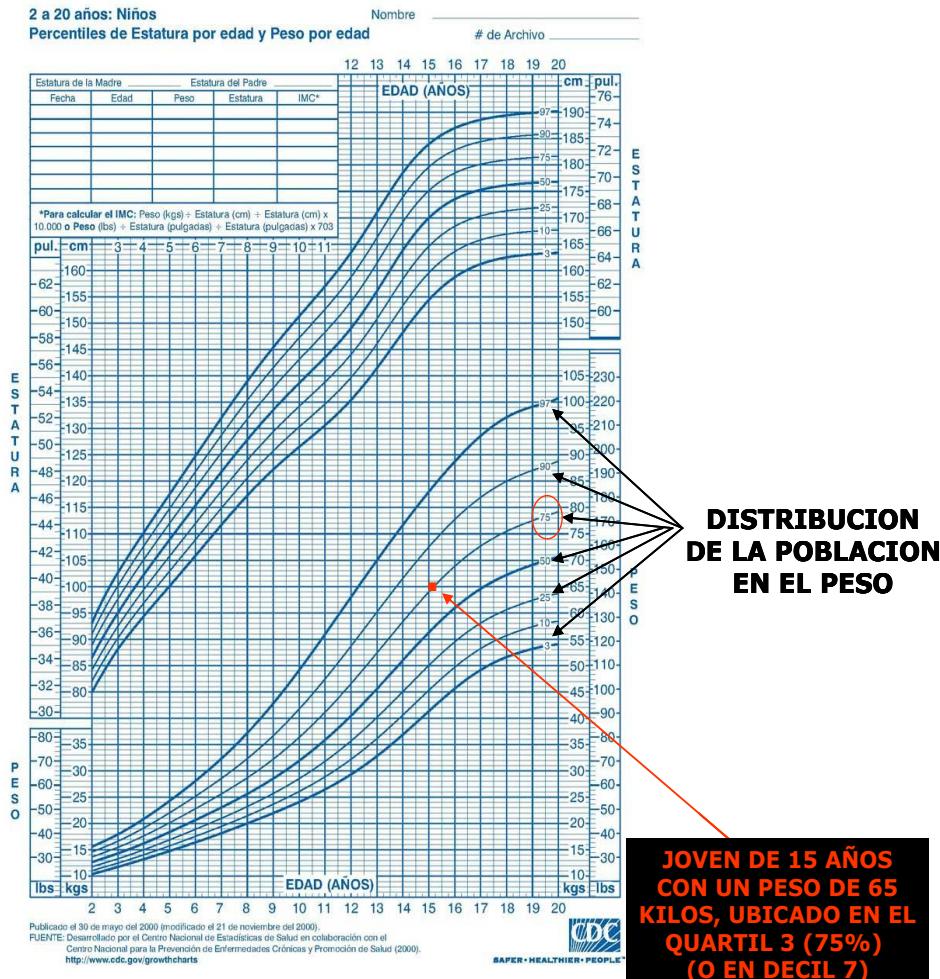
Imagínese que se ha diseñado una INTERVENCION EDUCATIVA a través de la cual se espera retrasar el inicio de la vida sexual de los adolescentes. Se decide probar usando un curso de 36 jóvenes de un colegio con características A, B y C. Usted debe escoger el curso de control... Resulta evidente que debe tener características A, B y C similares al curso que se va intervenir. Pues bien... lo que se debe hacer es:

1. Tomar la característica A, B y C del curso intervenido y ver cuáles son sus curvas de distribución.
2. Luego se hace lo propio con el curso de control.
3. Se COMPARAN las curvas del grupo intervenido y el de control y se espera que sean SIMILARES ESTADISTICAMENTE.

¿Por qué? Sencillamente porque esa es la única manera de estar seguro que lo que se encuentre al terminar la investigación es debido a la INTERVENCION EDUCATIVA y no a que desde antes existían diferencias entre los grupos.

La detección de la DIFERENCIA entre las medidas es la razón de la estadística analítica.

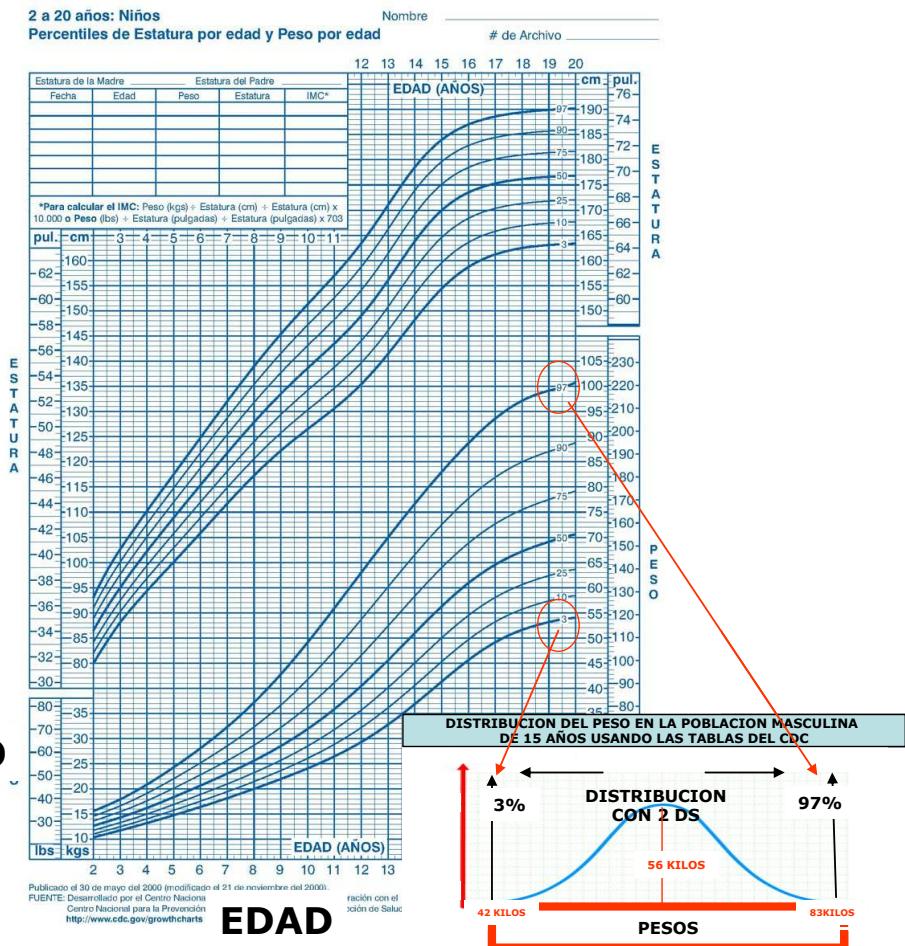
4.3.5.1. Explicación de la curva de distribución y su utilidad a través de las tablas de peso y talla



Una herramienta (instrumento) muy utilizado en el ejercicio clínico es el uso de escalas o tablas donde se puede confrontar lo que se espera con lo que los pacientes tienen. Las tablas de crecimiento son un buen ejemplo. Se construyen con base a las distribuciones de las poblaciones en cada edad, las que sumadas muestran las tendencias y así permiten construir los límites donde se ubican las poblaciones. Antes de exponer el ejemplo (joven de 15 años con...) se debe describir lo que compone la **tabla de percentiles**:

TALLA

PESO



1. La Tabla relaciona los valores para la edad (X) con los parámetros de peso y talla (Y). Para hacer más fácil la explicación sólo nos vamos a referir al peso.
2. De cada edad se obtiene la curva de distribución normal del peso con dos desviaciones estándar. Por ejemplo, con los jóvenes de 15 años:

**DISTRIBUCION DEL PESO EN LA POBLACION MASCULINA
DE 15 AÑOS USANDO LAS TABLAS DEL CDC**

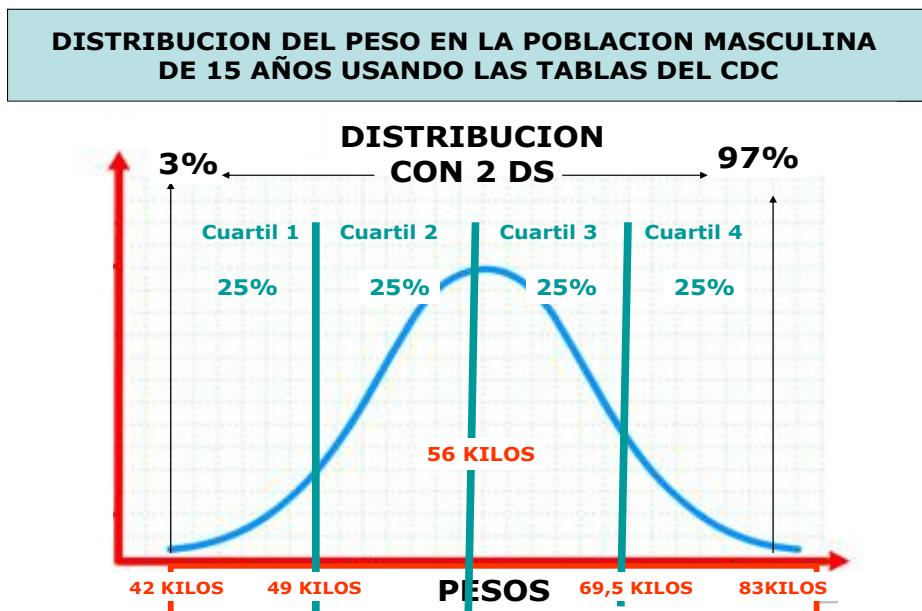


A una muestra representativa de jóvenes de 15 años sanos del mismo sexo (masculino en este ejemplo) de un país se pesó y se obtuvo que el peso medio era de 56 kilos. Al aplicar dos desviaciones estándar, dicho peso medio se dispersó hasta 83 kilos (límite superior del intervalo de confianza) y hasta 42 kilos (límite inferior del intervalo de confianza). Todos los pesos que estén dentro de 42 kilos y 83 kilos en jóvenes de 15 años, de ese lugar, se consideran NORMALES ESTADÍSTICAMENTE (una manera de decirlo) o IGUALES ESTADÍSTICAMENTE (otra manera de decirlo y según el contexto).

3. Dado que se ve por edades consecutivas, al unir los límites superiores (e inferiores) de los intervalos de confianza del 95% de cada edad se forman esas líneas continuas (tendencias).
4. Dado que el 95% de los datos quedará comprendido entre las dos desviaciones estándar, los límites de esos dos intervalos son 97% (superior) y 3 % (inferior). Ello quiere decir que el 95% de los pesos obtenidos de los jóvenes de tal edad están comprendidos entre esos dos límites. La forma como se subdividen estos dos límites es lo que establece:

Cuartiles si se dividen por proporciones del 25% los pesos encontrados. Ello quiere decir que en el primer cuartil está el 25% de los primeros pesos (entre 42 y 49 kilos), que en el segundo cuartil, está ese 25% más el siguiente 25% (y con ello se ha llegado a la mitad de la población), es decir, entre 42 kilos y 56 kilos; que en el tercer cuartil se agrega otro 25% (entre 42 kilos y 69,5 Kilos) y en el último cuartil se incluye el restante 25% de los pesos ubicados en la curva de distribución.

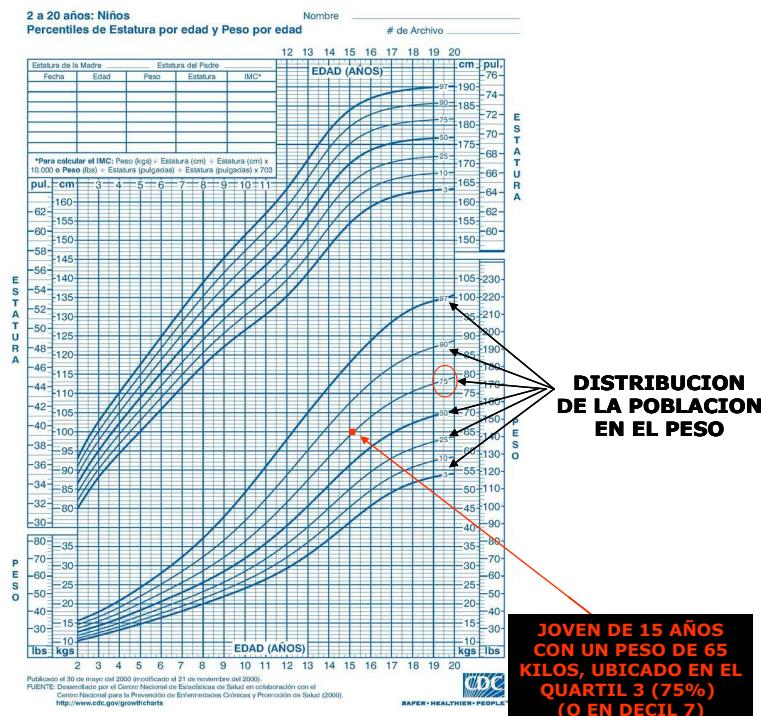
Igualmente se puede afirmar que entre 49 kilos y 56 kilos están ubicados los pesos de los dos cuartiles inferiores y que entre 56 y 83 están los pesos con cuartiles superiores. En fin... lo importante es que esta medida se usa para comparar tendiendo de referente la curva de distribución.



Por ejemplo, cuando incluyendo a todas las entidades de educación (100%) se hace un examen y con base en los resultados se mira la ubicación de una entidad con respecto a todas las demás, se puede afirmar que está en el cuartil inferior, o en el medio inferior o medio superior o en el cuartil superior.

Deciles (percentiles) si lo que se hace es dividir el 100% de los pesos de la población en 10%, 20%, 30%, 40%, 50%, etc. Evidentemente esta es otra forma de clasificar la curva que se usa para poder tener referentes.

Lo cierto es que así se forma la tabla y se puede identificar si el peso de una persona de tal edad está dentro de lo **normal esperado**.



Por ejemplo, (y volviendo a la imagen donde aparece la primera tabla) el punto rojo señala el peso de un joven de 15 años. ¿Dónde está ubicado en torno a la curva de distribución normal para su edad? Está situado en la línea que señala el 75. ¿Ello qué quiere decir? En primer lugar que el peso de este joven está dentro de la curva de distribución normal del peso para la edad y el sexo (evidente por estar dentro de los dos límites que establecen las líneas que miden la superior 97 y la inferior 3). Ahora bien... dentro de la curva el percentil 50 es aquel peso que divide a la población de esta edad en dos mitades (los pesos ubicados ahí señalan la **MEDIANA**) donde una mitad puede tener más que ese peso y la otra mitad menos. En este caso y, si la imagen permite visualizarlo, el peso del percentil 50 a los 15 años, es 56 kilos. Dado que este joven tiene un peso de 65 kilos se puede afirmar que:

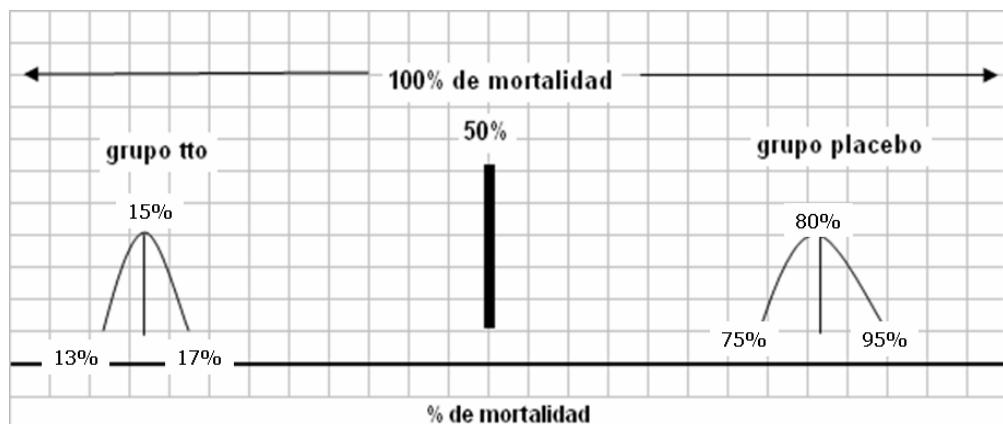
- Su peso está dentro de la curva de normalidad para su sexo y edad y región (en este caso y puesto que las tablas son del CDC seguramente de población norteamericana).

- Que su peso (65 kilos) está en el percentil 75%, es decir, que los jóvenes con peso normal pero, sin embargo, por encima de él (más peso) son un 25% y que él tiene un peso mayor al 75% de los jóvenes iguales a él en edad sexo y región (normales estadísticamente). Si se quisiera expresar su ubicación con respecto a la población en cuartiles, es el tres donde se ubica.

Obviamente la importancia de esta aplicación de las curvas de distribución normal ha sido y es inmensa. Por ejemplo, al hacer seguimiento de los niños por edad, podemos ver cómo va ocurriendo el incremento del peso en relación con la edad y ver si se hace dentro de los límites esperados; si no es así, nos hace sospechar que algo no está bien en el niño... y se busca.

4.3.5.2. La curva de distribución y la similitud y diferencias estadísticas

En epidemiología sí que se trabaja con las desviaciones estándar y las curvas de distribución normal; sin duda porque lo que se estudia son frecuencias. A continuación se enseñará cómo a través de las curvas se identifica la similitud o la diferencia estadística.



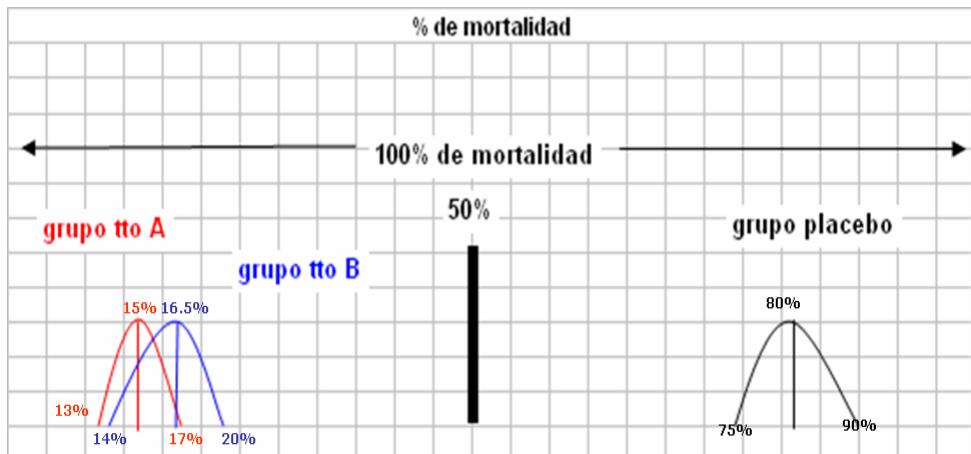
Retomando la intervención con antirretrovirales comparada con los placebos, mencionemos que la media de mortalidad al año en un grupo de antirretrovirales fue de un 15% y al aplicar las dos desviaciones estándar la curva de distribución quedó comprendida entre 13,5% (al restar dos DS) y 17,5% (al sumar dos DS). Ello quiere decir que las mortalidades desde 13,5 a 17,5 % son estadísticamente similares en este grupo de pacientes (los que reciben tratamiento). Ahora bien, en el grupo de comparación, al que se le suministró placebos, la media de la mortalidad fue de 80% y al aplicarle las dos desviaciones estándar quedó entre 75% y 95%, ó como quien dice que a un año, entre el 75%

y el 95% de los pacientes se mueren. Al comparar los dos grupos se puede ver que las dos curvas **no se encuentran** en momento alguno, por ende, se puede decir que definitivamente la mortalidad de estos dos grupos se comporta de manera **SIGNIFICATIVAMENTE DIFERENTE**, queriendo expresar el término significativo que la diferencia está confirmada por la estadística (de hecho tal término surge al aplicar una prueba estadística). De una vez se puede aclarar (y repetir la explicación) que en el grupo de antirretrovirales los datos 13,5 % (límite inferior) y 17,5% (límite superior) se denominan INTERVALOS DE CONFIANZA DEL 95%, así como en la otra curva, 75% (**Límite inferior**) y 95% (**Límite superior**) expresan los respectivos intervalos del 95%.



Deténgase un momento y repase el concepto de los intervalos de confianza... Se usa prácticamente en la totalidad de las investigaciones, por tanto, queda claro que será de gran valor.

Ahora, y para completar el ejercicio (y esperamos que el entendimiento) ¿qué pasaría si se incluyese un tercer grupo de población al cual se le da otro antirretroviral B? El comportamiento de los tres grupos en torno a la mortalidad queda expresado en la siguiente grafica:



¿Qué se observa? En primer lugar y describiendo:

1. Se observan las tres curvas de distribución del comportamiento de la mortalidad.
2. El grupo A tienen una media de mortalidad de 15% con IC 95% de 13%-17%.

3. El grupo B tiene una medida de mortalidad de 16,5% con IC 95% de 14%-20%.
4. El grupo de control o de placebo tiene una media de mortalidad de 80% con IC de 95% de 75%-90%.
5. Gráficamente los grupos A y B se diferencian del grupo placebo. Las curvas no se encuentran, por lo cual las diferencias son significativas. Pero, ¿qué sucede entre el grupo A y el grupo B?
6. Las dos curvas de distribución se interponen entre sí. La ciencia ha establecido que para mantener una DIFERENCIA SIGNIFICATIVA se acepta que un 5% de las curvas se mezclen pero si es más, entonces ya entre las curvas no existe diferencia significativa.

Ese 5% se expresa en el lenguaje de probabilidades como 0.05 o, lo que es lo mismo, $p = 0.05$. Claramente el grupo de TTO A con el grupo de TTO B no presentan diferencias significativas. Es decir,... lo mismo dar uno u otro antirretroviral. Y eso se afirma dada la mezcla entre las dos curvas que hace que en algunos momentos dé lo mismo pertenecer a A o a B.

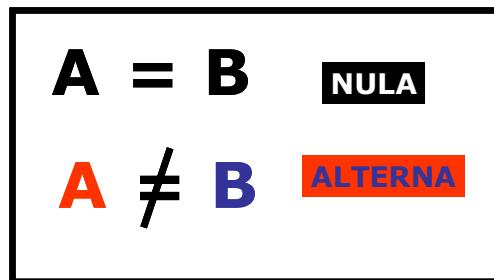
Ciertamente existen unas pruebas estadísticas (la prueba del χ^2 , entre otras) que expresan si existe o no diferencia significativa usando como indicador la p de 0.05. Si al aplicar la prueba la p da igual o menos de 0.05 (que es lo mismo que el 5%) entonces existen **DIFERENCIAS SIGNIFICATIVAS**. Quiere ello decir que lo que se está comparando tiene una diferencia de más del 95% o que coinciden (se mezclan las curvas) en menos de un 5%. Si la prueba da una p mayor de 0.05 (mayor de un 5%) entonces desde el punto de vista científico **no existe diferencia estadística**, ya que arbitrariamente en ciencia se dice que cuando se comparan dos elementos, si coinciden en más de un 5% son similares estadísticamente. En el caso del TTO A y B esa es la situación.

4.3.5.3. La curva de distribución y las hipótesis estadísticas

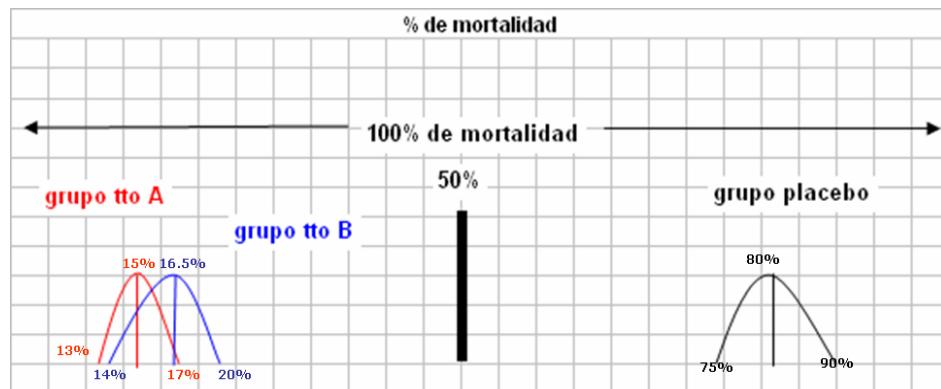
La hipótesis estadística surge al comparar dos elementos. Al hacerlo, sólo existen dos posibilidades:

1. O son iguales (denominada la HIPOTESIS NULA).
2. O son diferentes (denominada la HIPÓTESIS ALTERNA).

Gráficamente se expresa:



Para encontrar la respuesta se aplican las denominadas PRUEBAS ESTADÍSTICAS DE HIPÓTESIS, pruebas que se escogen de acuerdo a lo que se esté comparando. Al aplicar este concepto al ejemplo que se ha venido trabajando y recordando la imagen:



Primera comparación:

Grupo A vs. grupo placebo que asume la letra B. La pregunta hipotética es:

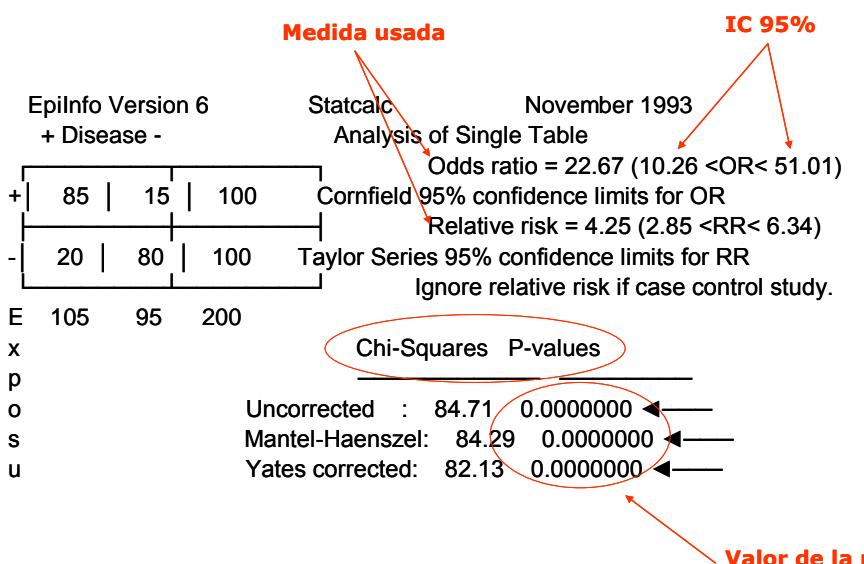
¿Existe diferencia en la mortalidad entre el grupo A y el grupo B? Usando las medias de mortalidad de una vez se ve que existe una diferencia clínica muy evidente: el grupo A tiene una media de mortalidad a un año del 15% en tanto el grupo B (o placebo) tiene una

media de mortalidad al año de 80%. Pero, lo que se quiere saber es si desde el punto de vista estadístico esta diferencia clínica persiste. La diferencia estadística permite asegurar si existe diferencia y que, al menos en un 95% de las veces en que en iguales condiciones se haga esta investigación, los resultados evidenciarán diferencias (ó no, de acuerdo al resultado). En cambio con una diferencia clínica no se puede estar seguro que ella fue producto del azar o lo que es lo mismo que no se tiene certeza de que se repita. La certeza lo establece que de 100 veces al menos en 95 exista esa diferencia. Entonces los pasos de la comparación entre A y B son:

1. Construir y colocar las cifras de la tabla comparativa:

		MORTALIDAD		
		No	Si	
A=grupo a	85	15		100
	20	80		100
		105	95	

2. Obtener la medida de comparación: en este caso es OR de 22. Quiere decir que por cada persona que recibió placebo y no murió, existen 22 que recibieron el tratamiento del grupo A y no murieron.
3. Se aplica la prueba estadística para establecer si este resultado es o no SIGNIFICATIVO. Entonces, y usando la tabla estadística del programa EPI INFO:



El programa arroja varias pruebas estadísticas. Están las de los IC de 95% para los "Odds Ratios" o los "relative risk". En este ejemplo, las medidas son SIGNIFICATIVAS.

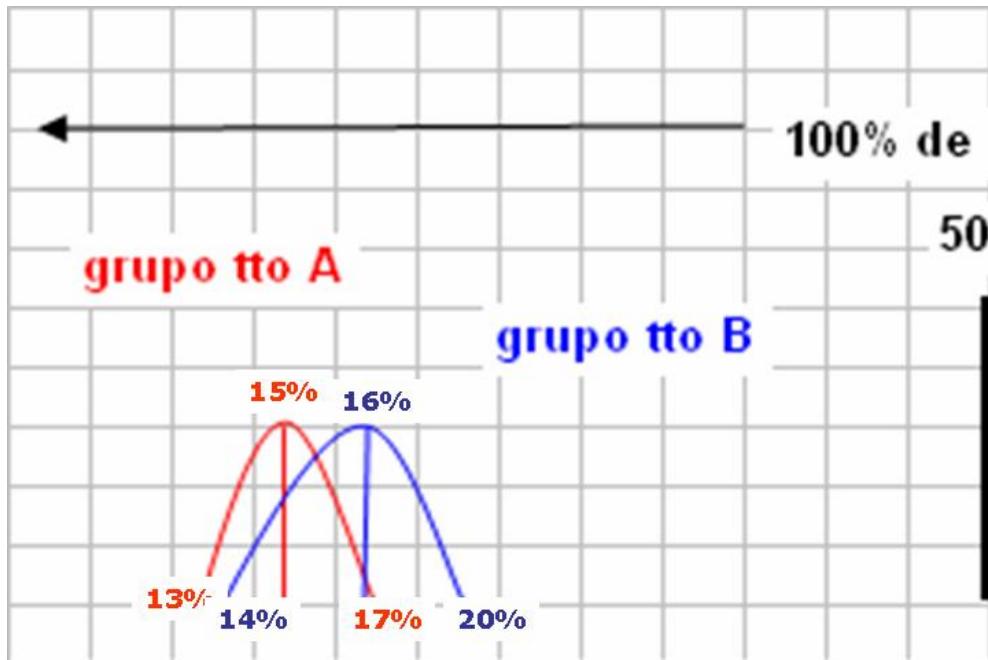
También puede usar el Chi (χ^2) cuadrado, evidencia que las p son menores de 0.05 (de hecho son menores de 0.000000) o lo que es lo mismo que la p (posibilidad) de que este resultado sea debido al azar es menor de 0.000000, es decir,... prácticamente imposible.

4. La interpretación que se hace es que A es diferente de B estadísticamente (Hipótesis alterna) y traspasado al idioma de la investigación que los pacientes tratados con A vs los tratados con placebo mueren menos estadísticamente. Como quien dice A beneficia a los pacientes... al ser comparado con el placebo.



¿Cuánto lo beneficia? Pues observen la diferencia: 85% no mueren con TTO A menos 20% que no mueren con placebo; es decir, que recibir TTO con A hace que mueran menos con una diferencia de 65%.

Ahora, y para completar el ejemplo, se va a comparar el grupo A con el grupo B. Recuerde la imagen:



Entonces los pasos son:

1. Construir y poner las cifras en la tabla comparativa:

		MORTALIDAD		
		No	Si	
A=grupo a	85	15	100	
	B=grupo b	84	16	100
		169	31	

2. Obtener la medida de la relación que es de OR de 1.08. Quiere decir que por cada persona que recibió TTO del grupo B y no murió, 1.08 personas del grupo A no murieron. Como quien dice un "poquitico" más no murieron.
3. Se aplica la prueba estadística para establecer si este resultado es o no significativo. Entonces, y usando la tabla estadística del programa EPINNFO:

+ Disease -			Analysis of Single Table		
<hr/>			Odds ratio = 1.08 (0.47 <OR< 2.48)		
+ 85 15 100			Cornfield 95% confidence limits for OR		
<hr/>			Relative risk = 1.01 (0.90 <RR< 1.14)		
- 84 16 100			Taylor Series 95% confidence limits for RR		
Ignore relative risk if case control study.					
E 169	31	200			
x			Chi-Squares	P-values	
p			<hr/>	<hr/>	
o	Uncorrected :	0.04	0.8450917		
s	Mantel-Haenszel:	0.04	0.8454746		
u	Yates corrected:	0.00	1.0000000		
r					
e	F2 More Strata; <Enter> No More Strata; F10 Quit				

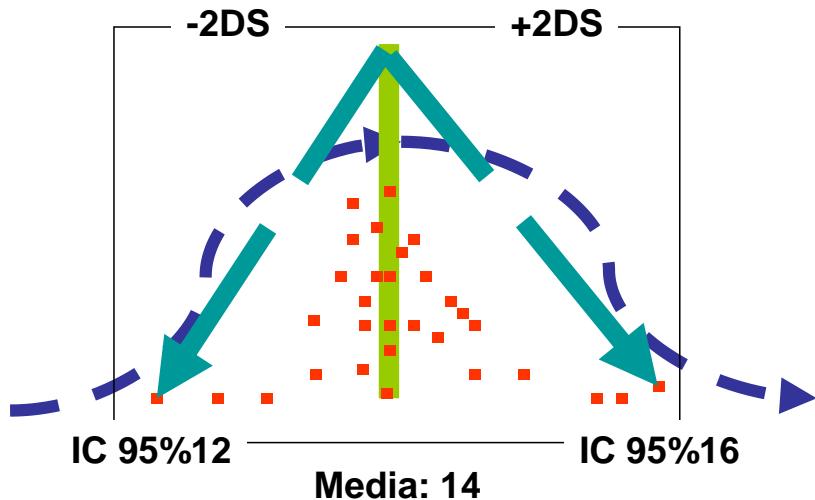
El programa evidencia que NO ES SIGNIFICATIVA la diferencia. ¿Por que? Pues por la interpretación:

- La prueba de IC de 95% muestra cómo los límites de los IC de 95%, cruzan el 1.
 - La prueba de la p del Chi cuadrado es mayor de 0.05.
4. La interpretación que se hace es que no existen diferencias estadísticas en tratar a los pacientes con A o con B. Por ende, A = B o como quien dice que se demostró la HIPOTESIS NULA.

Finalmente observe la interpretación similar entre la prueba p y los intervalos de confianza. Por supuesto que es lo esperado porque tanto una como otras son solo pruebas estadísticas en donde escoger una u otra depende de la modalidad de estudio. Completemos esta visión con otro ejemplo:

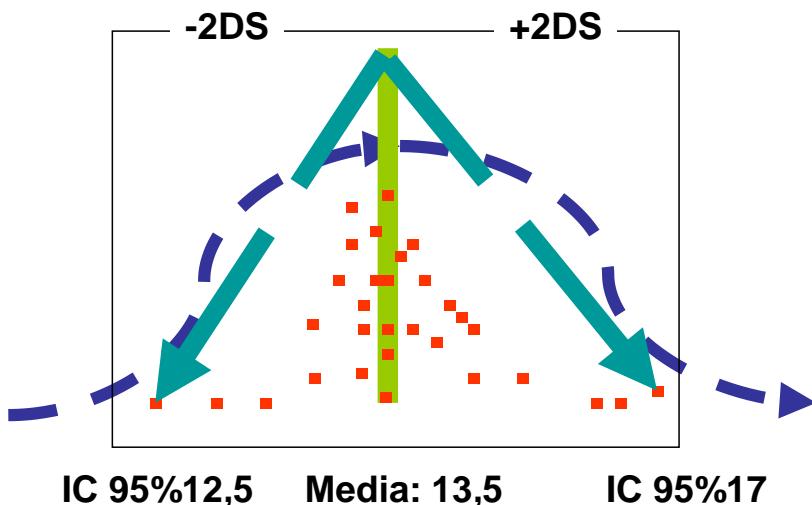


Imagínense que se desea saber y comparar cuál es la prevalencia (número de jóvenes que han tenido relaciones sexuales/total de jóvenes del colegio X 100) de relaciones sexuales en adolescentes de dos colegios. Resulta obvio que se necesita identificar muy bien a quién se le va preguntar en cada colegio. Es predecible que si le pregunta a jóvenes de 12 años y a otros de 15 si han tenido relaciones sexuales, puede (y de hecho existen) francas diferencias (un año más en la adolescencia ciertamente establece grandes contrastes). Resulta por completo racional que la pregunta ha de hacerse a jóvenes de las mismas edades. Así, se pueden comparar. Por ende, el paso obvio es conseguir a los adolescentes de uno y otro colegio, y obtener la edad promedio. Se espera que sean SIMILARES para poder ser comparables. Entonces al sacar la curva de distribución de edades del colegio A se obtuvo:

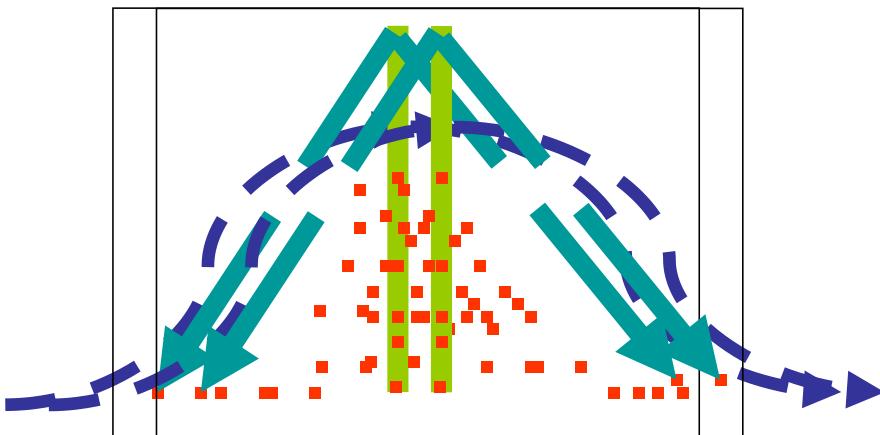


- La edad media de los adolescentes del colegio A fue de 14 años y el 95% del grupo está entre las edades de 12 años (límite inferior del IC de 95%) y 16 años (límite superior del IC de 95%).

En el colegio B el resultado de la curva fue:

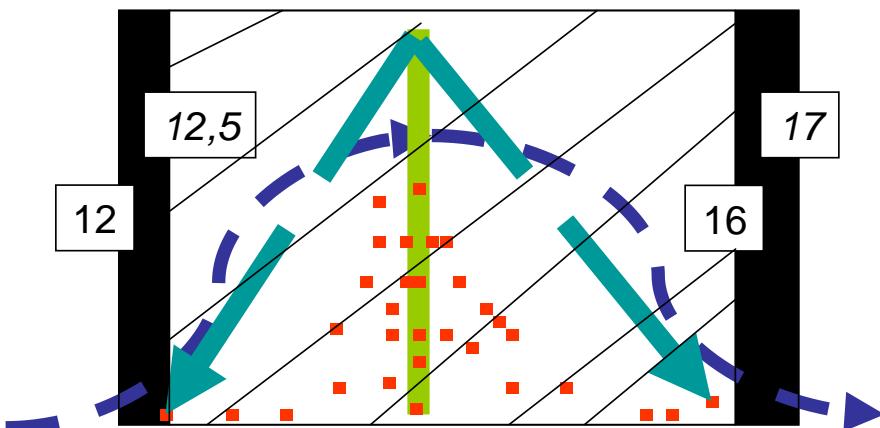


- La edad media de los adolescentes del colegio B fue de 13,5 años. El 95% de los jóvenes están entre 12,5 años y 17 años (límites de los intervalos de confianza).



Ahora la pregunta es de si estas dos poblaciones son **IGUALES** en torno a la edad, ello con el fin de saber si se pueden comparar. Entonces, y para responder, obsérvese lo que sucede al interponer las dos curvas:

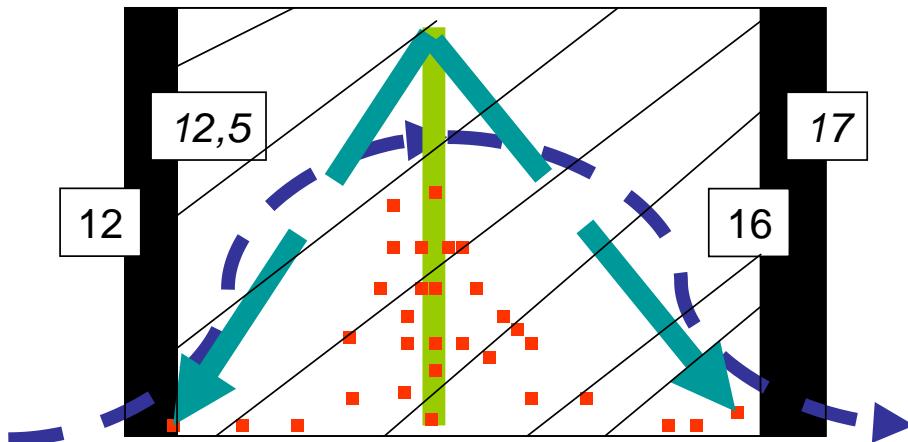
Como se observa existe una gran parte en donde las dos curvas se encuentran y dos pequeñas franjas en donde no se encuentran. La siguiente gráfica lo representa:



Resulta obvio que en las franjas oscuras estas dos poblaciones no son comparables en lo referido a la edad. ¿Por qué? Pues el colegio B no tiene adolescentes entre los 12 y los 12,5 años y el colegio A no los tiene entre 16 y 17 años (¿cómo se compara lo que no se tiene?). Pero, así mismo, resulta evidente que comparten las edades entre 12,5 y 16 años que por cierto son la mayoría de la población de ambos colegios (proporción mayor de ambas curvas). Ahora, y trasladando esto al lenguaje de la estadística, la pregunta que se hace es: Tomando como referente las edades...

1. A es = B

2. A es diferente de B



Estadísticamente se ha establecido que dos poblaciones (A y B) son similares en torno a un criterio (la edad en este caso, entonces la edad promedio de A es = a la edad promedio de B) si las diferencias entre ellas no son mayores de un 5%. Ello quiere decir que A es = a B si las dos franjas oscuras no son mayores de un 5 %. Otra manera de expresarlo es si las curvas coinciden en más de un 95%.

Ahora pruébese y construya la tabla comparativa y saque la prueba p para dar respuesta oficial de si A es = B o A es diferente a B.

4.4. La estadística analítica

Recuerde: la estadística puede entenderse como referentes que sean útiles para ver e interpretar la (o las) respuesta a las preguntas de investigación expresadas numéricamente. Las preguntas de investigación se contestan de tres maneras:

- Describiendo
- Comparando
- Interviniendo

Para describir se usa el lenguaje derivado de la estadística descriptiva. Para comparar e intervenir (que no es otra cosa que comparar) se usa la estadística analítica. La comparación de una situación A con otra (B) sólo puede dar lugar a dos posibles respuestas:

- a) A es = B
- b) A es diferente a B

La estadística analítica se dedica precisamente a ello: a identificar si al comparar dos situaciones éstas son similares o diferentes. Estas dos posibilidades constituyen las respuestas y son las probables hipótesis. Si A es = a B se denomina **HIPÓTESIS NULA**. Si A es diferente a B se denomina **HIPÓTESIS ALTERNA**.

Son muchísimos los momentos en epidemiología en que comparamos. De hecho, cuando se escogen diseños intervencionales (o experimentales), en todos ellos existirá un grupo de comparación y se trata de ver los resultados a las preguntas de investigación comparando el efecto o la medida en el grupo de intervención versus el de control. En los diseños observaciones, en la subclasificación de analíticos, igualmente se compara (sólo las investigaciones descriptivas con sus preguntas no comparan).



Cuando se hacen experimentos se trabaja con un 'pedazo' de la realidad y se busca de una u otra manera probar que una situación es igual a otra o que es diferente. Por ejemplo, si se da un medicamento para una enfermedad, es necesario que a un grupo se le dé el medicamento (A) y a otro no (B). La pregunta será: ¿El grupo A se comporta (comportamiento = morbimortalidad) igual que el grupo B? Sólo pueden haber dos respuestas hipotéticas: que **no existe diferencia** entre los comportamientos (A es = B) o que si existe (A es diferente a B). Claro... si no existe diferencia quiere decir que da lo mismo dar o no dar el medicamento, es decir: ique el medicamento en cuestión no sirve! La estadística analítica se dedica a analizar si al comparar una situación con otra ellas son similares o son diferentes. Por supuesto que dado que

se trabaja con muestras de la realidad, siempre existirá una posibilidad de que en sus resultados exista un error. Dado que la tabla de contingencia es la mejor herramienta para comparar, la estadística analítica usará una serie de pruebas para demostrar sus resultados. De igual manera, en las próximas líneas se intentará hacer entender cómo se usa la curva de distribución para probar las hipótesis y se hará una mención de los errores que se pueden presentar al plantear la hipótesis nula y alterna.



Ciertamente esto es una repetición del subíndice 4.3.5.2. Pero, repitiendo se aprende... ¿cierto?

1.4.1. La herramienta esencial de la comparación (Tabla de contingencia o de 2X2) y su aplicación con IC del 95% y prueba de la p

La epidemiología ha desarrollado la herramienta más útil de la comparación: se trata de la tabla de contingencia.

		VARIABLE DEPENDIENTE	
		SI PRESENTE	NO PRESENTE
VARIABLE INDEPENDIENTE	Si presente	a	b
	No presente	c	d
		a+c	b+d
		REALIDAD	
		(A es = B)	(A es diferente a B)
HIPÓTESIS NULA (A es = B)	Se acepta (A es = B)	Planteamiento correcto (A es = B)	Error tipo II (Se acepta siendo diferente)
	Se rechaza (A es diferente a B)	Error tipo I (se rechaza siendo igual)	Planteamiento correcto (A es diferente a B)
		ENFERMO	
		SI	NO
PRUEBA	POSITIVA	SI ENFERMO, PRUEBA (+) = VERDERO ENFERMO	NO ENFERMO, PRUEBA (+) = FALSO POSITIVO
	NEGATIVA	SI ENFERMO, PRUEBA (-) = FALSO NEGATIVO	NO ENFERMO, PRUEBA (-) = VERDERO NEGATIVO

La gráfica expresa tres formas diferentes de usar la tabla de contingencia o de comparación. Para relacionar variables, para probar hipótesis o para evaluar pruebas. Al fin y al cabo de lo que se trata es de aprender a través de la comparación.

La tabla como tal, compara la presencia o ausencia de una situación (variable dependiente) con exponerse a una variable independiente. Recuérdese que se compara la razón existente entre los no expuestos con los expuestos. La pregunta hipotética es si exponerse (A) es igual que no exponerse (B) o si es diferente (A es = B o A es diferente a B). Las dos medidas más utilizadas en estas comparaciones son:

Riesgo Relativo (RR):

$$RR = \frac{\frac{a}{(a+b)}}{\frac{c}{(c+d)}}$$

y el Odds Ratio (OR):

$$OR = \frac{a \times d}{b \times c}$$

En ambas medidas el referente es el 1 que siempre expone la proporción de la población total NO expuesta y que SÍ tiene la variable dependiente [—] con el total de la EXPUESTA y que SÍ tiene la variable dependiente [—] y cuyo resultado se da según sea el caso como RR o OR.

1.4.1.1. Ejemplo a través de un brote del uso de la tabla

Para entender lo interesante de la comparación y cómo permite ver, imagínense que luego de una fiesta donde estuvieron 54 personas, se presentó una enfermedad manifestada por diarrea y vómito. Es clara la sospecha de que se hayan intoxicado.

COMPARACION ENTRE COMIDA DE ENFERMOS Y DE SANOS									
CASOS	ENFERMOS				SANOS				
	ARROZ	ENSA	SALC	POST	CASOS	ARROZ	EN	SAL	POS
1	1	1	1	1	1	1	0	1	1
2	1	1	1	1	2	0	0	1	1
3	1	1	1	1	3	0	0	1	1
4	1	1	1	1	4	0	0	1	1
5	1	1	0	1	5	0	0	1	1
6	1	1	0	1	6	0	0	0	1
7	1	1	1	1	7	1	0	0	1
8	1	1	0	1	8	1	1	1	1
9	1	1	0	1	9	1	1	0	0
10	1	1	0	1	10	1	0	0	1
11	0	0	0	0	11	0	0	0	1
12	1	1	1	1	12	1	0	0	1
13	1	1	1	1	13	0	0	0	1
14	1	1	0	1	14	0	0	1	1
15	1	1	0	1	15	0	1	0	1
16	1	1	1	1	16	0	0	1	1
17	1	1	1	1	17	1	0	1	1
18	1	1	0	1	18	0	0	1	0
19	1	1	0	1	19	0	0	1	1
20	1	1	1	1	20	1	0	0	1
21	1	1	1	1	TOTAL	8	3	11	18
22	0	0	0	0	NO CONTACTO	12	17	9	2
23	1	1	1	1					
24	1	1	1	1					
25	1	1	1	1	ENFERMO	NO ENFERMO			
26	1	1	1	1	CON COMIDA	CON COMIDA			
27	1	1	0	1	SIN COMIDA	SIN COMIDA			
28	1	1	1	1					
29	1	1	1	1		30	8		
30	0	1	0	1	ARROZ	4	12		
31	1	1	1	1					
32	1	1	1	1		32	3		
33	1	1	1	1	ENSALADA	2	17		
34	0	1	0	1					
TOTAL	30	32	20	32		20	11		
NO CONTACTO	4	2	14	2	SALCHICHA	14	9		
						32	18		
					POSTRE	2	2		

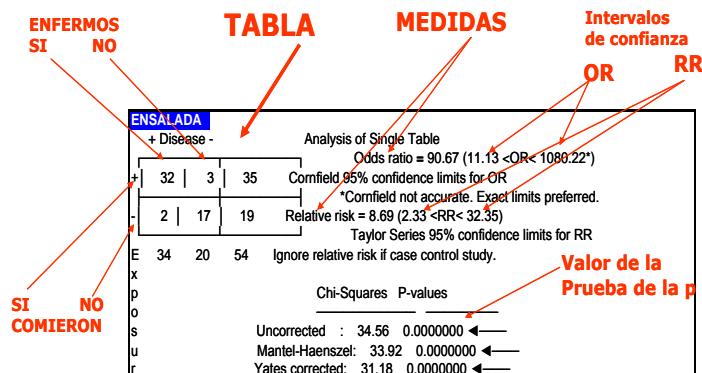
La gráfica muestra la base de datos de los sanos y enfermos y las comparaciones básicas entre los distintos alimentos. En la parte inferior derecha se compara con cada alimento quiénes se enfermaron y quiénes no.

Luego de ser atendidos los casos (que fueron 34 los que se enfermaron) en urgencias y descartar la gravedad del episodio para sus vidas, se hace necesario identificar cuál pudo ser el alimento responsable de esta intoxicación. Entonces se ha de comparar las personas que se enfermaron (casos = 34) con las que no se enfermaron (20 en este ejemplo). Ellas constituyen la variable dependiente. Y luego se deberá comparar con cada uno de los alimentos expuestos. ¿Qué se espera encontrar? Que uno de esos alimentos esté presente en los casos (más presente) y ausente en los que no se enfermaron (menos presente). Ello permite sospechar que sea el responsable. Para ello la tabla es particularmente útil.

Analice cómo se organiza la tabla de comparación: la variable independiente es en este caso arroz, ensalada, salchicha y postre, que son los alimentos consumidos, y deberán pasar cada uno como variables independientes o de exposición por los participantes en esa fiesta. Lo que se va a comparar es si comer versus no comer es un riesgo para enfermarse (variable dependiente). La pregunta es que si comer (A) da lo mismo que no comer (B) para que se presente la situación de enfermarse o lo que es lo mismo:

1. Ver si comer tal alimento (A) es = a (B) no comer tal alimento; es decir, el alimento no determina la diferencia entre enfermarse o no).
2. A es diferente a B (el alimento sí marca la diferencia entre enfermarse o no).

Lo cierto es que se deberá hacer una comparación de cada uno de los alimentos. Al usar el programa EPI INFO con uno de los alimentos, en el cual se puede hacer uso de la tabla de 2x2 y que obtiene las medidas de RR, de OR con sus respectivos intervalos de confianza y, adicionalmente con la prueba chi cuadrado, se obtuvo:



El programa EPI INFO sólo requiere que se llenen las casillas de DISEASE y EXPOSURE y arroja las MEDIDAS (OR y RR) con sus respectivos intervalos de confianza, así como el valor de la prueba p. Luego de aplicarlo y obtener las comparaciones (datos que usted puede sacar de la base de datos previa) se encontró:

ALIMENTO	MEDIDAS	PRUEBAS ESTADISTICAS	
ENSALADA		Intervalos de confianza	
OR	90,67	Limite inferior	Límite superior
RR	8,69	2,33	32,5
		valor prueba p	0,0000000
ARROZ		Intervalos de confianza	
OR	11,25	2,42	57,57
RR	3,16	1,33	7,5
		valor prueba p	0,00001779
SALCHICHA		Intervalos de confianza	
OR	1,17	0,33	4,12
RR	1,06	0,70	1,61
		valor prueba p	0,78370000
POSTRE		Intervalos de confianza	
OR	1,78	0,16	19,75
RR	1,28	0,47	3,49
		valor prueba p	0,57680000

Recuérdese que las dos medidas de asociación son el OR y el RR. El uso de una u otra depende de la modalidad de investigación. Por ejemplo, casos y controles o estudios transversales usan el OR. Las cohortes o los ensayos clínicos usan el RR. Además existe una relación entre los IC de 95% y la prueba p. Si los IC de 95% no son significativos, la prueba de la p será mayor de 0.05 (salchicha y postre) y si los IC de 95 % son significativos la prueba p será menos de 0.05 (ensalada y arroz).

El usar como medida el OR o el RR lo determina el tipo de estudio. En este caso, el **RR** es una mejor medida (por ser la incidencia lo que se busca). Se encontró que la ensalada produce la medida más fuerte con un RR de 8,69. Ello se interpreta que aquellos que comieron ensalada, se enfermaron 8,69 veces más que aquellos que se enfermaron, pero no comieron la ensalada. La conclusión es que la ensalada parece ser determinante en la aparición de la enfermedad. Pero... ¿se puede estar seguro de esta afirmación? O: ¿este resultado es SIGNIFICATIVO desde el punto de vista estadístico? Y estas preguntas se hacen porque este resultado (RR de 8,69) puede ser debido o bien:

- Realmente a la ensalada (porque está contaminada o porque está dañada en fin...)
- O al azar... queriendo decir con ello que a pesar de salir ese riesgo, no es la ensalada el alimento responsable, pues de repetirse el estudio casi con seguridad saldría otro alimento.

Lo cierto es que esa duda se plantea ante cualquier resultado: lo encontrado puede deberse o bien porque es así o por el azar. Las pruebas de significancia estadística son las que enfrentan este problema y procuran decírnos si lo encontrado es porque es así o por una causalidad del azar. Una de las maneras es usar las probabilidades...

A continuación se explicara cómo se hace esta diferenciación.

(Y se está repitiendo...)

1.4.1.1.1. El concepto de significancia a través del ejemplo

La significancia es un término estadístico que implica tener la certeza científica de que la manera como se interpreta un dato es válida (real desde el punto de vista de la estadística). Se usa constantemente en las investigaciones para comparar dos poblaciones (por ejemplo) en donde se debe tener la certeza de que ambas son iguales para poder ser comparables o cuando se encuentra el resultado de la intervención y se hace indispensable comparar lo hallado en un grupo u otro y saber si existe o no **DIFERENCIA SIGNIFICATIVA**.



Ciertamente esos dos usos son con mucho los más frecuentes al menos en el contexto de las investigaciones. El comparar dos grupos de población y comparar dos resultados, resultan ser en la práctica de la biometría los campos donde se aplican con mayor frecuencia estas pruebas de significancia.

El concepto de significativo implica (didácticamente) que, bajo la situación del ejemplo, de 100 veces que se analice la ENSALADA al menos en 95 saldrá como alimento con fuerte riesgo de ser el responsable de provocar la enfermedad (y que la posibilidad de que el resultado sea debido al azar es de mucho menos de 5 veces de 100). Según los resultados de los **Intervalos de Confianza**, por cada vez que no coman ensalada y se enfermen, por comer ensalada se enfermarían entre 2.33 y 32.5 veces (según el RR). Esto expresa dónde

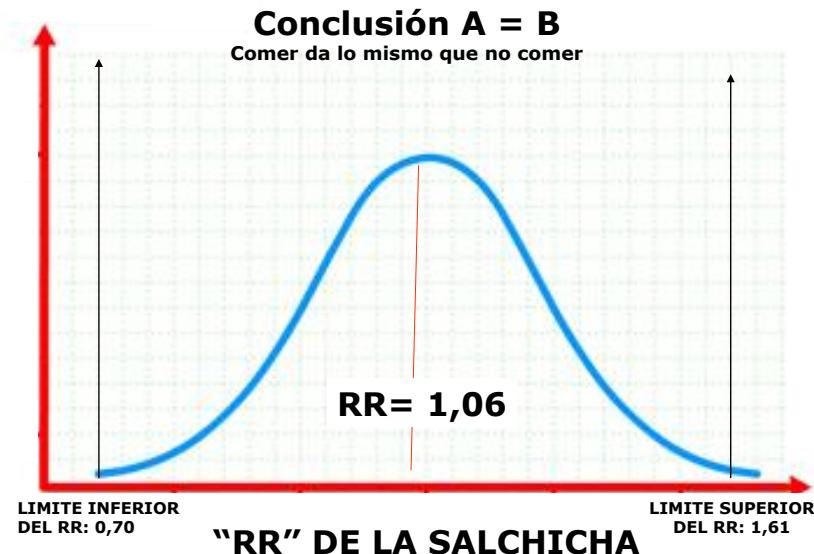
se ubican los Intervalos de Confianza del 95% y lo hace evidentemente en el terreno del riesgo. Según la **prueba de la p**, comer ensalada hace que se enfermen con una probabilidad de 100 (total de probabilidades) – 0.00000000 (resultado de la prueba de la p le da al azar) es decir, ciento por ciento seguro (más adelante se explica).

Expresado con un contraste de hipótesis es responder a la pregunta: ¿Comer ensalada (A) comparado con no comer (B) está relacionado o no con la enfermedad? Las únicas respuestas pueden ser:

- A=B quiere decir que no hubo diferencia en la aparición de la enfermedad entre los que comieron ensalada y no comieron. Por ende, el RR muy seguramente es muy cercano a 1 y los IC de 95% han de tener su límite inferior menor de 1 y el límite superior mayor de 1. Correspondiendo a esto, de seguro una prueba de la p sitúa el resultado mayor de 0.05.
- En contraste de lo anterior y como es este el caso, A (relación entre ensalada y la enfermedad) es diferente a B (no comer ensalada y su relación con la enfermedad) y, por tanto, los RR muestran unos límites siempre en el riesgo y una p menor de 0.05 (la situación A es significativamente diferente a B).

Esperamos que quede entendido que tanto los IC del 95% como la prueba p son dos formas de interpretar lo mismo. Ambas son pruebas de SIGNIFICANCIA y juntas se correlacionan. Si los IC del 95% expresan que existe diferencia significativa, entonces la prueba de la p es menos de 0,05 y si la prueba de p es mayor de 0,05 entonces se interpreta como que no existe diferencia significativa y los límites de los IC de 95% de seguro cruzan el 1.

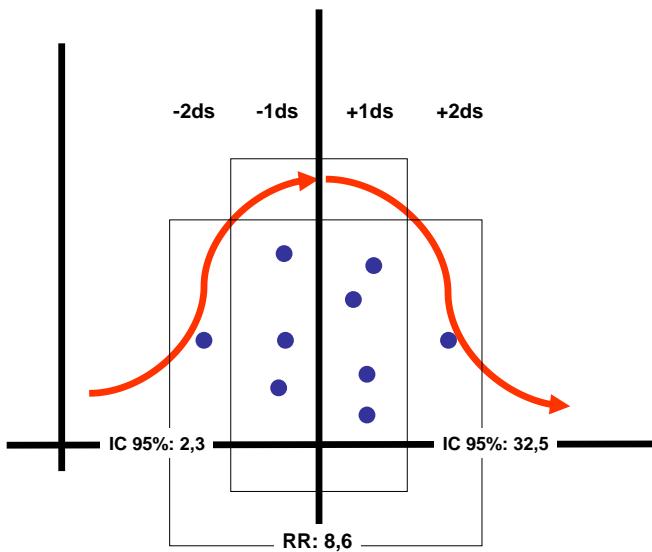
PRUEBA DE LA p PARA LA SALCHICHA =0,78 que quiere decir que la posibilidad de que la salchicha este "relacionado" con la enfermedad es más de 0,05 lo que es decir que comieron salchicha "igual" los que se enfermaron como los que no.



La gráfica expresa la situación de la salchicha en este ejemplo: al comparar los que enfermaron con los que no y que comieron o no salchicha, se obtuvo un RR de 1,06 (un "poquitico" de riesgo). Sin embargo, al ver la distribución de la población (en un 95%) se pudo ver cómo hubo personas que incluso parece que la salchicha las protegiese de la enfermedad... ¿Por qué? Pues porque el límite del RR de IC del 95% estuvo por debajo de 1 (en este caso 0,70). El límite superior estuvo en RR de 1,61. La interpretación es que la salchicha unas veces se comporta como riesgo y otras como protección. Claro, lo esperado (para decir que es o riesgo o protección) es que lo que sea riesgo sea siempre riesgo y lo que sea protección siempre lo sea (o que los límites de los IC de 95% ambos queden más de 1 o que ambos queden menos de 1); por consiguiente, la conclusión es que **no influye** en la enfermedad o lo que es lo mismo que el RR no es SIGNIFICATIVO. Si se usa la prueba de la p (otra manera estadística de hacer lo mismo) y dado que el resultado es de 0,78, se puede sacar la misma conclusión: comer o no comer salchicha da lo mismo en torno a la enfermedad. En hipótesis, $A = a$, B .

Entonces de esto se concluye que los intervalos de confianza son un concepto que surge de la curva de distribución normal. Se obtienen a través de procedimientos estadísticos específicos.

Volviendo al ejemplo (y repitiendo) es obvio que en una situación como la descrita, cada alimento debe ser confrontado entre enfermos y no enfermos para así poder identificar el alimento sospechoso de la enfermedad. Al preguntarles a las personas, sus respuestas permitirán obtener los datos (RR, en este caso). Al distribuir esas respuestas (datos) se forma una curva de distribución.



La gráfica muestra una curva de distribución normal, con los sitios donde se ubican los datos (y el RR) y muestra los límites en que se encuentra el 95% de los datos.

- En la ensalada, el límite inferior daría un RR de 2,3 y el límite superior daría un RR 32,5 (usando el programa la prueba de Cornfield). La interpretación de la **SIGNIFICANCIA** en el **RR** y en el **OR** es que los intervalos con sus límites o:
 - a) queden ambos en el riesgo (más de 1)
 - b) queden ambos en la protección (ambos menos de 1).

Ello es lo mismo que decir que estadísticamente siempre que se comió ensalada, las personas se enfermaron con un **riesgo relativo** entre 2,3 hasta 32,5 (siempre riesgo). Entonces en este ejemplo existe SIGNIFICANCIA del riesgo.

- El arroz (ver figura) tuvo un RR de 3,16 con límite inferior del IC 95% de RR 1,33 y límite superior de RR 7,5. Por ende, y al igual que la ensalada, el RR es SIGNIFICATIVO (también es un alimento RIESGO tal vez menos que la ensalada, pero ciertamente riesgo).
- En la salchicha el RR NO ES SIGNIFICATIVO.
- En el postre el RR es de 1,28 con un IC de 95% entre 0,47-3,49, por lo cual tampoco es significativo.

La conclusión es que la ensalada y el arroz mostraron SIGNIFICANCIA estadística... claro, más la ensalada o lo que es lo mismo, que éstos alimentos están relacionados significativamente con la enfermedad.

4.4.1.1.1.2. Ampliando la prueba p

Otra manera de ver la significancia es por la prueba de las posibilidades o prueba p. Vale la pena recordar que una probabilidad o posibilidad es la certeza que se tiene que suceda algo. Se expresan desde el 0 hasta el 1 donde, uno es igual al 100% de certeza de que suceda.

- Una **probabilidad (p)** de 0,5 quiere decir que existe un 50% de posibilidades que suceda. Por ende, existe un 50% de posibilidades de que no suceda (q).
- Una probabilidad de 0,05 quiere decir que existe un 5% de que suceda y un 95% de posibilidades de que no suceda que es = q .
- Y una probabilidad de 1 quiere decir que existe un 100% de posibilidades que suceda (certeza absoluta). Por tanto, la posibilidad que no suceda es 0 (q). Claro $p + q$ es = 1 y es la fórmula básica de las posibilidades. Entonces, y en el ejemplo (repitiendo):

ALIMENTO	MEDIDAS	PRUEBAS ESTADISTICAS	
		Intervalos de confianza	
ENSALADA		Límite inferior	Límite superior
OR	90,67	11,3	1082
RR	8,69	2,33	32,5
		valor prueba p	0,0000000

Al aplicar una prueba estadística para determinar la posibilidad de que la ensalada sea el alimento riesgo se encontró el valor de p , igual a $1 - \text{el valor de } p$ que sale en la prueba, por lo que p es = $1 - 0.000$; es decir, es 1. Entonces se interpreta así:

- a. Se parte de: Posibilidad es uno entonces $1 = p + q$.
- b. Si p es = 1 entonces: $1 - p$ es = q . es decir q es = 0
- c. Por ende, $p = 1$ (certeza total) - 0 (posibilidad que no sea) = 1
- d. Si 1 es = 100% la posibilidad de que la ensalada sea un alimento riesgo es del 100%.

Al comparar el resto de alimentos se encuentra:

ALIMENTO	MEDIDAS	PRUEBAS ESTADISTICAS	
ARROZ	Intervalos de confianza		
	Limite inferior	Límite superior	
OR	11,25	2,42	57,57
RR	3,16	1,33	7,5
	valor prueba p	0,00001779	
SALCHICHA	Intervalos de confianza		
	Limite inferior	Límite superior	
OR	1,17	0,33	4,12
RR	1,06	0,70	1,61
	valor prueba p	0,78370000	
POSTRE	Intervalos de confianza		
	Limite inferior	Límite superior	
OR	1,78	0,16	19,75
RR	1,28	0,47	3,49
	valor prueba p	0,57680000	

En el caso del arroz $p = 1 - \text{valor de } p$ encontrado estadísticamente = 0.99. Si 1 es 100% de probable que el arroz sea de riesgo, entonces 0.99 será que el arroz tiene un riesgo de 99%. Como es más de 95% entonces el arroz también es de riesgo.

En el caso de la salchicha, valor de p es = 0.78; entonces, p es = $1 - 0.78 = 0.22$ o lo que es lo mismo que existe un 22% de posibilidades que la salchicha sea un alimento riesgo. Por ende, estadísticamente no es riesgo.

En el caso del postre el riesgo es de 43%. Dado que para que sea SIGNIFICATIVO la p ha de ser igual o menos de 5% (o en número igual o menos de 0.05) se considera que estos dos alimentos no presentan un riesgo SIGNIFICATIVO).

Todo lo anterior sólo quiere decir que de los cuatro alimentos, la ensalada presenta una fuerte relación con la enfermedad, que el arroz también aunque menos fuerte y que se deben orientar las acciones de búsqueda del agente causal en la ensalada y en el arroz. ¿Pueden dos alimentos ser los responsables de la enfermedad? Claro que sí... explicarlo ya será cuestión de revisar los hechos, pero ciertamente es posible.



Una explicación didáctica de las probabilidades, de la significancia y de las pruebas de hipótesis, es la encontrada en el curso de estadística de la Universidad de Málaga del Dr. Francisco Javier Barón. Dice así:

Imagínese que tiene fuertes sospechas que una moneda esté cargada para que sólo salga cara cuando se lance al aire y caiga. Los planteamientos hipotéticos son así: cara es = A y sello es = B. La posibilidad que al lanzarla salga cara o sello es igual de 0.5 respectivamente. La hipótesis nula al lanzar la moneda es que la posibilidad de que salga cara o sello es igual. Entonces la hipótesis nula es A es = B (A y B tienen la misma p de salir). Lo que se planteó al comienzo es que A es diferente a B (al decir que está cargada, la posibilidad de que salgan más caras es mayor. Por ejemplo, de un 95% o más). ¿Cómo se prueba una u otra hipótesis? Sencillo... se ha de lanzar la moneda al aire y se pregunta:



1. ¿Qué posibilidad que al lanzarla salga cara? Pues sin duda la probabilidad es de un 50% (en el lenguaje de probabilidades: p de 0.5).
2. Pero si se vuelve a lanzar, ¿qué posibilidad existe de que vuelva a salir cara? Ahora es de un 25 % (p de 0.25 producto de 0.5 x 0.5) Si sale cara... si bien es

- sospechoso, no lo es como para asegurar que está cargada, ya que no cumple con el requisito de tener una $p = 0$ menor de 0.05.
3. Y al lanzarla por tercera vez, ¿qué posibilidad existe de salga nuevamente cara? Ya acá la posibilidad de que salga cara es sólo de un 6.25 % (p de 0.062) resultado de multiplicar el 0.25×0.25). La sospecha de que esté cargada es grande si llega a salir cara nuevamente (la posibilidad de que no saliera (q) sería de un 93.8% resultado de restar $100\% - 6.25\%$).
 4. ¿Y qué posibilidad que al lanzarla por cuarta vez, vuelva y salga cara? Ahora la probabilidad de que sea por el azar se ha reducido a un 0.39 % (p de 0.0039 = 0.062×0.062). Si sale cara, prácticamente ya no tenemos duda de que está cargada (la posibilidad de que esté cargada es = $100\% - 0.39\% = 99.6\%$).
 5. Y qué posibilidad existe de que salga cara si se lanza por quinta vez. Es tan sólo de una p de 0.0000015 (resultado de multiplicar 0.0039×0.0039). O sea que la posibilidad de que esté cargada es de un 99, 99984%.

Entonces lo que se plantea desde las hipótesis es que A (probabilidad que salga cara) es diferente a B (probabilidad que salga sello) con una p de 0.0039 que se da a partir de la cuarta vez que da cara de manera consecutiva. En general, y cuando se aplican las pruebas estadísticas que arrojan las probabilidades, se dice que A es = a B si la p es mayor de 0.05 y cuando es igual o menos de 0.05 es SIGNIFICATIVAMENTE DIFERENTE.

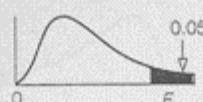
Obviamente las pruebas de hipótesis se usan con muchísima frecuencia cuando se estudian los medicamentos. Se analiza en concreto si su uso provoca en las poblaciones diferencias significativas cuando se compara con los placebos. Lo esperado es que exista diferencia estadística (A es diferente a B) en torno a las medidas (con sus respectivos datos).

Pero, además, las probabilidades se usan muchísimo para establecer probabilidades de riesgo. ¿Qué probabilidad existe de morir por una anestesia? Y esta probabilidad se incrementa por el tipo de cirugía que se vaya hacer... calcularlo es algo que se hace cada vez más y hasta el momento los seguros son los que más lo han desarrollado. En este caso la p en vez de multiplicarse se suma (se suma el riesgo de morir en anestesia + el riesgo derivado de la cirugía).

4.1.1.1.3. La prueba Ji (Chi cuadrado)

Como se debe haber dado cuenta, son muchos los fenómenos en la medicina que ameritan ser comparados. De acuerdo a las medidas y a sus correspondientes datos empleados, existen pruebas estadísticas unas mejores que otras. La prueba de chi cuadrado se usa con medidas de categorías. Su fórmula es igual a la suma del dato esperado menos el dato observado, todo ello elevado al cuadrado y el resultado dividido entre el dato esperado. El resultado se deberá observar en la tabla de distribución estadística del chi cuadrado, (teniendo en cuenta para ello los grados de libertad), y comparando el dato obtenido de la fórmula con el que sale en la tabla se afirma:

- Si el dato obtenido es mayor (al correspondiente a p de 0,05) que el que aparece entonces existe diferencia significativa.
- Si el dato obtenido es menor al número que aparece en la tabla, entonces se afirma que NO EXISTE DIFERENCIA SIGNIFICATIVA.



	Grados de libertad para el numerador															
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59
4	7.71	6.94	6.59	6.39	6.26	6.15	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.43
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69
50	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50
*	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39

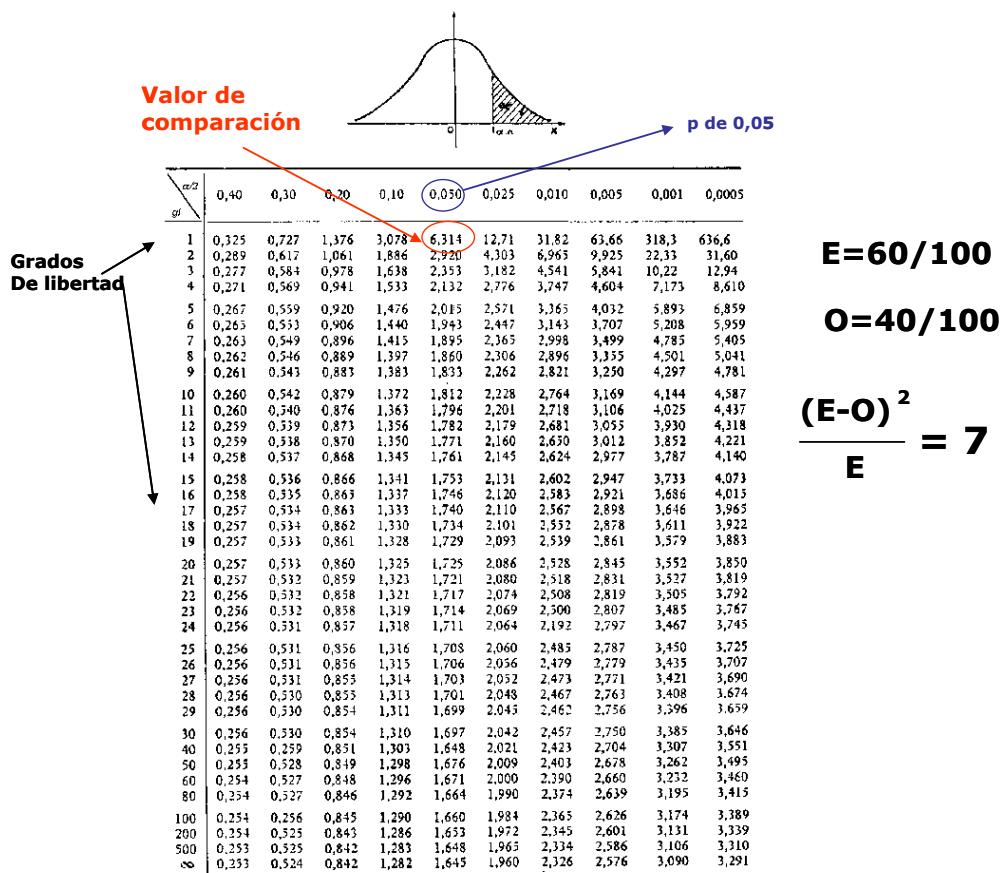
Las tablas estadísticas aparecen al final de los libros (de estadística). De acuerdo a la prueba que se emplee se usa la tabla de referente.

A manera de ejemplo imagínese que se quiere saber si un club de hipertensos ha logrado incrementar la adherencia de su población al tratamiento. Para ello se va a comparar con pacientes hipertensos que no acuden al club. Entonces se consiguen 100 hipertensos del club de los cuales 60 mostraron la medida de adherencia. Y de 100 hipertensos que no van al club se encontró que 40 cumplían con la medida de adherencia. Es evidente la diferencia clínica (60 en los del club y 40 en los que no van) pero, la pregunta es: ¿Esta diferencia es SIGNIFICATIVA o es el resultado del azar? Entonces se aplica la prueba chi cuadrado. Lo esperado es 60, lo encontrado fue 40. El resultado de aplicar la fórmula fue: 7. Al buscar el referente en la tabla del chi cuadrado con una p de

0,05 se encuentra que es 3,96. Por ende, el 7 es mayor con lo que se puede afirmar que esta diferencia es significativa y se concluye que el club de hipertensos en este caso aumenta la adherencia (y que la diferencia encontrada no fue resultado del azar).

NOTA: Si bien este texto no se detendrá a explicar cómo se usan las tablas estadísticas, el concepto que espera quede en el lector, es cómo se interpreta.

4.1.1.1.4. Entendiendo la prueba T



Esta es otra prueba muy utilizada para saber si existen diferencias significativas entre dos observaciones. Imagine que se va hacer una investigación en una población y la tensión arterial es una de las medidas y dos investigadores (A y B) serán los encargados de la toma de la tensión arterial. La pregunta obvia que se debe hacer antes de

medir la TA es: ¿Existirán diferencias en la toma de TA entre A y B? Resulta evidente su importancia: A y B deben tomar la TA igual y obtener los mismos resultados cuando le tomen la TA a unas mismas personas. De no ser así, no se puede estar seguro de lo que se vaya a obtener en la investigación. Entonces, y para responderla lo que se hace es:

1. Se escoge un pequeño grupo de personas, ideal que sea con las características de la población a estudiar.
2. Ambos A y B se ponen de acuerdo en cómo y con qué van a tomar la TA.
3. Luego A les toma la TA a las personas. Obtiene el promedio de la sistólica y diastólica.
4. Luego lo hace B (a las mismas personas).
5. Ahora debo comparar que las tomas de A y B son estadísticamente similares o, lo que es lo mismo, que no tienen DIFERENCIA SIGNIFICATIVA. Acá es donde se aplica la prueba T que es:

1. La comparación

Población	Investigador A		Investigador B	
	Sistólica	Diastólica	Sistólica	Diastólica
1	130	70	130	70
2	120	80	130	80
3	140	90	140	90
4	125	85	120	90
5	120	80	130	75
6	140	60	130	65
7	150	80	140	80
8	120	80	130	80
9	130	80	130	80
10	120	70	120	75
Promedios	129,5	77,5	130	78,5

¿Será que existe diferencia Entre investigador A con B?

2. Aplicando la prueba T

Sistólica dm A	Diastólica dm A	
	Varianza	Varianza
0,5	0,25	-7,5
-9,5	90,25	2,5
10,5	110,25	12,5
-4,5	20,25	7,5
-9,5	90,25	2,5
10,5	110,25	-17,5
20,5	420,25	2,5
-9,5	90,25	2,5
0,5	0,25	2,5
-9,5	90,25	-7,5
	102,25	66,25
DS A	10,1118742	DS A 8,1394103
Prueba t	0,02472341	Prueba t 0,12285902

- 2.1. Se obtiene la desviación media de A.
- 2.2. Se obtiene la varianza de A.
- 2.3. Se saca la Des Estándar de A.
- 2.4. Se aplica la formula
- 2.5. Se busca el número de la prueba T en la tabla estadística T y se interpreta

No existe diferencia significativa entre El A y el B en la toma de tensión

Nota: Si número es = a más o menos 6,31, la diferencia es significativa. Si es Menos no existe diferencia

Prueba T = la media de A menos la media de B elevado al cuadrado y el resultado dividido por la Desviación Estándar de A. El número que dé se compara en la tabla estadística t y ahí se sabrá si existe o no diferencia entre A y B.



En este ejemplo el investigador A tomó la TA a diez personas y registró la sistólica y diastólica. Obtuvo el promedio. Lo propio hizo el investigador B. Luego se obtuvo: la desviación media de la sistólica en el investigador A (sistólica de la primera persona menos promedio y así sucesivamente). Luego obtuvo la varianza y con ella obtuvo la Desviación Estándar. De esa manera obtuvo todos los elementos para poder comparar a saber:

1. Promedio de A.
2. Promedio de B.
3. Desviación estándar de A.

Aplica la fórmula y obtuvo en el caso de la TA sistólica un número de 0,024. Para la diastólica obtuvo un número de 0,12. Esos resultados se buscan en la tabla estadística de la t (claro, se debe saber buscar) y si estos números son mayores que los referenciados en la tabla, entonces se dice que EXISTE DIFERENCIA SIGNIFICATIVA. Si son menores se afirma que NO EXISTE DIFERENCIA SIGNIFICATIVA. Más como un dato para abreviar (y no mirar la tabla) se afirma que si el resultado de la prueba t es más o menos de 6,3 existe DIFERENCIA SIGNIFICATIVA. Si es menos no existe. En el caso entre A y B no existe...

1. La comparación

Población	Investigador A		Investigador C	
	Sistólica	Diastólica	Sistólica	Diastólica
1	130	70	150	90
2	120	80	160	90
3	140	90	140	90
4	125	85	140	90
5	120	80	150	100
6	140	60	130	100
7	150	80	140	90
8	120	80	130	80
9	130	80	130	80
10	120	70	120	70
Promedios	129,5	77,5	139	88

¿Será que existe diferencia Entre investigador A con C ?

2. Aplicando la prueba T

Sistolica dm A	Varianza dm A	Diastolica dm A	Varianza dm A
0,5	0,25	-7,5	56,25
-9,5	90,25	2,5	6,25
10,5	110,25	12,5	156,25
-4,5	20,25	7,5	56,25
-9,5	90,25	2,5	6,25
10,5	110,25	-17,5	306,25
20,5	420,25	2,5	6,25
-9,5	90,25	2,5	6,25
0,5	0,25	2,5	6,25
-9,5	90,25	-7,5	56,25
	102,25		66,25
DS A	10,1118742	DS A	8,1394103
Prueba t	8,92515058	Prueba t	13,5452073

Si existe diferencia significativa entre El A y el C en la toma de tensión

Ahora imagine que se trae a un investigador C y surge, por supuesto, la misma pregunta. Dado que A y B son iguales da lo mismo comparar a C con alguno de ellos dos. Se aplican todos los pasos para la prueba t y se obtienen los números del círculo. Dado que son mayores de 6,5 entonces SI EXISTE DIFERENCIA EN LAS CIFRAS OBTENIDAS ENTRE A y C. Ello es clave porque si no lo hacen de manera similar, no podremos estar seguros que las diferencias en los resultados sean por la investigación o por esta diferencia.

La prueba t se aplica en numerosos procesos de **estadardización** o **calibración**. En los estudios pilotos se usa para probar instrumentos. En los laboratorios se usa para calibrar los equipos.

4.4.2. El error tipo I y tipo II

Hasta el momento se ha intentado explicar cómo la estadística (a través de la curva de distribución, por ejemplo) es útil para demostrar la hipótesis nula ($A = B$) o la hipótesis alterna (A es diferente a B). Pero, puede suceder que se equivoque. Es decir que resulte:

1. Que A es diferente a B cuando realmente es igual (error alfa o tipo I). Esto se conoce como rechazar la hipótesis nula cuando ella es cierta (o aceptar la alterna cuando no es verdadera).
2. Que A es = a B cuando realmente no lo es (error beta o tipo II). Esto se conoce como aceptar la hipótesis nula cuando es falsa (o rechazar la hipótesis alterna cuando ella es verdadera).

¿Por qué aparecen los errores tipo I y tipo II? Pueden aparecer como consecuencia de fallas en el manejo de la **información** recibida, por **selección** inapropiada de la población o sencillamente por la **casualidad**. Obviamente los dos primeros aspectos dependen del investigador. El tercero del azar.

		REALIDAD	
		(A es = B)	(A es diferente a B)
HIPOTESIS NULA (A es = a B)	Se acepta (A es = B)	Planteamiento correcto (A es = B)	Error tipo II (Se acepta siendo diferente)
	Se rechaza (A es diferente a B)	Error tipo I (se rechaza siendo igual)	Planteamiento correcto (A es diferente a B)

		ENFERMO	
		SI	NO
PRUEBA	POSITIVA	SI ENFERMO, PRUEBA (+) = VERDERO ENFERMO	NO ENFERMO, PRUEBA (+) = FALSO POSITIVO
	NEGATIVA	SI ENFERMO, PRUEBA (-) = FALSO NEGATIVO	NO ENFERMO, PRUEBA (-) = VERDERO NEGATIVO

Al observar el planteamiento de la hipótesis nula a través de esta tabla se observa dónde caen los errores. Tal vez una forma práctica de observar la aplicación de esta tabla es a través del cuadro de pruebas diagnósticas:

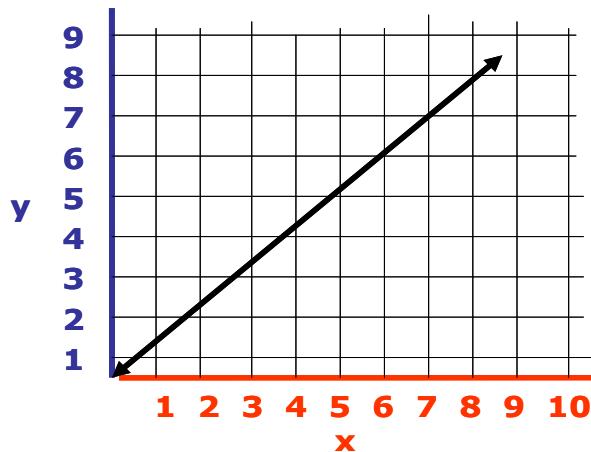
El falso negativo y el falso positivo expresan los errores tipo I y tipo II.

Son muchas las posibilidades de que aparezcan los errores. El más frecuente en torno a medicamentos es afirmar que sirve cuando no sirve o que no sirve cuando realmente sirve. Pero, probablemente donde más se ven (de forma muy evidente) en el mundo de la medicina sea en el uso de las pruebas diagnósticas. A la larga, los seres humanos nos la pasamos planteado hipótesis y con base en ellas tomamos las decisiones.

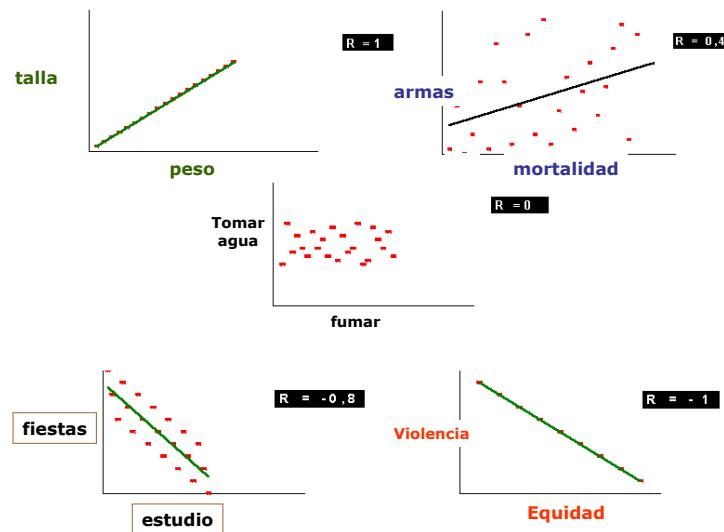
Ciertamente estos errores se observan en investigaciones. Con el fin de ponerlas bajo el contexto de la clínica unos ejemplos pueden ser:

- Un ejemplo de A es diferente a B cuando realmente son iguales** puede ser el siguiente: la prueba de tuberculina detecta anticuerpos contra la TBC. Si es positiva el sujeto ha tenido o tiene contacto con el bacilo de la TBC. Entonces A es = contacto con TBC. Por ende, al aplicar la prueba a un sujeto con TBC (B es = a sujeto con TBC) ha de dar positiva. Por ende B es = contacto con TBC. Por ello A es = B. Pero Por ejemplo, en muchos casos de TBC miliar la prueba de tuberculina sale negativa (A es diferente a B), a pesar de que el individuo está lleno de bacilos tuberculosos. Hoy sabemos que en tal entidad (TBC miliar) el sujeto padece de una inmunosupresión que impide respuestas de los anticuerpos.
- En torno a decir que A es diferente a B cuando en realidad no lo es.** Por ejemplo, cuando afirmamos que un sujeto tiene sífilis (A es = sífilis) por una serología positiva (B es = serología positiva) y la realidad es que no tiene. ¿Por qué sale la prueba positiva? Una explicación puede ser porque tiene una enfermedad inmunológica que falsea la prueba. Acá decimos que A es = sífilis y B es = a serología positiva. Entonces la hipótesis nula es A es = B (que es lo se deduce de la prueba), pero la realidad es que no.

4.4.3. Las correlaciones



Resultan muy evidentes las relaciones que existen entre distintas características. Por ejemplo, la relación entre comer y peso, o entre fumar y cáncer del pulmón. Algunas de estas relaciones son tan sólidas que si se incrementa (o se disminuye) una de las características, se incrementa (o se disminuye) la otra. En otros casos la relación es inversa. Entre más de esto menos de aquello, pero ciertamente sólida ya que si se afecta una se afecta la otra. Estas relaciones entre variables es lo que se conoce con el nombre de correlaciones y su estudio es particularmente importante pues a través de ello se pueden tomar decisiones. Por ejemplo, terapéuticas que modifican sustancialmente el pronóstico.



Estadísticamente existen muchas pruebas para analizar si existe o no correlación entre dos variables. Sus resultados se expresan del 0 donde no existe correlación hasta el 1 donde existe una correlación perfecta. Las gráficas de dispersión con (dos variables) son la forma de ver si existen visualmente correlaciones.

El eje de la X y el eje de la Y representan dos variables y se quiere saber si están correlacionadas. Donde dice $R = 1$ la correlación es directa: a más X, más Y, lo es de manera proporcionada. Ese R se obtiene a través de las pruebas estadísticas de correlación. Un ejemplo puede ser la correlación entre talla y peso en la infancia, donde a más talla, más peso (claro que llega una edad en donde no se crece más y el peso sí se puede aumentar). La siguiente gráfica con un R de 0,4, expresa que, si bien tienen relación, esta no es tan proporcionada como la anterior. Por ejemplo, armas de fuego y mortalidad. Sin duda existe una relación, pero no siempre guarda proporción directa. USA tal vez es el país con más armas de fuego y su mortalidad no es mayor a la de otros países con menos armas. Cuando el R es = 0 el diagrama de dispersión es como se ve ahí. Por ejemplo, relación entre tomar agua y fumar... no existe relación. Si uno aumenta el otro no varía. Relaciones negativas puede ser la establecida entre fiestas y estudio. Entre más fiestas menos estudio. Por ejemplo, en los jóvenes de los colegios. El R es = -0,8. Y la relación entre equidad con la violencia puede establecer una correlación negativa perfecta donde a más equidad menos violencia ($R = -1$).

Una fórmula para establecerla es: La sumatoria del producto de multiplicar la desviación media (DM) de una variable con la desviación media de la otra variable, y este resultado dividido por el producto de la raíz cuadrada originado con la suma de la desviación media de una variable al cuadrado con la de la otra variable al cuadrado.

$$\text{Correlación lineal} = \frac{\sum(DM_x * DM_y)}{\sqrt{\sum DM_x^2} * \sqrt{\sum DM_y^2}}$$

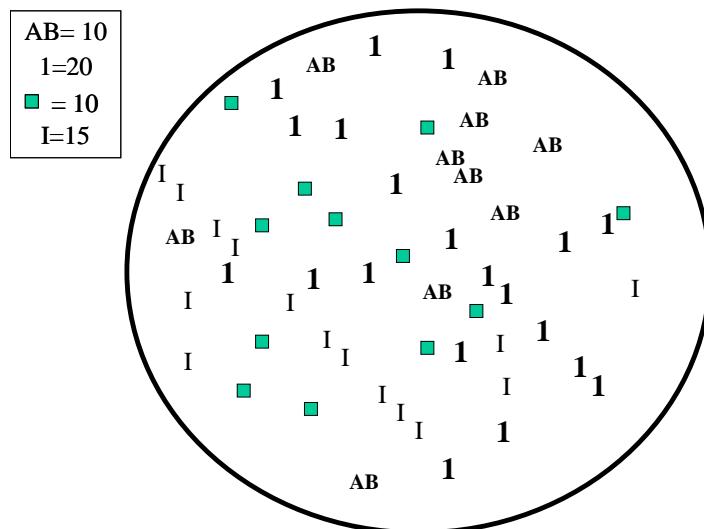
La búsqueda de correlaciones y su interpretación resultan particularmente importantes en el mundo de la salud. Es importante *no incurrir en el error de asignarle a las correlaciones una interpretación causal*. La causalidad se deriva de otras interpretaciones y, si bien las correlaciones puedan dar ideas, es peligroso (por la inmensa posibilidad de cometer errores) concluir causalidad. Por ejemplo, el texto de Norman y Streiner menciona que algún investigador encontró una correlación entre hipolipemia y cáncer de estomago. Dedujo que no se debían bajar los niveles de triglicéridos y así se podría prevenir el cáncer, pero luego descubrió que es el cáncer el que puede bajar los niveles de lípidos.

4.4.4. La inferencia estadística

Inferir es convertir un resultado investigativo en una generalización. Por ejemplo, y con base al caso citado previamente de los antirretrovirales en el SIDA, es concluir que dado que fueron útiles en la población de esta investigación (muestra), lo será en todos los pacientes con SIDA (población). De hecho el acto médico de la intervención busca predecir que si los enfermos se toman los medicamentos, se puede predecir (inferir) que se mejorarán. Obviamente las inferencias parten de la certeza que la población de la investigación (muestra) representa al resto de la población que presenta la circunstancia y, por lo tanto, los resultados encontrados en el grupo investigado se pueden extrapolar al resto de población que esté bajo las mismas circunstancias. La estadística es particularmente útil para ello. La inferencia estadística parte de dos componentes:

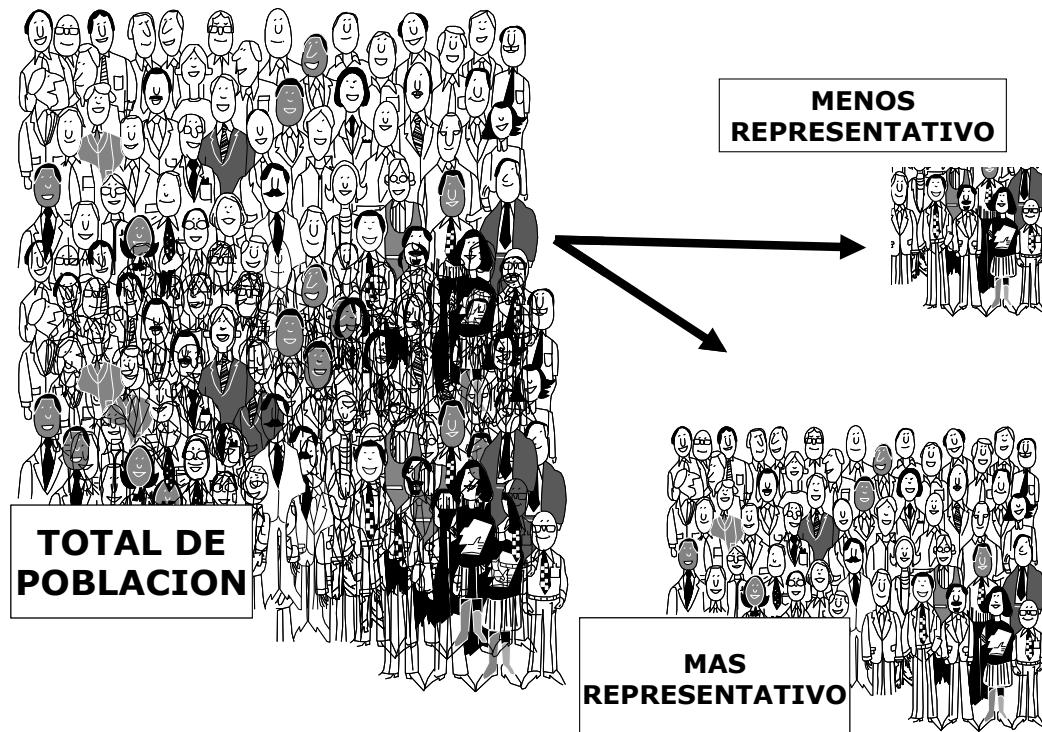
1. El primero es que la muestra con que se trabaja en la investigación ciertamente representa al total.
2. El segundo es que el número utilizado en la investigación permite calcular el grado de error. Al respecto, es obvio que entre menor sea la distancia entre el número con que se trabaja en una muestra y el total, más pequeño será el error probable. Esto se conoce como potencia.

A manera de ejemplo:



Imagínese que el círculo con su contenido representa el total de una población con sus respectivas características. El cuadro describe cuántos AB, hay, así como cuántos 1, cuántos cuadros y cuántos I. Una muestra representativa deberá tener todos estos componentes los cuales además, deberán estar en una misma proporción (más 1 que AB, por ejemplo). Para que sea representativa se deberá introducir por ejemplo, el total de características en una bolsa, revolverla y sacar el número (uno por uno) correspondiente de características. Así se les permitió a todas esas características la misma probabilidad de salir. El número a sacar se obtiene a través de fórmulas, calculando el margen de error. Claro... siempre existirá un margen de error, el que se disminuye en la forma que el número de la muestra se acerca al total de las características y, por ende, al total de la población.

Es por ello que los ensayos clínicos o los ensayos de campo, en donde se selecciona la población de cada grupo (el de intervención y el de comparación) de manera aleatoria, son ideales para hacer inferencias. La potencia la dará el tamaño de la muestra.



En los estudios de prevalencia, donde se trabaja con muestra, el hacer una selección aleatoria de la población objeto es lo que permite que lo que se encuentre en la muestra se pueda generalizar al total.

Los diseños cualitativos renuncian a las inferencias. Ellos pretenden esencialmente entender fenómenos y no buscan predecir.

Lo cierto es que existen numerosas pruebas estadísticas en las investigaciones cuantitativas. Existen pruebas para probar las hipótesis. Pruebas para seleccionar muestras y pruebas para hacer inferencias. De acuerdo al diseño empleado o a las medidas utilizadas se usarán unas u otras. Un buen investigador debe hacer una selección apropiada de su población objeto, así como pensar en las características que ha de tener una muestra para ser apropiada. Luego podrá solicitar la ayuda a un profesional estadístico quien le podrá recomendar qué prueba(s) es o son la o las mejores a utilizar.

Es evidente que uno de los intereses de los investigadores que usan diseños cuantitativos es aplicar sus resultados al total de la población objeto. Por ejemplo, se hace un estudio sobre un grupo de pacientes con SIDA, pues se anhela poder aplicarlo a todos. Ello forma parte de la predicción, donde lo encontrado acá se puede predecir como útil en todos. Para poderlo hacer se ha de tener en cuenta numerosos elementos. Uno de ellos es el tamaño de la muestra o población en la que se va desarrollar la investigación. Resulta obvio que siempre existe un margen de error como resultado de esa diferencia entre el trabajo con la muestra y el total de casos. Una explicación es que puede ocurrir que en la muestra falte población que tenga alguna característica que tienan el total de población. Esa falta puede ser o porque en la muestra no está una representación adecuada del total de población o porque la entidad tiene muchas características distintas en distintas personas (variabilidad). Claro, esto se corrige a través del tamaño de la muestra (entre más grande, menor margen de error). También se ha de tener en cuenta el universo o población de referente para el estudio.

Suponiendo que se va hacer un estudio de intervención (probar un medicamento) en personas con SIDA y se escogen personas afroamericanas, ¿se podrá inferir los resultados a la población africana? Interesante... ¿verdad? Al respecto, la respuesta sería sí, siempre y cuando las características de la población sean parecidas. Por ejemplo, que el estado inmunológico y nutricional (variables muy importantes en esta enfermedad) de la población muestra de la investigación sea parecido al de la población africana.

A manera de conclusión...



Este texto espera haber logrado acercar a la realidad (y a la práctica) los conceptos de la epidemiología, la investigación y la estadística. Si bien, y para entender, se separan y se explican dichos conceptos, la realidad es que la forma como se presentan en el mundo práctico es integral. Muchas definiciones en este texto se han ejemplificado en el mundo de la medicina, pero igual se podría haber escogido el mundo de la enfermería o el de cualquier área de la salud. Ciertamente de lo que se trata es de aplicar una forma de pensar en donde, y en particular, a través de las líneas se ha insistido en el método científico.

Recuerden que este texto ha sacrificado muchos conceptos buscando la didáctica. No pretende competir con los sólidos textos de epidemiología o estadística. De hecho, pretende ser un primer referente y que a lo mejor aquellos que nos lean, puedan luego adentrarse en textos más completos.

A continuación vienen unos anexos en donde se han construido ejercicios de acuerdo a las áreas o disciplinas, procurando crear un contexto que permita entender.

Esperamos que el texto haya contribuido en el proceso de formación en que, sin duda, están los lectores... los invitamos a que nos retroalimenten a través del correo

jucagonqui@yahoo.es.

Lecturas recomendadas

1. Beaglehole R, Kjellstrom T. Epidemiología Básica. Publicación científica 551. OPS 1994.
2. Álvarez F, Álvarez A. Investigación y Epidemiología. ECOE Ediciones 1998. Cali Colombia.
3. Back C, Llopis A, Najera E, Terris M. El desafío de la epidemiología. Problemas y lecturas seleccionadas. Publicación científica número 25. Editorial de la OPS 1988.
4. Lapierre D. Más grandes que el amor. Editorial Planeta Seix Barral Vigésima segunda edición. 1999.
5. Pollak K. Los discípulos de Hipócrates. Una historia de la medicina. Círculo de Lectores S.A. España 1970.
6. Vásquez L, Romero M, León R. Contreras. Introducción a la bioestadística y a la epidemiología. 1998. Editorial Mcgraw-hill Interamericana de Venezuela, S.A.
7. Falbo GH, Buzzetti R, Cattaneo A. El homicidio entre los niños y adolescentes: estudio de casos y controles en Recife (Brasil). Bulletin of the World Health Organization. 2001; 79 (1).
8. Ardila E, Sánchez R, Echeverry J. Estrategias de investigación en medicina clínica. Manual Moderno 2001.
9. Mac Mahon B, Pugh T. Principios y métodos de la epidemiología. La prensa médica mexicana. Segunda edición 1975.
10. Rey Calero J. Método epidemiológico y salud de la comunidad Editorial. McGRAW-HILL Interamericana de España. Impresión de 1992.
11. López MS, Garrido LF, Hernández AM. Desarrollo histórico de la epidemiología: su formación como disciplina científica. Salud pública de México, 2000; 42(2).
12. Polit D, Hungler B. Investigación científica en ciencias de la salud. Año 2000 sexta edición. McGraw-Hill Interamericana.
13. Colimon KM. Fundamentos de epidemiología. Ediciones Díaz de Santos SA. 1990. Madrid (España).
14. Moreno AA, López MS, Corcho BA. Principales medidas en epidemiología. Salud pública de México, 2000; 42(4).
15. Barón FJ. Bioestadística, métodos y aplicaciones. Curso virtual de la Universidad de Málaga.
16. González JC, De la Hoz F. Determinación de relaciones entre los comportamientos de riesgo psicosociales y la familia de la población adolescente escolar de los colegios públicos de Suba. <http://www.bdigital.unal.edu.co/657/>.
17. González JC, Benavides R, Bonilla O. Determinación de las probabilidades de apendicitis en pacientes con dolor abdominal que acudieron a la CJNC en el año 2002. Carta Comunitaria FUJNC. Número 85 Junio-Julio del 2006.