



REDLIC Red Editorial
Latinoamericana de
Investigación Contemporánea

Lo esencial en **BIOESTADÍSTICA MÉDICA**

Fausto Bolívar Maldonado Reyes, Robert Iván Álvarez Ochoa, Gabriela Cordero Cordero,
Andrés Alexis Ramírez Coronel, Miguel Ángel Capote Llanares

REDLIC | Red Editorial
Latinoamericana de
Investigación Contemporánea

Lo esencial en
BIOESTADÍSTICA MÉDICA

2024

Lo esencial en **BIOESTADÍSTICA MÉDICA**

Fausto Bolívar Maldonado Reyes, Robert Iván Álvarez Ochoa,
Gabriela Cordero Cordero, Andrés Alexis Ramírez Coronel,
Miguel Ángel Capote Llanares

Indexaciones



Bases de datos



Licencia



FICHA TÉCNICA

Título: Lo esencial en
BIOESTADÍSTICA MÉDICA

Autores: Fausto Bolívar Maldonado Reyes, Robert Iván Álvarez
Ochoa, Gabriela Cordero Cordero, Andrés Alexis Ramírez
Coronel, Miguel Ángel Capote Llanares.

Prólogo: Dr. Wilber Adolfo Gómez Vargas, PhD

Primera edición: marzo de 2024

e-ISBN: 978-9942-659-00-2

DOI: <https://doi.org/10.58995/lb.redlic.33>

Link: <https://redliclibros.com/index.php/publicaciones/catalog/book/33>

Producción editorial y coordinación técnica

© Red Editorial Latinoamericana de Investigación
Contemporánea REDLIC S.A.S. (978-9942-7063)

📍 Avenida 3 de noviembre y segunda Transversal

🌐 www.editorialredlic.com

✉ rev.investigacioncontemporanea@gmail.com

✉ contactos@editorialredlic.com

☎ 098 001 0698

Coordinador editorial:

Dra. Marcia Iliana Criollo Vargas, PhD

Diseño de portada: [Creative](#)

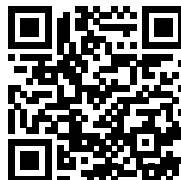
Diseño y diagramación: [Creative](#)

Obra sometida al arbitraje por pares dobles ciego
externos a la editorial.



Este material está registrado bajo licencia Creative Commons International, con permiso para reproducirlo, publicarlo, descargarlo y/o distribuirlo en su totalidad únicamente con fines educativos y/o asistenciales sin ánimo de lucro, siempre que se cite como fuente al autor (es) y a la Red Editorial Latinoamericana de investigación contemporánea.

Cuenca-Ecuador



DOI



AUTORES:

Fausto Bolívar Maldonado Reyes, Robert Iván Álvarez
Ochoa, Gabriela Cordero Cordero, Andrés Alexis
Ramírez Coronel, Miguel Ángel Capote Llanares

¿CÓMO CITAR?

Maldonado Reyes FB, Álvarez Ochoa RI, Cordero Cordero G, Ramírez Coronel AA, Capote Llanares MÁ. Lo esencial en BIOESTADÍSTICA MÉDICA. Cuenca, Ecuador: Red Editorial Latinoamericana de Investigación Contemporánea; 2024. 232 p.

AVISO LEGAL

IMPORTANTE

Los contenidos presentados en este libro son para fines informativos y educativos. Los autores, editores y publicadores no se hacen responsables de cualquier consecuencia derivada del uso o interpretación de la información presentada en este libro.

Contenido

Prólogo.....	13
Introducción.....	15

Capítulo I

MEDIDAS EPIDEMIOLÓGICAS. ANÁLISIS CUANTITATIVO DEL PROBLEMA DE SALUD EN LA POBLACIÓN19

Conceptos	20
A. Proporción.....	20
B. Tasa.....	21
C. Razón	23
Medidas epidemiológicas.....	24
A. Prevalencia:.....	24
B. Tasa de prevalencia (TP).....	24
C. Incidencia.....	26
D. Tasa de incidencia (TS).....	26
E. Tasa de incidencia acumulativa.....	27
Relación entre medidas epidemiológicas	28
Tasa de prevalencia y tasa de incidencia	28
Patrón de enfermedades en una comunidad.....	29
Características de la población	30

Capítulo II

VALIDEZ Y TEST O PRUEBA DE REPRODUCIBILIDAD 41

Sensibilidad y reproducibilidad de un test o prueba	43
Valores predictivos	47
Influencia de la tasa de prevalencia sobre el valor predictivo de una prueba	49
Reproducibilidad de un test.....	56
Conclusión	61

Capítulo III

DISEÑOS DE ESTUDIO..... 69

Estudios experimentales	70
Estudios no experimentales.....	72
Estudios observacionales.....	73
A.- Estudios epidemiológicos descriptivos	73
B.- Estudios epidemiológicos analíticos	74
Estudios no experimentales.....	75
A.- Estudios transversales	75
B.- Estudios longitudinales (estudios de seguimiento).....	76
C.- Caso control	78

Capítulo IV

ESTADÍSTICA DESCRIPTIVA GENERALIDADES.....85

PARTE I

Enfoque exploratorio de los datos85

Variables87

La escala de medidas..... 88

PARTE II

Estadística descriptiva de una dimensión92

Métodos gráficos.....92

Presentación de datos cualitativos..... 94

 Diagrama de barras 94

 Diagrama de sector o de torta96

Presentación de datos cuantitativos (numéricos) 98

 El histograma105

 El polígono de frecuencia.....106

A.- Descripción de los datos cualitativos.....109

 1.- Datos nominales109

 2.- Datos ordinales.....110

B.- Organización de los datos cuantitativos110

 1.- Parámetros de tendencia central111

 A.- La media (la media aritmética)111

 B.- La mediana114

 C.- La moda.....116

2.- Parámetros de dispersión.....	117
A.- El rango.....	119
B.- Los cuartiles (y rango del intercuartil)	119
C.- La varianza y la desviación estándar.....	121

PARTE III

Análisis de los datos explorados.....	131
Técnica de Tallo y Hoja.....	132
Gráfico de Caja.....	136

PARTE IV

ESTADÍSTICA DESCRIPTIVA. DESCRIPCIÓN

ESTADÍSTICA DE DOS VARIABLES.....	143
Relación entre diferentes variables.....	143
Análisis de dos variables cualitativas.....	145
Asociación entre una variable cualitativa y una variable cuantitativa	149
Asociación entre variables cuantitativas.....	151
Transformación	154

Capítulo V

ESTADÍSTICA ANALÍTICA. INFERENCIA	161
Principios de inferencia estadística.....	161
Inferencia estadística y validación.....	161

Punto de estimación y estimación del intervalo de confianza.....	163
Estructura de la hipótesis.....	164
Distribución simple en forma teórica	168
Intervalo de confianza.....	170
Límite de confianza unilateral.....	173
¿Cómo enfocar un análisis estadístico?	174
Punto de estimación = test estadísticos.....	176
Z-test.....	180
Alternativa del Z-test: prueba chi cuadrado de Pearson χ^2	183
Test de Mantel-Haenszel χ^2	186

Capítulo VI

Introducción a la psicometría.....	189
Fiabilidad de la escala.....	189
Coeficientes alfa y omega.....	191
Error típico de medida.....	195
Calcular la puntuación observada.....	196
Dimensionalidad.....	201
Análisis factorial exploratorio	201
Supuestos.....	205
Rotación factorial	206
Análisis de contenido.....	214
Análisis factorial confirmatorio (AFC).....	216
Referencias	222

PRÓLOGO

La bioestadística médica es el lenguaje que habla la evidencia en el campo de la medicina. Es la herramienta que nos permite comprender, interpretar y aplicar los datos que emergen de la compleja red de información que caracteriza a la investigación en salud. Este libro es una guía indispensable para aquellos que buscan navegar por este vasto océano de datos con destreza y comprensión.

En un mundo donde la toma de decisiones clínicas y la formulación de políticas de salud dependen cada vez más de la evidencia científica, la bioestadística se erige como el cimiento sobre el cual se construye el conocimiento médico. Este libro no sólo proporciona las herramientas necesarias para entender los principios fundamentales de la bioestadística, sino que también explora cómo estas herramientas pueden ser aplicadas de manera práctica en la investigación médica.

A lo largo de estas páginas, el lector será guiado desde los conceptos más básicos hasta los métodos más avanzados, siempre con un enfoque claro en la relevancia clínica. Se abordarán temas que van desde la recopilación y descripción de datos hasta el diseño y análisis de estudios epidemiológicos y ensayos clínicos. Se destacarán casos prácticos y ejemplos reales para ilustrar cómo la bioestadística influye en la toma de decisiones médicas y en la mejora de la atención al paciente.

Este libro es el resultado de la colaboración de expertos en bioestadística y profesionales de la salud, unidos por la convicción de que el conocimiento estadístico es esencial para una práctica médica basada en la evidencia. Esperamos que este viaje a través de la bioestadística médica sea tan educativo como emocionante, y que al final, los lectores se sientan capacitados para explorar, entender y contribuir al vasto campo de la investigación médica.

Dr. Wilber Adolfo Gómez Vargas, PhD.

Universidad de Antioquia (Colombia)

INTRODUCCIÓN

La bioestadística médica es una disciplina fundamental en el ámbito de la investigación y la práctica clínica, proporcionando las herramientas estadísticas necesarias para comprender, analizar y aplicar los datos en el contexto de la medicina y la salud. Este libro busca ser un compañero indispensable para estudiantes, investigadores y profesionales de la salud, brindando una sólida base en los principios estadísticos aplicados a la biomedicina.

En el complejo entramado de la medicina moderna, la toma de decisiones basada en la evidencia se ha convertido en el pilar sobre el cual se sustentan los avances científicos y las mejores prácticas clínicas. La bioestadística surge como la brújula que guía a los profesionales de la salud en la interpretación y aplicación de los datos, permitiéndoles discernir entre la casualidad y la causalidad,

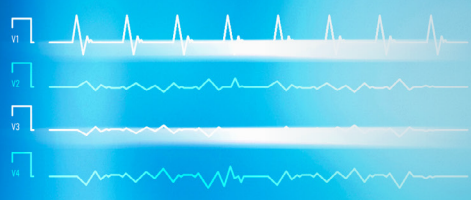
identificar patrones y asociaciones, y evaluar la eficacia de intervenciones médicas.

En cada capítulo, se presenta la relevancia directa de la bioestadística en la toma de decisiones médicas. Además, se destacarán las consideraciones éticas y prácticas que deben guiar la aplicación de estas herramientas estadísticas en el contexto de la investigación médica.

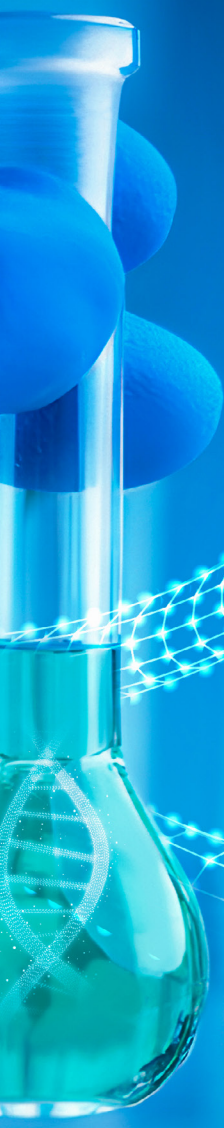
Al sumergirnos en las páginas de este libro, extendemos una invitación al lector para descubrir el poder transformador de la bioestadística médica y adquirir las habilidades esenciales necesarias no sólo para interpretar críticamente la literatura científica, sino también para diseñar estudios sólidos y contribuir de manera significativa al progreso de la medicina basada en evidencia. Desde los conceptos fundamentales hasta las aplicaciones más avanzadas, este libro aborda la bioestadística médica con un enfoque claro y práctico. A lo largo de sus páginas, exploraremos cómo diseñar estudios robustos, recopilar datos de manera efectiva y utilizar técnicas estadísticas pertinentes para

responder preguntas clínicas, impulsando así el avance del conocimiento médico.

¡Bienvenidos a un viaje apasionante en el mundo de la bioestadística médica!



10 mm/s 10 mm/mV 0.05-35



Capítulo I

MEDIDAS EPIDEMIOLÓGICAS

ANÁLISIS CUANTITATIVO DEL PROBLEMA DE SALUD EN LA POBLACIÓN

INTRODUCCIÓN

Los instrumentos de las medidas epidemiológicas son capaces de ayudarnos a cuantificar aspectos de la frecuencia de una enfermedad, por lo tanto, permiten la total descripción y comprensibilidad de su importancia en la salud pública.

CONCEPTOS

a. PROPORCIÓN

Es la relación entre el número de individuos que poseen una cierta característica y el total de la comunidad (Moreno-Altamirano et al., 2000):

$$\text{Proporción} = a / (a+b)$$

a: el número de individuos que poseen una característica o atributo

b: el número de individuos que no poseen el mismo atributo

a + b: el total de la población bajo nuestro estudio

Propiedades:

$$0 \leq a / a+b \leq 1$$

- El numerador está incluido en el denominador
- Si esta condición no existe, la proporción es = 0 entonces (a = 0)
- Si todos los miembros de la población tienen el atributo, la proporción es = 1 (b = 0)
- Generalmente es expresado como porcentaje = $(a / a+b) * 100$

Ejemplo:

En una comunidad con una población de 434 habitantes, 27 personas tienen colitis amebiana el 1 de enero del 2022.

Entonces:

$$27 / (27 + 407) * 100 = 6,2 \%$$

Interpretación: El resultado indica que el 6,2% de la población tiene colitis amebiana el 1 de enero del 2022. Esto significa que, de los 434 habitantes, 27 personas padecen esta enfermedad en ese momento.

b. TASA

La cantidad de cambios de un evento sobre el tiempo (unidad de tiempo).

En epidemiología los cambios de un evento relatan los nuevos casos (tasa de natalidad, tasa de morbilidad, tasa de mortalidad, etc). El numerador establece la frecuencia absoluta simple de casos del estudio, mientras que el denominador está formado por la sumatoria de los períodos individuales de riesgo a los que han estado expuestos los individuos susceptibles de la población de estudio (Moreno-Altamirano et al., 2000).

Tasa de incidencia de una enfermedad:

$$\text{Tasa de incidencia} = \frac{\text{Número de nuevos casos}}{\text{Población en riesgo}} \times \text{Periodo bajo observación}$$

Propiedades

El valor mínimo es = 0

El valor máximo es =/+

El valor de la tasa depende del tiempo de observación

El numerador está incluido en el denominador

Ejemplo: ¿Cuál es la tasa de mortalidad por COVID-19 en el municipio de Azogues en el año 2020?

Entonces:

Tasa de incidencia = número de muertes durante 2020 / (población en riesgo a 30 de junio) x (10ⁿ)

Tasa de incidencia = 125/13.450 x 100.000 Hab = 929,36 por cada 100.000 Hab

Interpretación: El resultado indica que la tasa de mortalidad por COVID-19 en el municipio de Azogues en el año 2020 fue de 929,36 por cada 100.000 habitantes. Esto significa que, durante ese año, hubo 125 muertes por COVID-19 en una población en riesgo de 13.450 habitantes. Esta tasa de incidencia expresa la cantidad de muertes por COVID-19 por cada 100.000 habitantes en el municipio.

c. RAZÓN

Es igual a la relación entre dos unidades, cada una de ellas posee diferentes cualidades (Moreno-Altamirano et al., 2000):

$$\text{Razón} = x / y$$

Propiedades:

El numerador no está incluido en el denominador

La razón es utilizada en las siguientes situaciones:

a) X y Y se refieren al mismo evento de salud en dos distintos grupos

Razón de género= $\frac{\text{\# de pacientes masculinos con SIDA/población masculina}}{\text{\# de pacientes femeninas con SIDA/población femenina}}$

También puede encontrarse:

Razón de género = $\frac{\text{\# de casos de género masculino}}{\text{\# de casos de género femenino}}$

b) X y Y se refieren a dos períodos de tiempo

Razón de género= $\frac{\text{\# de pacientes masculinos con SIDA/población masculina en 2004}}{\text{\# de pacientes femeninas con SIDA/población femenina en 2004}}$

c) X y Y se refieren a igual período de tiempo y población, pero no tiene un valor real

$$\frac{(\# \text{ de niños muertos por sarampión vistos en una comunidad en 2004})}{(\# \text{ de casos de mortalidad esperados sin campaña de vacunación})}$$

En salud pública uno puede encontrar diferentes usos de la razón como:

Camas/médicos, auxiliares/doctor, población/unidades de salud, etc.

MEDIDAS EPIDEMIOLÓGICAS

a. PREVALENCIA:

Número de casos que existen en cierto momento dado en una población determinada.

Se entiende por casos, todos los casos nuevos y viejos. Por ejemplo: la prevalencia de sarampión significa el número de casos totales que cumplen con el criterio diagnóstico de sarampión (Bonita et al., 2008; Lozano Poveda, 2012).

b. TASA DE PREVALENCIA (TP)

Tasa de Prevalencia= $\frac{\text{número de casos en un momento dado}}{\text{población total}}$

Ejemplo:

Número de pacientes con sarampión en la comunidad "X" al 1/12/2021= 74

Población total de la comunidad "X" al 1/12/2021= 324

$$TP = 74/324 = 0,23 \times 100 = 23 \%$$

Interpretación: El resultado indica que en la comunidad "X" al 1 de diciembre de 2021, el 23% de la población total ha sido diagnosticada con sarampión. Esto sugiere que existe un brote de sarampión en la comunidad, ya que el porcentaje de pacientes afectados es bastante alto en comparación con la población total. Es importante tomar medidas preventivas y de control para evitar la propagación del virus y proteger a la población.

Propiedades de la tasa de prevalencia:

- Es una medida de probabilidad. En este contexto existe la oportunidad de encontrar algo positivo para la enfermedad estudiada en un tiempo dado.
- Es una buena medida para planear intervenciones, al menos para enfermedades crónicas (da una estimación de la frecuencia relativa de un problema de salud).
- No es el mejor instrumento para evaluar el impacto de un programa de salud.

c. INCIDENCIA

Número de nuevos casos presentados durante un cierto periodo de tiempo en una población dada.

En Javier Loyola existieron 27 nuevos casos de sarampión el 1/01/1987 y el 31/12/1987

Por lo tanto, sólo existen 27 casos nuevos sin un denominador.

d. TASA DE INCIDENCIA (TS)

$$TS = \frac{\text{número de nuevos casos}}{\text{población en riesgo} \times \text{periodo de observación}}$$

La tasa de incidencia depende de una unidad de tiempo y esta no puede ser más que uno.

Ejemplo: ¿Cuál es la tasa de incidencia de diarrea en niños menores de 5 años en un Centro de Salud de Javier Loyola durante un año, considerando que se registraron 792 nuevos episodios de diarrea en un total de 100 niños seguidos durante este periodo?

$$TS = 792 / 100 \text{ niños} = 7,9 \text{ episodios de diarrea por niño.}$$

Interpretación: La tasa de incidencia de diarrea en niños menores de 5 años en el Centro de Salud de Javier Loyola es de 7,9 episodios por niño por año. Esto significa que, en promedio, cada niño menor de 5 años en este centro de salud experimenta aproximadamente 8 episodios de diarrea por año. Esta información es importante para evaluar la carga de enfermedad y la necesidad de intervenciones de salud pública en esta población específica.

e. TASA DE INCIDENCIA ACUMULATIVA (TIA)

$$\text{TIA} = \frac{\text{número de nuevos casos en un periodo de tiempo}}{\text{población total en riesgo}}$$

- El numerador es el número de pacientes afectados (casos) durante el periodo de observación.
- El denominador es la población en riesgo al inicio del periodo de observación.

TIA, indica el riesgo dado a cada una de las personas de la población de estudio durante el periodo de observación.

La tasa de ataque es una específica tasa de incidencia acumulativa, limitada al periodo del tiempo, durante el cual la población está expuesta a cierto riesgo.

Ejemplo: en una comunidad de 500 personas, se han registrado 30 casos nuevos de una enfermedad respiratoria en el transcurso de un año. ¿Cuál es la tasa de incidencia acumulativa de esa enfermedad respiratoria por 1000 habitantes en esa comunidad?

En este caso, la tasa de incidencia acumulativa sería:

Tasa de incidencia acumulativa = $(30 \text{ casos} / 500 \text{ personas}) \times 1000 = 60 \text{ casos por cada } 1000 \text{ personas}$

Interpretación: La tasa de incidencia acumulativa de la enfermedad respiratoria en esa comunidad es de 60 casos por cada 1000 habitantes. Esto significa que, en promedio,

60 personas de cada 1000 habitantes de la comunidad han desarrollado la enfermedad respiratoria durante el transcurso de un año.

Esta medida es útil para comprender la carga de la enfermedad en la comunidad y puede ser utilizada para implementar medidas preventivas y de control.

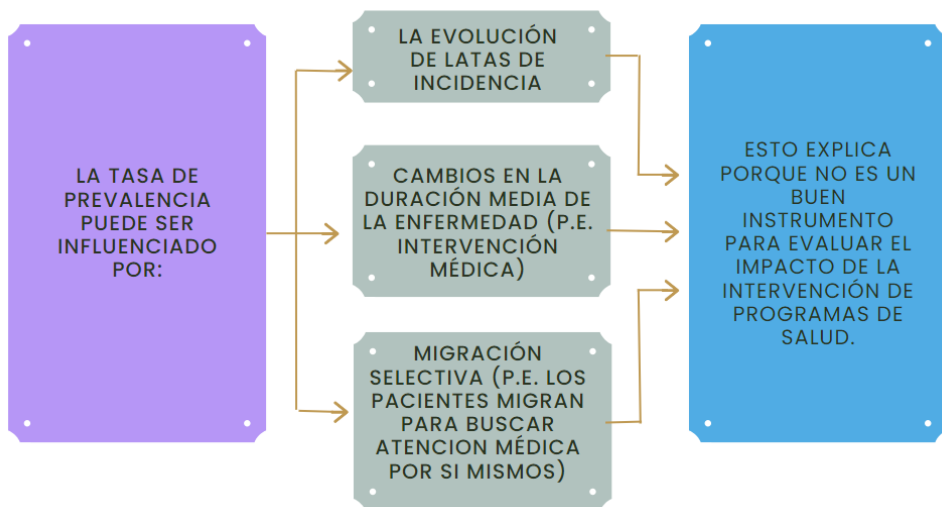
RELACIÓN ENTRE MEDIDAS EPIDEMIOLÓGICAS

TASA DE PREVALENCIA (TP) Y TASA DE INCIDENCIA (TI)

Se utiliza en casos raros de presentación de enfermedades y es igual a:

Tasa de prevalencia = tasa de incidencia x el tiempo medio de duración de la enfermedad.

Cuando dos de los tres parámetros son conocidos, el tercero puede ser encontrado. En la práctica este es usado calculando la tasa de incidencia cuando todos los casos presentados han sido registrados, la población es conocida y estable, y un estimativo de la duración media de la enfermedad es disponible.



PATRÓN DE ENFERMEADES EN UNA COMUNIDAD

La epidemiología descriptiva no contiene por sí misma una descripción de la frecuencia con que ocurre una enfermedad en una comunidad (prevalencia, tasa de prevalencia, incidencia, tasa de incidencia, etc.), por lo que la distribución de la enfermedad en una comunidad bajo estudio puede demostrar un patrón que está ligado a ciertas características (Organización Panamericana de la Salud, 2011).

Estas características son clasificadas como:



CARACTERÍSTICAS DE LA POBLACIÓN (PERSONAS)

Cada persona difiere de otra por factores endógenos (edad, género, raza) y por factores exógenos (situación familiar, condición socioeconómica, religión, etc.).

EDAD

La ocurrencia de infección, enfermedad o muerte está fuertemente asociada con la edad. La asociación es fuerte y para comparar correctamente dos poblaciones, uno necesita estandarizar la variable edad.

La edad puede ser definida cronológicamente o puede ser evaluada por criterio funcional. Cronológicamente la edad es ampliamente aplicada, pero en la práctica a veces no siempre se obtienen los datos precisos de la edad (Romero Sandoval et al., 2015).

GÉNERO

En algunas situaciones la ocurrencia de enfermedad es diferente de acuerdo al género. La proporción de hombres y mujeres en una población puede variar según la edad en diferentes grupos (nuevos nacimientos, niños jóvenes, adultos, tercera edad). Estas diferencias deben ser mantenidas presentes en la mente cuando se quiera interpretar los resultados globales (Romero Sandoval et al., 2015).

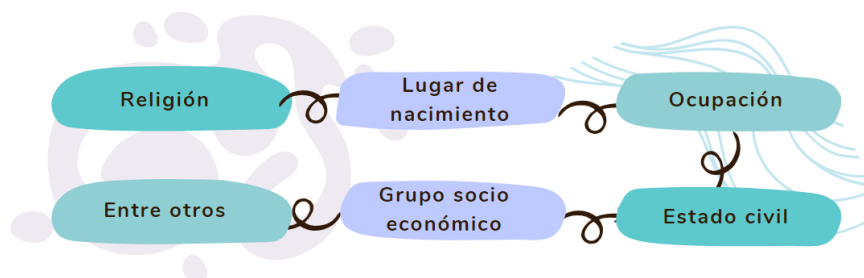
RAZA

Los factores hereditarios de riesgo relacionados a la raza pueden favorecer ciertas enfermedades. Es importante tomar esto en consideración cuando se conduce un estudio de epidemiología descriptiva. Es difícil entender qué problemas de salud están directamente ligados a la raza, debido a que frecuentemente es más fácil ligarlo con el estado socioeconómico, la cultura, la educación, etc. (Romero Sandoval et al., 2015).

GRUPO ÉTNICO

Desde el punto de vista epidemiológico, un grupo étnico es una reunión de individuos quienes han convivido juntos bajo las mismas condiciones de un particular medio ambiente, suficientemente largo, para haber adquirido las mismas características, tanto sociales como biológicas. En el estudio epidemiológico de las enfermedades y los servicios de salud es importante anotar el origen étnico (Romero Sandoval et al., 2015).

OTROS FACTORES



CARACTERÍSTICAS DEL LUGAR

Lugar, en términos geográficos, se refiere a la latitud y longitud, pero este también puede ser definido por las condiciones medio ambientales y sus habitantes.

La distinción entre el área urbana y rural es un elemento esencial para la descripción epidemiológica y/o la investigación. El lugar de origen es importante en algunas investigaciones, aunque a veces es difícil definirlo. La noción de lugar es particularmente importante en la vigilancia epidemiológica (Martínez Hernández, 2013).

CARACTERÍSTICAS DEL TIEMPO

Cambios temporales en el punto de un fenómeno de salud, a través de la variabilidad en el tiempo pueden ser causas medibles.

La ocurrencia en la variación del fenómeno puede ser.

- No periódico
- Periódico
- Cambios regulares

NO PERIÓDICO. Por lo general, se observa en forma de una epidemia o pandemia. Se refiere a un aumento notable por encima de los niveles "normales" de incidencia o prevalencia de una enfermedad en una población.

PERIÓDICO: Se refiere a la periodicidad de las funciones fisiológicas del individuo. Uno debe conocer esta periodicidad y tomar las medidas preventivas para evitar complicaciones mayores (Martínez Hernández, 2013).

CAMBIOS O TENDENCIAS: Es la evolución de un fenómeno sobre un largo período de tiempo (décadas, generaciones o centurias).

Como ejemplos podemos citar: la declinación de la mortalidad infantil en Europa, la desaparición del sarampión. Las causas de estos cambios seculares son frecuentes y están interrelacionados, por ejemplo: cambios en las condiciones económicas, en las condiciones sanitarias, en el estado nutricional, etc. El conocimiento y tener en cuenta las tendencias es importante cuando los impactos de los programas de salud están por ser evaluados (Martínez Hernández, 2013).

CONCLUSIONES

Los esenciales conceptos han sido descritos y deberán ser conocidos:

- Prevalencia y tasa de prevalencia
- Incidencia, tasa de incidencia, tasa de incidencia acumulativa

En próximos ejercicios, estas herramientas facilitarán una evaluación más precisa de un problema de salud y ayudarán a identificar el grupo de la población afectado por un fenómeno.

EJERCICIO

En agosto de 1995, se realizó una primera investigación en una comunidad rural con el objetivo de conocer el número de casos de parasitosis en dos zonas de la región, cuyos datos se muestran en la tabla 1.

La estructura de la población estudiada se encuentra en la tabla 2.

- **Pregunta 1:** ¿Qué medidas epidemiológicas pueden ser encontradas de los datos presentados?
- **Pregunta 2:** ¿Qué información ellos proveen?

Para comprender el estudio del problema en el tiempo y el número de nuevos casos detectados en la población total entre 7/95 y 5/97, **véase la tabla 3.**

En la **tabla 4 se observa la evolución de la estructura o grupos de la población,** durante el mismo periodo.

- **Pregunta 3:** ¿Cuál es la incidencia y la tasa de incidencia en ambas zonas?
 - a.- Por trimestre (sobre el total del periodo de observación)
 - b.- Por el año 1986
- **Pregunta 4:** ¿Hay diferencia entre la tasa específica de incidencia por comunidad y/o por trimestre?
- **Pregunta 5:** ¿A qué conclusiones operacionales puede usted llegar después de este análisis?

Para estudiar más ampliamente el impacto de la parasitosis en el estado de salud de la población; fueron recolectados los datos con los resultados sobre la mortalidad específica de la parasitosis en la **tabla 6**. Por su parte, la **tabla 7** presenta la composición o estructura de la población al final de junio de 1996.

- **Pregunta 6:** ¿Qué medidas epidemiológicas pueden ser calculadas con estos datos?
- **Pregunta 7:** ¿Podría esto ser de interés para conocer el número de casos por zona?
- **Pregunta 8:** ¿Su análisis permite identificar el alto riesgo de los grupos? (Compare su respuesta con la respuesta de la pregunta 2)

TABLA 1

NÚMERO DE CASOS DE PARASITOSIS DETECTADOS EN LA PRIMERA RECOLECCIÓN DE DATOS.

AGOSTO - 1995

EDAD EN AÑOS	ZONA A	ZONA B	AMBAS ZONAS
0	10	7	17
1 - 9	72	54	126
10 +	78	63	141

TABLA 2

**ESTRUCTURA DE LA POBLACIÓN DE LAS ZONAS
INVESTIGADAS**

AGOSTO – 1995

EDAD EN AÑOS	ZONA A	ZONA B	AMBAS ZONAS
0	452	520	972
1 – 4	1967	2248	4215
5 – 9	1954	2239	4193
10 – 19	2761	3158	5919
20 – 29	1750	2000	3750
30 – 39	1330	1525	2855
40 – 49	737	844	1581
50 – 59	520	596	1116
60 +	433	496	929
TOTAL	11904	13626	25530

TABLA 3

**NÚMERO DE TRIMESTRE DE LOS NUEVOS CASOS DE
PARASITOSIS EN LA POBLACIÓN TOTAL**

1995 - 1997

EDAD EN AÑOS	ZONA A	ZONA B	AMBAS ZONAS
3/95	4371	3950	8321
4/95	5030	5416	10446

1/96	4272	3512	7784
2/96	3607	2143	5750
3/96	3740	2569	6309
4/96	4369	4285	8654
1/97	3758	3525	7283
2/97	3485	2336	5821

TABLA 4

EVOLUCIÓN DE LA POBLACIÓN EN EL MEDIO TRIMESTRE DURANTE EL PERIODO DE ESTUDIO

1995 - 1997

EDAD EN AÑOS	ZONA A	ZONA B	AMBAS ZONAS
3/95	11904	13626	25530
4/95	12008	13745	25753
1/96	12113	13865	25978
2/96	12219	13986	26205
3/96	12326	14108	26434
4/96	12434	14231	26665
1/97	12543	14356	26898
2/97	12563	14482	27133

TABLA 5

**NÚMERO DE MUERTES ATRIBUIBLES A PARASITOSIS
(AMBAS ZONAS)**

1/1/1996 – 31/12/1996

EDAD EN AÑOS	ZONA A + ZONA B
0	18
1 – 4	26
5 – 9	8
10 +	25
TOTAL	77

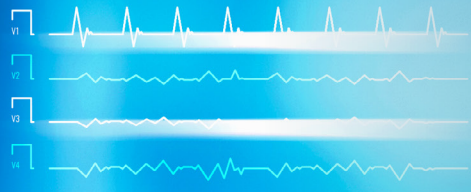
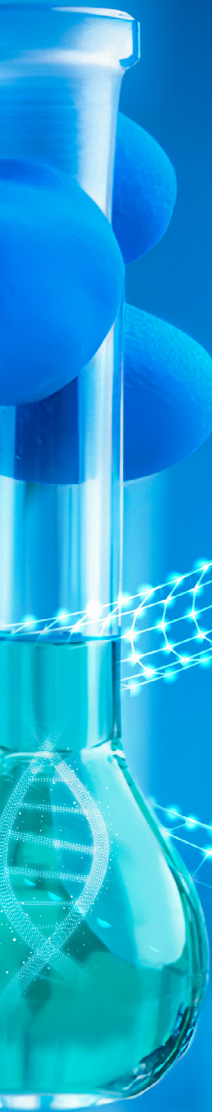
TABLA 6

**ESTRUCTURA DE LA POBLACIÓN (COMBINADAS AMBAS
POBLACIONES)**

FINAL DE JUNIO

EDAD EN AÑOS	ZONA A + ZONA B
0	1002
1 – 4	4346
5 – 9	4323
10 – 19	6103

20 – 29	3866
30 – 39	2944
40 – 49	1630
50 – 59	1151
60 +	958
TOTAL	26323



10 mm/s 10 mm/mV 0.05-35



The background is a solid teal color. In the top-left corner, there are five vertical bars of varying heights and a faint grid pattern. On the left side, there is a faint, stylized network diagram consisting of interconnected nodes and lines.

Capítulo II

VALIDEZ Y TEST O PRUEBA DE REPRODUCIBILIDAD

INTRODUCCIÓN

El enfoque de este capítulo es presentar técnicamente elementos que permitan realizar una evaluación diagnóstica y la búsqueda de métodos.

Empezamos. Para juzgar la validez de un test, uno debería evaluar:

- Su validez interna
- Su validez externa

LA VALIDEZ INTERNA DE UN TEST

Es la capacidad para evaluar de una forma o manera correcta el estado de salud de un individuo. Es la medida de calidad de un instrumento (Hernández Sampieri et al., 2010).



LA REPRODUCIBILIDAD

Es la capacidad de obtener el mismo resultado cuando se repite un test sobre la misma persona bajo idénticas condiciones.

Uno puede distinguir la reproducibilidad de un observador (= REPRODUCIBILIDAD INTRA-OBSERVACIÓN) y de varios observadores (REPRODUCIBILIDAD INTER-OBSERVADOR).

1.- SENSIBILIDAD Y REPRODUCIBILIDAD DE UN TEST O PRUEBA

Cuando una población está expuesta a un agente infeccioso, puede, por ejemplo, ser dividida en dos subpoblaciones: una subpoblación infectada y otra que no.

La prueba de laboratorio ideal sería sin error, la ambigüedad o el desconocimiento indican el valor del estado de salud.

Así, una prueba clasificaría a una **persona infectada como positiva o no infectada como negativa**. Lamentablemente, esta prueba ideal no existe. Por lo tanto, habrá ciertas subclasificaciones, como se refiere en la siguiente tabla (Vizcaíno-Salazar, 2017):

TEST	PACIENTE INFECTADO (E +)	PACIENTE NO INFECTADO (E-)	TOTAL
POSITIVO (V+)	A.- VERDADERO POSITIVO	B.- FALSO POSITIVO	A + B
NEGATIVO (V-)	C- FALSO NEGATIVO	D.- VERDADERO NEGATIVO	C + D
TOTAL	A + C	B + D	A+B+C+D

Fuente: (Vizcaíno-Salazar, 2017)

SENSIBILIDAD

La sensibilidad es la proporción de personas con el problema de salud reconocidas como tal (verdadero positivo) (Hernández Sampieri et al., 2010).

Sensibilidad = la probabilidad de ser diagnosticado como un paciente cuando uno realmente ha tenido la enfermedad = Paciente (verdadero positivo / enfermedad positiva).

$$\text{Sensibilidad} = \frac{\text{verdaderos positivos (V+)}}{\text{Total de la población enferma}}$$

$$\text{Sensibilidad} = \frac{\text{verdaderos positivos (V+)}}{\text{Verdaderos Positivos (V+) + Falsos Negativos (F-)}}$$

ESPECIFICIDAD

La especificidad es la proporción de personas no enfermas reconocidas como tal (verdaderos negativos) (Hernández Sampieri et al., 2010).

Especificidad = probabilidad de que los resultados de un test o prueba resulten negativos cuando realmente no se tiene la enfermedad = paciente (verdadero negativo/enfermedad negativa)

$$\text{Especificidad} = \frac{\text{verdaderos negativos (V-)}}{\text{Total, de población enferma}}$$

$$\text{Especificidad} = \frac{\text{Verdadero Negativo (V-)}}{\text{Verdaderos Negativos (V-) + Falsos Negativos (F-)}}$$

EJEMPLO:

146 pacientes se presentaron a consulta por la presencia de úlceras genitales, para lo cual fueron sometidos a pruebas para detectar presencia de anticuerpos del VIH, mediante ELISA (Pasteur) y con la referencia del test IBA (Immune Blood Essay). ¿Cuál es la sensibilidad y especificidad del test de ELISA?

De los 146 pacientes sometidos a pruebas, se encontró 58 IBA test positivo, 55 ELISA positivo y 1 seronegativo para ambos, IBA y ELISA.

TEST	IBA +	IBA -	TOTAL
ELISA +	54	1	55
ELISA -			
TOTAL	58		146

Los otros datos pueden ser fácilmente calculados:

TEST	IBA +	IBA -	TOTAL
ELISA +	54	1	55
ELISA -	4	87	91
TOTAL	58	88	146

Entonces:

$$\text{LA SENSIBILIDAD} = \frac{\mathbf{P\ ELISA+}}{\mathbf{IBA +}}$$

LA SENSIBILIDAD ES IGUAL: A LA PROBABILIDAD DE ENCONTRAR UN ELISA + CUANDO LA PRUEBA DE REFERENCIA ES POSITIVO.

$$\text{Sensibilidad} = \frac{54}{58} = 0,93 = 93 \%$$

$$\text{LA ESPECIFICIDAD : } \frac{\mathbf{P\ ELISA -}}{\mathbf{IBA -}}$$

LA ESPECIFICIDAD ES IGUAL: A LA PROBABILIDAD DE ENCONTRAR UN ELISA – CUANDO LA PRUEBA DE REFERENCIA ES NEGATIVA

$$\text{Especificidad} = \frac{87}{88} = 0,99 = 99 \%$$

La especificidad se expresa como un porcentaje y es importante porque nos indica cuán confiable es un test para descartar la presencia de una enfermedad en una población.

Interpretación: Una especificidad alta nos indica que el test tiene pocas probabilidades de dar un resultado falso positivo,

lo cual es importante para evitar diagnósticos erróneos y tratamientos innecesarios.

Una especificidad baja, por otro lado, indica que el test tiene una mayor probabilidad de dar falsos positivos, lo cual puede llevar a tratamientos innecesarios y a una mayor carga emocional para los pacientes.

* **P= probabilidad**

II.- VALORES PREDICTIVOS

Los valores predictivos de una prueba son una medida de funcionamiento de la prueba en una población dada (Vizcaíno-Salazar, 2017).

A.- VALOR PREDICTIVO DE UNA PRUEBA POSITIVA (VPP)

VPP = la probabilidad de tener la enfermedad cuando la prueba es positiva (**valor predictivo positivo VPP**)

$$VPP = \frac{\text{verdadero positivo (v+)}}{\text{Verdadero positivo (v+) + falso positivo (f+)}}$$

$$VPP = P(E+ / V+)$$

En nuestro ejemplo:

$$VPP = P(E+ / V+)$$

VPP= La probabilidad de ser realmente seropositivo cuando la prueba de ELISA fue Positivo (seropositivo = prueba de referencia de IBA es positiva)

$$VPP = \frac{54}{55} = 98 \%$$

B.- VALOR PREDICTIVO DE UNA PRUEBA NEGATIVA (VPN)

VPN = La probabilidad de ausencia de la enfermedad cuando la prueba es Negativa

(valor predictivo negativo VPN)

$$VPN = \frac{\text{verdadero negativo (v-)}}{\text{Verdadero negativo (v-) + falso negativo (f-)}}$$

$$VPN = \frac{V-}{V- + F-}$$

$$= P(E- / V-)$$

En nuestro ejemplo

VPN= P(E- / V-)
 = La probabilidad de ser realmente seronegativo cuando la prueba de ELISA fue Negativa (seronegativa = prueba de referencia IBA es negativo).

$$VPN = \frac{87}{91} = 96 \%$$

*** P= probabilidad, V=Verdadero , F= falso , E= prueba ELISA**

Fuente: (Vizcaino-Salazar, 2017)

III.- INFLUENCIA DE LA TASA DE PREVALENCIA SOBRE EL VALOR PREDICTIVO DE UNA PRUEBA

La sensibilidad y la especificidad son fijadas por una prueba dada; el valor predictivo va a depender de la validez interna de la prueba o test y la tasa de prevalencia de una enfermedad (Medina, 2011).

A.- EN EL EJEMPLO ANTERIOR LA PRUEBA DE ELISA TIENE UNA ESPECIFICIDAD DEL 99 % Y UNA SENSIBILIDAD DEL 93 %.

Entre los pacientes con úlceras genitales del Este de África, la tasa de prevalencia de VIH es:

$$\frac{58}{146} = 0,397 = 40 \%$$

Pero si se consultaran a 100.000 pacientes con úlcera genital, la siguiente tabla deberá ser construida:

TEST	SEROPOSITIVOS	SERONEGATIVOS	TOTAL
ELISA +	37.200	600	37.800
ELISA -	2.800	59.400	62.200
TOTAL	40.000	60.000	100.000

Entonces:

$$\text{El valor predictivo positivo sería: } = \frac{37.200}{37.800} = 98\%$$

$$\text{El valor predictivo negativo sería } = \frac{59.400}{62.200} = 95\%$$

En este caso hay 600 pacientes falsos positivos, pero también hay 2.800 falsos negativos.

Interpretación: En este caso, se está presentando un escenario donde se evalúa la prevalencia de VIH entre los pacientes con úlceras genitales en el Este de África. Se menciona que la tasa de prevalencia de VIH es del 40% entre los pacientes con úlceras genitales.

Posteriormente, se plantea una situación hipotética donde se consultan a 100.000 pacientes con úlcera genital y se construye una tabla con los resultados de las pruebas de ELISA realizadas para detectar la presencia de VIH. En la tabla se muestran los resultados de los pacientes seropositivos y seronegativos en función de los resultados de la prueba ELISA.

A partir de los datos de la tabla, se calculan el valor predictivo positivo y el valor predictivo negativo. El valor predictivo positivo se refiere a la probabilidad de que un paciente con un resultado positivo en la prueba tenga realmente

la condición que se está evaluando (en este caso, VIH). En este escenario, el valor predictivo positivo es del 98%, lo que significa que el 98% de los pacientes con un resultado positivo en la prueba ELISA realmente tienen VIH.

Por otro lado, el valor predictivo negativo se refiere a la probabilidad de que un paciente con un resultado negativo en la prueba no tenga la condición que se está evaluando. En este caso, el valor predictivo negativo es del 95%, lo que indica que el 95% de los pacientes con un resultado negativo en la prueba ELISA realmente no tienen VIH.

Estos valores de valor predictivo son importantes para evaluar la eficacia de la prueba en la detección de la condición (VIH en este caso) y para interpretar los resultados de manera más precisa en función de la prevalencia de la enfermedad en la población estudiada.

B.- PODRÍA USARSE EL MISMO VALOR DE LA PRUEBA DE ELISA CON UNA SENSIBILIDAD DE 93%, ESPECIFICIDAD DEL 99%, EN UNA POBLACIÓN DEL ÁREA RURAL DE AFRICANOS MONÓGAMOS CON UNA SEROPREVALENCIA DEL 0,1 %; ENTONCES:

TEST	SEROPOSITIVOS	SERONEGATIVOS	TOTAL
ELISA +	93	907	1.000
ELISA -	7	98.901	98.908
TOTAL	100	99.900	100.000

$$\text{VPP} = \frac{93}{1.000} = 9.3 \%$$

$$\text{VPN} = \frac{98.901}{98.908} = 99.99 \%$$

Interpretación:

- De cada 1.000 personas que se someten a la prueba ELISA, 93 son seropositivas y 907 son seronegativas.
- De las 93 personas que dan positivo en la prueba ELISA, 9,3% (9 personas) realmente tienen la enfermedad.
- De las 907 personas que dan negativo en la prueba ELISA, 99,99% (906 personas) realmente no tienen la enfermedad.

C.- UNO PUEDE EVITAR CONSTRUIR LA TABLA DE CONTINGENCIA PARA CADA SITUACIÓN DE LA PREVALENCIA, POR LO QUE PODEMOS USAR EL TEOREMA DE MAYES. ESTE TEOREMA PERMITE OBTENER LOS VALORES PREDICTIVOS POSITIVOS Y/O NEGATIVOS DE UNA PRUEBA USANDO LOS SIGUIENTES PARÁMETROS:

- La sensibilidad de la prueba
- La especificidad de la prueba
- La tasa de prevalencia

La fórmula para el valor predictivo positivo es:

$$\text{VPP} = \frac{\text{Sensibilidad} * \text{la Tasa de Prevalencia}}{\text{Sensibilidad} * \text{TP} + (1-\text{PR}) (1 - \text{Especificidad})}$$

TP: tasa de prevalencia, PR= prevalencia

Aplicando la fórmula:

1.- EN EL EJEMPLO ANTERIOR DE PACIENTES CON ÚLCERAS GENITALES Y CON VIH SEROPOSITIVOS:

- Sensibilidad = 93 %
- Especificidad = 99 %
- Tasa de prevalencia = 40 %

$$\text{VPP} = \frac{0,93 * 0,40}{0,93 * 0,40 + (1- 0,40) (1-0,99)}$$

$$\text{VPP} = \frac{0,93 * 0,40}{0,93 * 0,40 + (0,60) (0,01)}$$

$$\text{VPP} = \frac{0,372}{0,378} = 0,98 = 98 \%$$

Entonces, el valor predictivo positivo es de 98 %

Interpretación:

- El VPP de 98% indica que, de cada 100 pacientes con úlceras genitales que dan positivo en la prueba del VIH, 98 realmente tienen la infección.
- Esto significa que la prueba es muy precisa para identificar a los pacientes con VIH que tienen úlceras genitales.

2.- EN EL EJEMPLO DE LA POBLACIÓN MONOGÁMICA CON UNA TASA DE PREVALENCIA VIH SEROPOSITIVO DE 0,1 %, TENEMOS:

$$VPP = \frac{0,93 * 0,001}{0,93 * 0,001 + (0,999) (1-0,001)}$$

$$VPP = 0,00093 / 0,00999 = 0,093$$

$$VPP = 9,3 \%$$

Interpretación:

- El VPP de 9,3% indica que, de cada 100 personas monógamas que dan positivo en la prueba del VIH, 9,3 realmente tienen la infección.
- Esto significa que la prueba tiene una precisión moderada para identificar a los pacientes con VIH en esta población con baja prevalencia.

D.- EL VALOR PREDICTIVO NEGATIVO PUEDE TAMBIÉN SER CALCULADO CON LOS MISMOS TRES PARÁMETROS ANTERIORES:

$$\text{VPN} = \frac{\text{Especificidad} * (1 - \text{PR})}{(\text{Especificidad} * (1-\text{PR}) + (1 - \text{Sensibilidad}) * \text{TP})}$$

Explicación de la fórmula:

Especificidad: Probabilidad de que una persona sana dé un resultado negativo en la prueba.

TP: Tasa de verdaderos positivos, es decir, la proporción de personas con la enfermedad que dan positivo en la prueba.

PR: Tasa de prevalencia, es decir, la proporción de personas en la población que tiene la enfermedad.

Sensibilidad: Probabilidad de que una persona con la enfermedad dé un resultado positivo en la prueba.

Ejemplo:

- Sensibilidad = 93 %
- Especificidad = 99 %
- Tasa de prevalencia = 40 %

$$\text{VPN} = \frac{0,99 * (1-0,40)}{0,99 * (1-0,40) + (1- 0,93) * 0,40}$$

$$\text{VPN} = \frac{0,99 * 0,60}{0,99 * 0,60 + 0,07 * 0,40}$$

$$\text{VPN} = \frac{0,594}{0,594 + 0,028}$$

$$\text{VPN} = 0,95 = 95\%$$

Interpretación del VPN:

En este caso, el VPN del 95% indica que, de cada 100 personas que dan negativo en la prueba, 95 realmente no tienen la enfermedad. Esto significa que la prueba tiene una alta precisión para descartar la enfermedad en personas sanas.

IV.- REPRODUCIBILIDAD DE UN TEST

La reproducibilidad de una prueba o test, nos indica la probabilidad de obtener el mismo resultado cuando una prueba es repetida de una manera similar en igual sujeto u objeto y bajo las mismas condiciones (Manterola et al., 2018).



1.- VARIACIÓN BIOLÓGICA

La variación biológica hace referencia a todos los factores de los individuos. Por ejemplo: edad, género, raza, condición socioeconómica, cultura, experiencias médicas previas, etc.

Por lo tanto, si partimos de que cada individuo es diferente, los resultados de las medidas serán diferentes.

Ejemplo:

Una medida de la concentración de hemoglobina en 50 individuos, no será igual a que se repitan 5 veces en cada 10 personas.

2.- VARIACIÓN TEMPORAL

La variación temporal hace referencia a todos los factores que podrían causar variación en el tiempo: descanso o actividad, tiempo del día, etc. Entonces, uno debería medir bajo similares condiciones al mismo momento del día.

3.- VARIACIONES EN LAS MEDIDAS

Las variaciones en las medidas hacen referencia a los errores en la medición. En estos están todos los factores que de una u otra forma hacen o tienden a producir diferencias en las medidas. Estos errores pueden deberse a los instrumentos y/o al observador que usa esta medida. Los errores del observador pueden ser detectados cuando se reproducen las pruebas. Por lo tanto, debemos considerar las diferencias entre:

a.- La reproducibilidad únicamente en un observador: la reproducibilidad dentro del observador:

= Las fallas hechas por el observador quien repite las observaciones (ciego)

b.- La reproducibilidad dentro de varios observadores: La reproducibilidad inter – observador:

= Las fallas cometidas por varios observadores durante la ejecución y la interpretación de la prueba.



Uno puede cuantificar la reproducibilidad de la siguiente manera:

SEGUNDA LECTURA

PRIMERA LECTURA	+	-	TOTAL
+	a	b	a + b
-	c	d	c + d
TOTAL	a + c	b + d	a+b+c+d

Ejemplo:

Para conocer la reproducibilidad de un electrocardiograma (ECG), realizado y leído debido a una cardiopatía por chagas crónico, cuyo cardiólogo realiza la lectura de dos veces en estudio ciego de 78 ECG.

$$\text{La reproducibilidad} = \frac{a + d}{a + b + c + d}$$

Explicación de la fórmula:

- a. Número de casos en los que ambos observadores coinciden en que el resultado es positivo.
- b. Número de casos en los que el primer observador dice que el resultado es positivo y el segundo dice que es negativo.
- c. Número de casos en los que el primer observador dice que el resultado es negativo y el segundo dice que es positivo.
- d. Número de casos en los que ambos observadores coinciden en que el resultado es negativo.

Los resultados son los siguientes:

SEGUNDA LECTURA

PRIMERA LECTURA	+	-	TOTAL
+	34	6	40
-	6	32	38
TOTAL	40	38	78

La reproducibilidad = $\frac{34 + 32}{78}$

La reproducibilidad = 85 %

Interpretación: En este estudio, la reproducibilidad es del 85%. Esto significa que hay un alto grado de acuerdo entre los dos observadores.

V.- CONCLUSIÓN

Los siguientes conceptos son propuestos en este capítulo:

- Sensibilidad
- Especificidad
- Valor Predictivo Positivo
- Valor Predictivo negativo
- Valoración interna de una prueba o test
- Reproducibilidad

Usted deberá ser capaz de evaluar la calidad de una prueba de diagnóstico, comparándola con una prueba de referencia y aplicando el teorema de Bayes para juzgar la utilidad de la prueba en un diagnóstico y las estrategias de seguimiento.

Finalmente, Ud., deberá ser capaz de entender y calcular la reproducibilidad de una prueba o test.

AUTOEVALUACIÓN

Por favor, indicar las respuestas correctas: A, B o C sobre esta hoja.

Una nueva prueba ha sido desarrollada para detectar la leishmaniasis visceral. En condiciones ideales, una evaluación de esta prueba o test con 100 pacientes infectados y 100 pacientes no infectados da los siguientes resultados.

	<i>PRUEBA +</i>	<i>PRUEBA -</i>	<i>TOTAL</i>
LEISHMANIASIS +	67	33	100
NO LEISHMANIASIS	1	99	100
TOTAL	68	132	200



1. La especificidad de esta prueba es:

- A.- 75 %
- B.- 98 %
- C.- 99 %

2. La sensibilidad de esta prueba es:

- A.- 33 %
- B.- 67 %
- C.- 98 %

3. Nosotros vamos a usar esta prueba en una ciudad donde la prevalencia de leishmaniasis visceral es del 5 %. Supongamos lo siguiente: una mujer de esta ciudad da como resultado una prueba positiva, ¿Cuál sería la probabilidad de que ella realmente tenga leishmaniasis?

- A.- 98 %
- B.- 67 %
- C.- 78 %

EJERCICIO SOBRE: SENSIBILIDAD, ESPECIFICIDAD Y VALOR PREDICTIVO

1. En el laboratorio central, 120 pacientes fueron confirmados con infección por VIH mediante la prueba de "Western Blot"; mientras que, de los pacientes que se efectuaron la prueba de ELISA, 118 fueron positivos y 2 negativos. Por su parte, de 160 pacientes con VIH seronegativos que se realizaron las pruebas en el mismo laboratorio, 5 tienen un test de ELISA positivo. Pregunta: ¿Cuál es la sensibilidad y la especificidad del test de ELISA?

Para las siguientes preguntas, use los resultados obtenidos en la pregunta 1 y las tablas que se encuentran en el anexo 1, cuya prevalencia de VIH es diferente en la población.

2. Preguntas:
 - a. ¿Cuál es la probabilidad que una persona de 32 años con ELISA POSITIVO?: ¿es realmente seropositivo?
 - b. ¿Cuál es la probabilidad que una persona de 20 años con ELISA POSITIVO?: ¿es realmente seropositivo?
 - c. ¿Cuál es la probabilidad que un niño de 10 años con ELISA NEGATIVO?: ¿es realmente seronegativo?
3. Ud., organiza un seguimiento para VIH con la prueba de ELISA en un sector del área urbana, 10.000

personas entre los 25 y 30 años de edad utilizaron los servicios. ¿Cuántas personas seropositivas han sido perdidas o pasadas por alto?

4. Supongamos que Ud. organiza el seguimiento en áreas rurales del país.

¿Ud. encontraría más verdaderos positivos (VP) que falsos positivos (FP)?

¿Ud. pensaría que es digno de atención organizar un seguimiento debido a este problema?

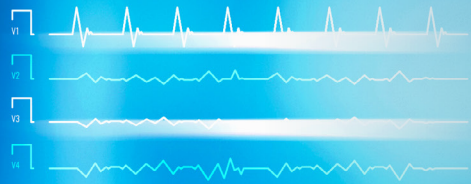
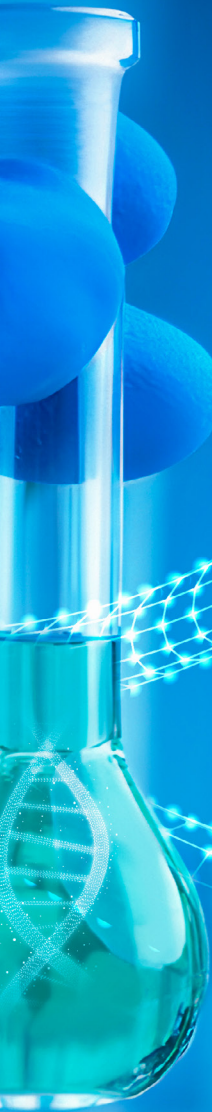
ANEXO 1

TABLA 1

<i>GRUPO DE EDAD</i>	<i>NEGATIVO</i>	<i>POSITIVO</i>	<i>% POSITIVO</i>
0 - 4	355	12	3,3
5-14	453	4	0,9
15-24	834	95	10,3
25-34	442	108	19,5
35-44	477	111	18,9
> 45	338	39	10,3
GÉNERO			
MASCULINO	1.504	140	8,5
FEMENINO	1.394	229	14,1
TOTAL	2.899	369	11,3

TABLA 2

<i>GÉNERO</i>	<i>NEGATIVO</i>	<i>POSITIVO</i>	<i>% POSITIVO</i>
MASCULINO	467	2	0,4
FEMENINO	479	5	1,0
TOTAL	946	7	0,7



10 mm/s 10 mm/mV 0.05-35



Capítulo III

DISEÑOS DE ESTUDIO

El diseño de un estudio consiste en los pasos y estrategias que un investigador utiliza para escoger a los sujetos de estudio, recopilar datos, analizar la información obtenida y dar significado a los hallazgos. Estos diseños buscan controlar variables y minimizar posibles sesgos para obtener resultados válidos y generalizables. Existen diferentes tipos de diseños de estudio, como experimentales, cuasiexperimentales, observacionales y descriptivos, cada uno con sus propias características y metodologías. Los diseños de estudio son la base para realizar investigaciones rigurosas y obtener conclusiones sólidas en diversas disciplinas académicas.

ESTUDIOS EXPERIMENTALES

Los estudios experimentales, determinan el impacto de una intervención a nivel individual o comunitario.

Ejemplos:

- Ensayos Clínicos
- Ensayos de vacunación
- Esquema de la modificación de la conducta (Maldonado Reyes et al., 2023)

Principios:

1. Las unidades experimentales son elegidas de las bases de la inclusión o de exclusión
2. Estas unidades son asignadas en forma aleatoria a diferentes esquemas de tratamiento.
3. Los resultados finales de los grupos son comparados. Si alguna diferencia observada es considerada, se atribuye la diferencia en el esquema de tratamiento.

Nota:

- La selección de riesgo se remedia cuando la asignación del esquema de tratamiento es en forma aleatoria.
- Si las muestras son grandes, los esquemas de los factores de confusión son igualmente distribuidos sobre

el grupo de tratamiento y el sesgo de confusión es prevenido. Si el esquema de tratamiento es aplicado a un grupo de control pequeño, los factores de confusión deben también ser evitados necesariamente.

- Siempre hay un sesgo de observación (que puede ser minimizado por el estudio ciego), con pérdida selectiva para la siguiente prueba.

En la práctica: la distribución de los esquemas de tratamiento en las comunidades de prueba está frecuentemente basado sobre las consideraciones de la naturaleza operativa, tales como: la disposición de cierto grupo de personas que van a participar en el estudio, o el número de comunidades que uno pueda manejar. Sin embargo, si dos grupos no son iguales antes del tratamiento, las diferencias en los resultados no pueden ser atribuidas a las diferencias en el tratamiento.

Los estudios epidemiológicos experimentales se usan para argumentar y probar el impacto de un "tratamiento" y se realizan en los laboratorios experimentales porque el investigador tiene el control sobre:



ESTUDIOS NO EXPERIMENTALES

En los estudios no experimentales, el factor de exposición no es manipulado por el investigador.

Las personas, por sí mismas, deciden ser (o pasan a ser) expuestas a ciertas influencias dañinas o a ciertas influencias benéficas. Si nosotros observamos, por ejemplo, la diferencia significativa en la tasa de cáncer entre los fumadores y no fumadores, uno no puede inmediatamente inferior o atribuir la relación entre la causa y la enfermedad. En realidad, algunas diferencias existen entre los fumadores y no fumadores (edad,

género, estado socio económico, estrés, hábitos nutricionales, bebedor, etc.) y algunos de estos factores están, por sí mismos, asociados al cáncer. **Ellos son llamados: “factores de confusión”** (Maldonado Reyes et al., 2023).

Por lo tanto, está claro que la dificultad de interpretación es mucho más difícil en un estudio no experimental que en un estudio experimental.

ESTUDIOS OBSERVACIONALES

Existen dos tipos de estudios observacionales: estudios descriptivos y estudios analíticos.

A.- ESTUDIOS EPIDEMIOLÓGICOS DESCRIPTIVOS

Describe la situación usando los diferentes parámetros epidemiológicos:

Tiempo: estación, periodicidad, tendencias.

Lugar: geográfico, distribución, urbano, rural.

Persona: edad, género, ingresos económicos, raza, etc.

Los estudios descriptivos, pueden ser de corte transversal (usando la tasa de prevalencia como descriptor) o corte longitudinal (seguimiento, usando la tasa de incidencia o la tasa de incidencia acumulativa como descriptor). Los resultados serán tasas (específicas), gráficos o mapas sobre

los cuales uno puede distinguir las características y; por lo tanto, la evolución de un fenómeno (Villa Romero et al., 2012).

El estudio descriptivo no responde a la pregunta porqué, pero puede generar una hipótesis, por ejemplo:

- Cuando buscamos casos de cólera durante una epidemia, ellos pueden ser encontrados en las comunidades alrededor de las fuentes de agua.
- Cuando buscamos la prevalencia o la incidencia de una enfermedad para un programa de control que gradualmente viene siendo implementado, uno puede observar el decrecimiento de las tasas en áreas que fueron cubiertas por las nuevas medidas.

Estas hipótesis tienen que ser confirmadas por los estudios analíticos.

B.- ESTUDIOS EPIDEMIOLÓGICOS ANALÍTICOS

Trata de encontrar la respuesta a la pregunta "**¿por qué ocurrió el evento?**"

En estos estudios, uno intenta establecer la asociación entre los factores de exposición X (= presumibles causas) y el resultado (= enfermedad, efecto) mientras controlamos los otros factores que pueden confundir la asociación.

El resumen estadístico de esta asociación son la tasa y el ratio, el odds ratio (OR) y el riesgo atribuible (RA), los cuales permiten medir el impacto del factor X sobre el resultado.

Para inferir la causa, es importante establecer que la exposición tuvo lugar antes que el evento de interés ocurriera. Algunos diseños de estudios permiten establecer los momentos de la exposición más precisamente que otros (Bonita et al., 2008; Villa Romero et al., 2012).

ESTUDIOS NO EXPERIMENTALES

A.- ESTUDIOS TRANSVERSALES

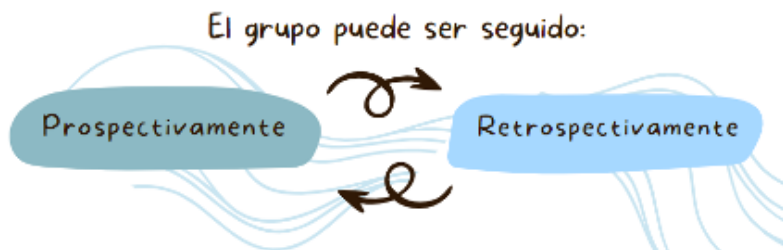
Son los estudios en los cuales la información del estado de salud y los factores de exposición son reunidos simultáneamente y por una sola vez (Espinoza Casco et al., 2023; Maldonado Reyes et al., 2023). Esto se denomina una ***“fotografía epidemiológica”***.

Ejemplo: En un estudio de la enfermedad de chagas, las personas (tasa de infección, tasa de enfermedad crónica) y el vector (tasa de infestación de las casas) son estudiadas al mismo tiempo. Pero; ¿hay que considerar que muchas más personas podrían haber sido infectadas hace 30 años atrás y la tasa de infestación de las casas, es de ahora o éstas pudieron haber sido infestadas antes?

Este ejemplo demuestra las dificultades agrupadas cuando interpretamos el estudio transversal. Otros diseños de estudios dirigen el problema en el tiempo de la exposición más la ocurrencia de la enfermedad.

B.- ESTUDIOS LONGITUDINALES (ESTUDIOS DE SEGUIMIENTO)

Un grupo de personas es seleccionado sobre las bases de la presencia o ausencia de los factores de riesgo (edad, género, hábitos particulares, etc.) y seguida sobre un periodo de tiempo. Dicho grupo constituye un COHORTE. La aparición de un problema específico de salud es observada en el tiempo (Hernández Sampieri et al., 2010).



ESTUDIO PROSPECTIVO

El investigador tiene que ir observando lo que está pasando en el grupo de estudio en forma conjunta, es decir, observar el futuro de la cohorte. Por ejemplo, el investigador y el grupo de edad (Hernández Sampieri et al., 2010; Maldonado Reyes et al., 2023)



Ejemplo: Para estudiar la malnutrición como un factor de riesgo de mortalidad relacionado con sarampión; se realiza una toma desde el 1/06/94 en el siguiente grupo de niños para el estudio:

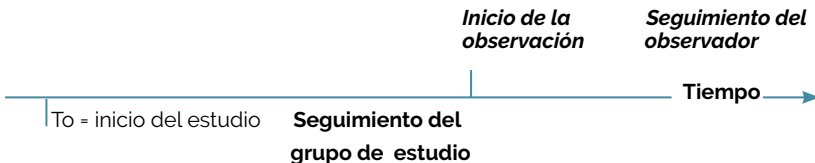
Una cohorte de niños bien alimentados y una cohorte de niños desnutridos

Posteriormente se realiza el seguimiento hasta el 31/05/95

Por lo tanto; tendremos en la cohorte una pequeña selección de riesgo ya que el evento no toma lugar en ese momento, sino, hasta después de empezar el estudio. Sin embargo, el estudio sería muy costoso y existe el riesgo de que algunos de los niños se pierdan en el seguimiento (aunque uno haya seleccionado un grupo de estudio).

ESTUDIO RETROSPECTIVO O HISTÓRICO

Cuando el problema de estudio ha tomado ya lugar o este ya se ha producido.



Ejemplo: para realizar el estudio de niños con problemas de desnutrición, todas las historias clínicas de los niños que fueron atendidos durante al menos los últimos 5 años (desde 1 / 6 / 89) son examinadas y analizadas desde el 1/6/94 (y en los siguientes días).

Un estudio retrospectivo es fácil de realizar y los costos son mucho más bajos que los estudios prospectivos. Sin embargo, hay una seria posibilidad de sesgo, por lo tanto; el investigador necesita tener datos reales y uniformes en la historia clínica de todos los elementos del estudio.

C.- CASO CONTROL

En este tipo de estudio, uno toma un grupo de personas que exhiben un fenómeno de salud (casos) y un grupo que tiene el particular rasgo (control) y se pregunta a ambos grupos acerca de su exposición a un posible factor causal (Donis, 2013; Hernández-Avila et al., 2000).

Ejemplo:

CÁNCER DE PULMÓN

<i>FUMADOR</i>	<i>SI (+)</i>	<i>NO (-)</i>
<i>SI (+)</i>	A (+) (+)	B (+) (-)
<i>NO (-)</i>	C (-)(+)	D (-)(-)

En un estudio de caso control, uno no puede determinar las oportunidades de tener una enfermedad cuando está expuesto a un factor de riesgo.

Entonces, uno compara la probabilidad de exposición en los casos y los controles.

Indicaciones:

1. Largo periodo de latencia entre la exposición y la enfermedad.
2. La reunión de la información es difícil, cara y dolorosa (ejemplo: frecuente contaminación de la muestra tomada), sin embargo; únicamente unos pocos problemas pueden proveer información valiosa.
3. Pérdida de la hipótesis.

La siguiente tabla exhibe la diferencia entre un estudio longitudinal y un caso control.

<i>CARACTERÍSTICAS</i>	<i>LONGITUDINAL</i>	<i>CASO CONTROL</i>
Selección de participantes	Sobre el estado de exposición	Sobre el estado de la enfermedad
Observación	Incidencia de la enfermedad	Exposición pasada
Duración del estudio	Largo	Corto
Tamaño de la muestra	Grande	Pequeña
Facilidades operacionales	Bajo	Alto
Costos	Alto	Bajo
Peligro de selección de riesgo al empezar el estudio	No (prospectivo) Sí (retrospectivo)	Sí
Pérdida del seguimiento	Frecuente	Ninguna
Peligro de sesgo en la observación	Sí	Sí
Necesidad de control de los factores de confusión	Sí	Sí
Generación de hipótesis	No	Sí
Usa para la prueba de hipótesis	Sí	Sí
Tasas estadísticas	Incidencia, ratio	No incidencia – Odds ratio

CONCLUSIÓN

Los siguientes conceptos han sido discutidos en este capítulo:

- | | |
|-----------------------------|---|
| Estudios Observacionales: | - Descriptivos
- Analíticos |
| Diseños de estudio: | - Experimentales
- No Experimentales |
| Estudios no experimentales: | - Estudios transversales
- Estudios longitudinales
- Caso control |

Por lo tanto Ud., estará en la capacidad de seleccionar un apropiado diseño de estudio epidemiológico para evaluar un problema de salud en función de:

- La frecuencia del fenómeno
- Su población objetivo
- El tipo del factor de riesgo
- Su presupuesto y medidas técnicas

EJERCICIO DE DISEÑO DE ESTUDIOS

Reconozca los siguientes diseños:

- 1.1.- En una población, todos los hombres adultos fueron visitados en sus lugares de vivienda, para estudiar los casos de ocurrencia de hernia inguinal. "De los 364 individuos incluidos en la investigación llevada a cabo en abril de 1994, 86 tenían al menos una hernia inguinal o cicatriz quirúrgica".
- 1.2.- Para estudiar la ocurrencia de tuberculosis en monos de un zoológico, una prueba de tuberculosis fue llevada a cabo sobre los animales que existían (41 animales de diferentes especies), de los cuales; 3 animales resultaron positivos.
- 1.3.- Para detectar la resistencia a la penicilina en *Neisseria Gonorrhoeae*, 256 prostitutas fueron estudiadas bisemanalmente durante seis meses (51 casos fueron encontrados).
- 1.4.- Un estudio para valorar la eficacia residual de fenitroton, en una concentración de 1 g/m² y 2 gr/m², ha sido llevado a cabo.

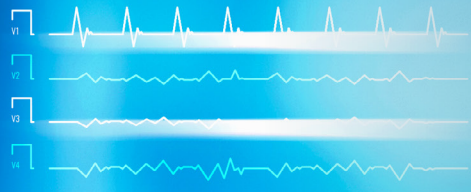
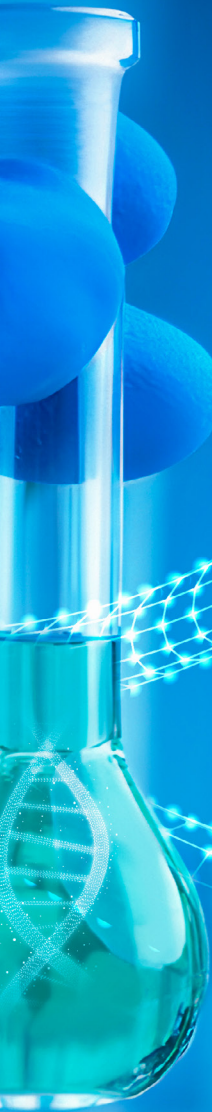
Los efectos residuales han sido determinados por una prueba biológica aplicada cada dos semanas luego de la fumigación y durante un periodo de 12 semanas. El control de las pruebas ha sido lle-

vado a cabo en similares condiciones en casas no fumigadas.

1.5.- Resumen del autor.

"Un estudio llevado a cabo en Bolivia, en un valle de nivel intermedio, siguiendo la cordillera, al plano del Amazonas, se estudió 113 casos de leishmaniasis. Mientras se realizaba el estudio y se distribuían los 113 casos de acuerdo a la edad, género y raza, el grupo de investigación notó que había una altísima frecuencia de leishmaniasis entre los 20 y 30 años de edad y entre la población de blancos".

1.6.- En la ciudad X, 39 casos de SIDA han sido detectados en noviembre de 1994. Una investigación fue llevada a cabo considerando las características serológicas de esos 39 casos y su conducta sociológica ha sido estudiada en detalle. Similares observaciones han sido realizadas sobre una muestra de los pacientes del hospital en la misma ciudad.



10 mm/s 10 mm/mV 0.05-35



Capítulo IV

ESTADÍSTICA DESCRIPTIVA GENERALIDADES

PARTE I

ENFOQUE EXPLORATORIO DE LOS DATOS

La estadística puede ser clasificada como Descriptiva (este capítulo) o Inferencial (próximo capítulo).

La **estadística descriptiva** se refiere a los procedimientos y formas de organizar, resumir, presentar (gráficamente) e interpretar los datos numéricos.

- a) *¿Cuáles son los pasos a seguir cuando examinamos un conjunto de datos?*
- b) *Cuando los datos han sido ya recolectados, ¿cuál de los casos es el más frecuente?, ¿Cómo nosotros estudiaremos correctamente ellos?*

Considerar el siguiente,

PLAN DE ANÁLISIS:

PASO 1: EVALUAR EL GRUPO DE DATOS

- El número de las variables
- La escala de medidas de cada una de las variables
- La forma de los datos: ¿estos son datos originales?
¿Estos han sido agrupados o resumidos ya?

PASO 2: DECIDIR CUÁLES VARIABLES VAN A SER ESTUDIADAS

¿Está Ud., interesado en las variables aisladamente o en la relación entre ellas?

PASO 3: HACER UNA APROPIADA REPRESENTACIÓN GRÁFICA DE LOS DATOS

Analizar e interpretar esos datos.

PASO 4: SELECCIONAR LA APROPIADA DESCRIPCIÓN DE LOS PARÁMETROS

Resumir los datos. Analizar y explicar los datos numéricos encontrados.

VARIABLES

En un estudio epidemiológico o bioestadístico, los atributos o cualidades de la población objetivo se denominan variables; estos varían entre los diferentes sujetos, por el mismo sujeto o pueden variar también en el tiempo. Cuando un atributo en una población no varía, es denominado CONSTANTE.

Variables según medición

- **Variables cualitativas:** son aquellas que representan características o cualidades no cuantificables mediante números. Estas variables pueden tomar forma de categorías, clases o grupos. Algunos ejemplos de variables cualitativas incluyen el género, la nacionalidad, el estado civil, el color de ojos, entre otros.

Estas variables se utilizan comúnmente en estudios de investigación para clasificar y describir diferentes atributos de la población en estudio.

- **Variables cuantitativas:** son aquellas que representan cantidades numéricas que pueden ser medidas y clasificadas. Estas variables pueden ser de dos tipos: las variables continuas que toman un valor en un rango específico, como la altura o el peso de una persona, y las variables discretas que representan cantidades con un número finito de posibles valores, como la cantidad de hijos de una familia.

VARIABLES SEGÚN SU INFLUENCIA

- **VARIABLES INDEPENDIENTES:** Son aquellas variables que se manipulan o controlan en un experimento y que se cree que pueden influir en el resultado. Por ejemplo, la dosis de un medicamento en un estudio clínico.
- **VARIABLES DEPENDIENTES:** Son aquellas variables cuyo valor depende de la variable independiente y que se utilizan para medir el efecto de la variable independiente en el experimento. Por ejemplo, la mejoría de los síntomas en los pacientes que reciben el medicamento en el estudio clínico.

B.- LA ESCALA DE MEDIDAS

La escala de medida es la serie de clases o categorías que permiten indicar la posición del atributo de un objeto de estudio. Las clases son detalladas y mutuamente excluyentes.

LA ESCALA NOMINAL

Las clases son nominadas y no tienen un orden. Su forma más simple de presentación es la escala dicotómica (únicamente dos clases). El atributo de cualquiera de las dos puede estar presente o ausente. Ejemplo: grupo sanguíneo, grupo étnico, lenguaje, estado civil, etc. (Espinoza Casco et al., 2023; Hernández Sampieri et al., 2010; Maldonado Reyes et al., 2023).

LA ESCALA ORDINAL

Las clases son también nominadas, pero estas guardan un determinado orden que naturalmente existe entre ellas. Ejemplo: nivel de escolaridad, grupo socioeconómico, grado de desnutrición, etc. (Espinoza Casco et al., 2023; Hernández Sampieri et al., 2010; Maldonado Reyes et al., 2023).

ESCALAS NUMÉRICAS

Contiene todos los valores numéricos que pueden ser contados o medidos.

1.- Escala numérica discreta

Los valores observados son obtenidos por conteo, ellos son los números enteros.

Ejemplo: el pulso por minuto, la paridad de una mujer, número de visitas a una unidad de salud por año, etc.

2.- Escala numérica continua

Los valores no son contados, pero son medidos. El resultado es un número real ordenado desde menos infinito a más infinito (en la práctica de cero infinitos a más infinito).

El número de decimales del valor está delimitado por la precisión del dispositivo de la medida. Ejemplo: peso, talla, concentración de hemoglobina, colesterolemia, circunferencia de brazo, etc.

Es importante distinguir las diferencias de la escala de las medidas, porque la selección del enfoque del resumen y la representación gráfica dependen de ello. (Espinoza Casco et al., 2023; Hernández Sampieri et al., 2010; Maldonado Reyes et al., 2023).

3.- Escala de intervalo

Las unidades de medida tienen igual magnitud entre sí, pero no tienen un punto de origen natural o cero absoluto. Esto significa que se pueden realizar operaciones matemáticas como la suma y la resta, pero no se puede calcular la razón entre los valores. Un ejemplo de escala de intervalo es la escala de temperatura Celsius, donde el cero no indica una ausencia de temperatura.

4.- Escala de razón

Las unidades de medida tienen igual magnitud entre sí y también tienen un punto de origen natural o cero absoluto. Esto significa que se pueden realizar operaciones matemáticas como la multiplicación y la división, y se puede calcular la razón entre los valores. Un ejemplo de escala de razón es la escala de peso en kilogramos, donde el cero indica la ausencia de peso.

3.- CONCLUSIONES

Los siguientes conceptos han sido explicados:

- Variable cualitativa
- Variable cuantitativa
- Escala de medidas
- Escala nominal
- Escala ordinal
- Escala numérica discreta
- Escala numérica continua
- Escala de intervalo
- Escala de razón

Usted deberá conocer qué pasos debe tomar para evaluar el grupo de datos y ser capaz de llevar a cabo el primer paso: clasificar las variables y reconocer la escala de medida.

ESTADÍSTICA DESCRIPTIVA DE UNA DIMENSIÓN

PARTE II

MÉTODOS GRÁFICOS

A.- OBJETIVOS

- a. Sirve para visualizar la distribución de los datos
 - Su forma general
 - La concentración de ciertos valores
 - Sus extremos
 - Para encontrar anomalías en la distribución
- b. Para fácilmente comparar diferentes distribuciones.
- c. Para seleccionar correctamente los parámetros de la reducción de los datos.



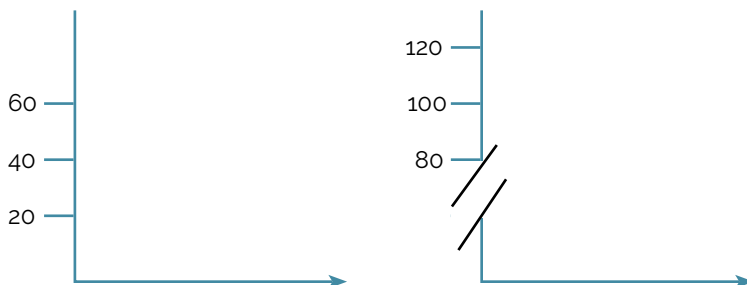
La representación gráfica de los datos cualitativos difiere de la representación gráfica de los datos cuantitativos.

B.- ROLES BÁSICOS

Algunos roles básicos deben ser aplicados en ambos tipos de representación gráfica.

- a. La figura tiene que llevar un título que claramente describe lo que el gráfico representa y que pueda ser entendido sin la necesidad de leer el texto (persona, lugar, tiempo).
- b. El total general y el total en cada uno de los parámetros tienen que ser indicados. Esto es particularmente importante cuando uno organiza los datos en porcentajes (400 de 800 equivale al 50 %, como 3 de 6 es el 50%. El conocimiento de los totales es importante como números absolutos).
- c. Las unidades de medida tienen que ser claramente indicadas.
- d. Las coordenadas tienen que ser marcadas y nombradas.
- e. Los ejes tienen que empezar en cero, caso contrario deberán ser marcadas con el signo correspondiente “//”

f.



Fuente: Elaborado por los autores

B.- PRESENTACIÓN DE DATOS CUALITATIVOS

DIAGRAMA DE BARRAS

Es ampliamente usado para la presentación de datos cualitativos (Ardila Duarte, 2022).

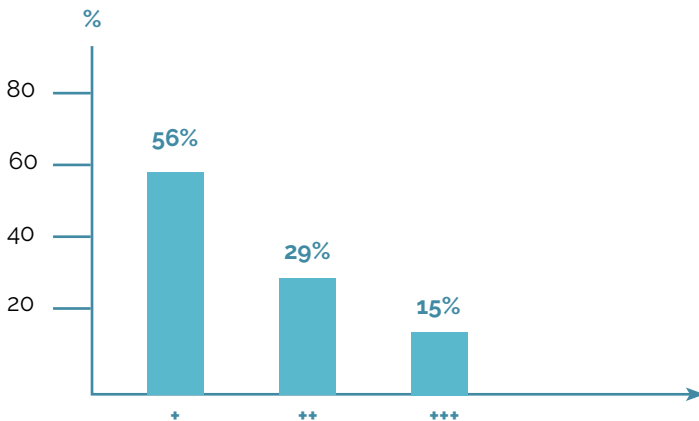
Procedimiento:

- Todos los parámetros son divididos en clases
- Los rectángulos de igual ancho son elevados desde el eje X. Su altura es proporcional a la frecuencia relativa o absoluta de las clases.
- El total del largo de las barras es 1 (o su equivalente al 100 %) para la presentación de la frecuencia relativa y N (total del número de observaciones) para la frecuencia absoluta.

Ejemplo: presencia de bacterias en la orina de 90 pacientes comprendidos en las edades de 7 a 15 años en el Centro de Salud de Azogues

PRESENCIA DE BACTERIAS	NÚMERO OBSERVADO	%
+	50	56
++	26	29
+++	14	15
TOTAL	90	100

Presencia de bacterias en la orina de 90 pacientes en las edades de 7 a 15 años del Centro de Salud de Azogues. Enero 2005



Interpretación:

- 56% (50 pacientes) de la muestra presenta una cruz "+" de bacterias en la orina, lo que indica una cantidad baja o moderada.
- 29% (26 pacientes) presenta dos cruces "++" de bacterias, indicando una cantidad moderada o alta.
- 15% (14 pacientes) presenta tres cruces "+++" de bacterias, lo que indica una cantidad alta de bacterias en la orina.

DIAGRAMA DE SECTOR O DE TORTA

Este diagrama es frecuentemente usado para ilustrar los datos cualitativos. Alguna gente encuentra más atractivo que el diagrama de barras, pero se deben utilizar cuando hay pocas categorías (Ardila Duarte, 2022).

La técnica consiste en convertir las frecuencias en ángulos:

Ejemplo:

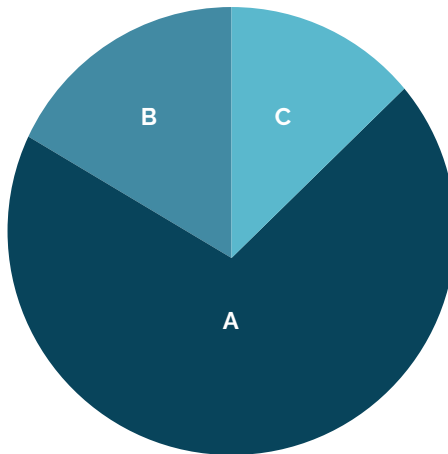
Ocupación	Número	Ángulos
Farmacéutico	12	A
Médico	43	B
Enfermera	25	C
TOTAL	188	360 °

Los ángulos son calculados de la siguiente manera:

$$A = \frac{120 * 360}{188} = 230^\circ$$

$$B = \frac{43 * 360}{188} = 82^\circ$$

$$C = \frac{25 * 360}{188} = 48^\circ$$



Interpretación:

- Los médicos son el grupo más numeroso (43 personas), lo que representa 82° del gráfico circular.

- Las enfermeras son el segundo grupo más numeroso (25 personas), lo que representa 48° del gráfico circular.
- Los farmacéuticos son el grupo menos numeroso (12 personas), lo que representa 23° del gráfico circular.

PRESENTACIÓN DE DATOS CUANTITATIVOS (NUMÉRICOS)

1.- DISTRIBUCIÓN DE FRECUENCIA

En primer lugar, lo que debemos hacer es reagrupar los datos para realizar la distribución de frecuencias.

Esto consiste en predeterminar un número determinado de grupos, como, por ejemplo: los grupos de edad, y un conteo de los números de observaciones que tiene cada grupo o clase. Estos grupos deben ser ordenados y constituidos en cada uno de tal forma que cada valor esté dentro de cada grupo o clase, que sea mutuamente excluyente siempre que se le considere un evento y que todos los rangos de valores estén cubiertos.

Por ejemplo: los latidos cardiacos

Los latidos cardíacos de los estudiantes en el primer año de medicina son:

59, 72, 58, 65, 77, 83, 72, 77, 62, 62 latidos por minuto

La distribución de los valores puede uno distribuirlos de la siguiente manera:

55 – 64 latidos/minuto	4	50 – 59 latidos/minuto	2
65 – 74	3	o 60 – 69	3
75 – 84	3	70 – 79	4
		80 – 89	1

Términos que se consideran para la redistribución de los valores en los grupos:

Frecuencia Absoluta (if): hace referencia al número de observaciones (n) encontrados en un grupo o clase.

En nuestro ejemplo, la frecuencia absoluta en el grupo de 65 a 74 es 3

Frecuencia Relativa (fi): es igual al número de un grupo sobre el total.

En nuestro ejemplo: 65 – 74 es de 3, por lo tanto:

$$f_i = \frac{3}{10} = 33\%$$

Frecuencia Acumulativa (Ni): es igual a la suma de la frecuencia (absoluta o relativa) de uno de los grupos, más la suma de todos los grupos que lo antecedieron.

En nuestro ejemplo:

La frecuencia acumulativa en el grupo de 65 a 74 es de 7, que equivale a $7/10$ (70 %) resultante de la suma del grupo que antecede (55 a 64) de 3. Si añadimos el grupo 75 a 84, la frecuencia acumulativa sería del 100 %.

Guía para la selección de grupos en la Distribución de Frecuencias

1. El número de grupos no debería ser ni muy grande ni muy pequeño. Debemos considerar que pocos grupos ensombrecen la información y demasiados grupos, por el contrario, no clarifican el problema. Uno puede encontrar grupos vacíos o grupos con muy pocas observaciones. Es más práctico empezar con una distribución más bien de pocos grupos debido a que estos pueden ser barridos en los grandes grupos (para cambiar de grupos grandes a pequeños se tendría que rehacer de nuevo los datos).

Para estimar el número óptimo de grupos, considerar la siguiente regla propuesta por Herbert Sturges en 1926, la cual establece que el número de intervalos debe ser aproximadamente igual a:

Regla de Sturges $\longrightarrow n = 1 + 3,3 * \log N$
(N: número de observaciones)



Ejemplo: observaciones de 100 individuos

$\text{Log } 100 = 2$

$n = 1 + 3,3 * 2 = 8$ grupos

2. Los límites inferior y superior de cada grupo corresponden a la ocurrencia de los datos crudos. Si los datos de la edad son recogidos en meses, los límites de los grupos no pueden ser en semanas.
3. Los intervalos deben ser de igual número (la diferencia entre el límite menor de un grupo y el límite menor del siguiente grupo) son convenientes y facilitan futuros cálculos, aunque esto no es necesario. Uno puede encontrar grupos de edad como, por ejemplo:

5 – 9 años

10 – 14 años

15 – 44 años

45 – 69 años

4. Los grupos tienen que ser mutuamente excluyentes. Esto frecuentemente sucede en la distribución de los grupos de edad. Los grupos de edad deberían leerse así: < 1 año o (0 años), 1 – 4 años, 5 – 9 años, etc. Un niño estaría en la edad de 5 – 9 años desde el cumplimiento de su quinto cumpleaños hasta el día anterior de su décimo cumpleaños.

5. Los intervalos finales abiertos deberían ser evitados, por ejemplo: grupo de 65 + años.
6. Para mejor comprensión de la marca de clase se sugiere establecer la diferencia entre los dos límites del mismo intervalo.

EJEMPLO

Se tienen los siguientes datos de altura de un grupo de estudiantes:

150 cm, 152 cm, 154 cm, 156 cm, 158 cm, 160 cm, 162 cm,
164 cm, 166 cm, 168 cm, 170 cm

Solución:

1. Aplicación de la Regla de Sturges:

Calcular el número de intervalos (n):

$$n = 1 + 3,322 * \log(n)$$

$$n = 1 + 3,322 * \log(11)$$

$$n \approx 5,32$$

Redondeando a un número entero, se establecen 5 intervalos.

2. Calcular la amplitud de clase (A):

Calcular el rango (R):

$$R = \text{Valor máximo} - \text{Valor mínimo}$$

$$R = 170 \text{ cm} - 150 \text{ cm}$$

$$R = 20 \text{ cm}$$

Calcular la amplitud de clase:

$$A = R / k$$

$$A = 20 \text{ cm} / 5$$

$$A = 4 \text{ cm}$$

3. Establecer los intervalos:**Primer intervalo:**

Límite inferior (Li): 150 cm

Límite superior (Ls): 154 cm ($Li + A - 1$)

Marca de clase (MC): $(Li + Ls) / 2 = 152 \text{ cm}$

Segundo intervalo:

Li: 154 cm

Ls: 158 cm ($Li + A - 1$)

MC: 156 cm

Tercer intervalo:

Li: 158 cm

Ls: 162 cm ($Li + A - 1$)

MC: 160 cm

Cuarto intervalo:

Li: 162 cm

Ls: 166 cm ($Li + A - 1$)

MC: 164 cm

Quinto intervalo:

Li: 166 cm

Ls: 170 cm ($Li + A - 1$)

MC: 168 cm

4. Tabla de frecuencias con marca de clase y diferencia entre límites:

Intervalo	Límite inferior	Límite superior	Marca de clase	Diferencia entre límites	frecuencia
1	150 cm	154cm	152cm	4cm	3
2	154 cm	158cm	156cm	4cm	4
3	158 cm	162cm	160cm	4cm	2
4	162 cm	166cm	164cm	4cm	1
5	166 cm	170cm	168cm	4cm	1

**Observación:**

La diferencia entre los dos límites del mismo intervalo es siempre la amplitud de clase (A). En este caso, la diferencia entre 150 cm y 154 cm (primer intervalo) es de 4 cm, al igual que la diferencia entre 154 cm y 158 cm (segundo intervalo), y así sucesivamente.

Conclusión:

La Regla de Sturges permite determinar el número de intervalos adecuado para una distribución de datos. La amplitud de clase se calcula dividiendo el rango entre el número de intervalos.

La marca de clase se calcula como la media del límite inferior y el límite superior de cada intervalo. La diferencia entre los dos límites del mismo intervalo es siempre la amplitud de clase.

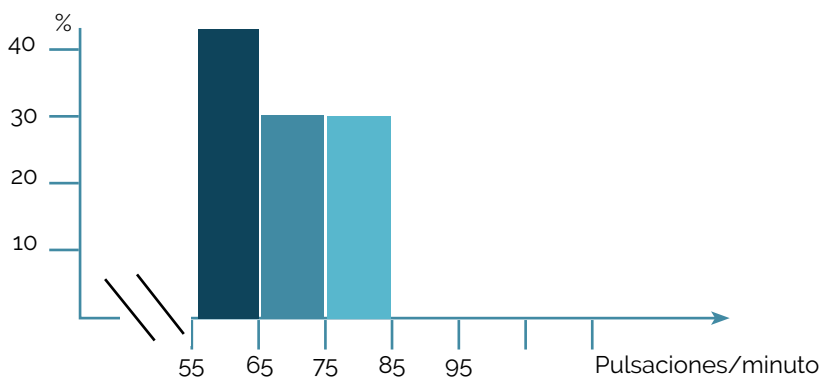
2. TIPOS DE REPRESENTACIÓN GRÁFICA**EL HISTOGRAMA**

El histograma es una representación gráfica de barras donde el ancho de los rectángulos corresponde a la posición de los grupos en el eje numérico X. Los rectángulos en este caso serían contiguos, es decir, estarían tocándose. La superficie del rectángulo es proporcional a la frecuencia del grupo (Ardila Duarte, 2022; Martín Andrés & Luna del Castillo, 2004).

Cuando el ancho de la frecuencia es variable, la altura de las columnas no refleja la frecuencia absoluta o relativa.

Ejemplo de los latidos cardiacos:

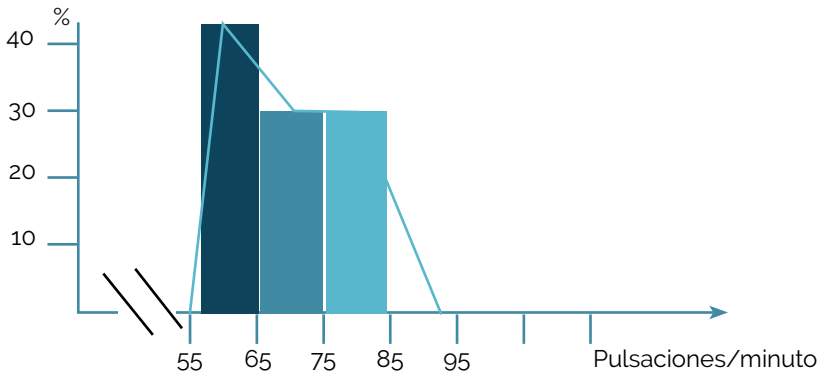
HISTOGRAMA DE LA FRECUENCIA DE LOS LATIDOS CARDIACOS EN 10 ESTUDIANTES



EL POLÍGONO DE FRECUENCIA

Este gráfico se obtiene cuando conectamos los puntos medios de los grupos en la parte alta de las barras del histograma grupo (Ardila Duarte, 2022; Martín Andrés & Luna del Castillo, 2004). El polígono de frecuencia permite desplegar varias distribuciones en una figura.

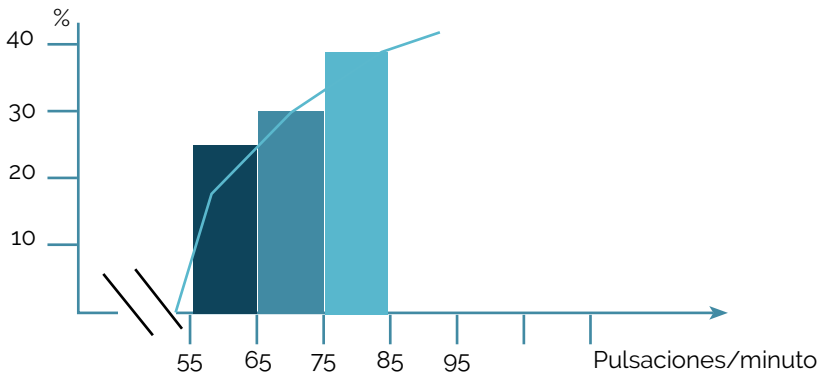
Obsérvese el polígono de frecuencia en el ejemplo de los latidos cardiacos de los estudiantes.



POLÍGONO DE FRECUENCIA ACUMULATIVA

En algunos casos es importante demostrar la frecuencia acumulativa de una serie de observaciones en un gráfico.

Para conocer cuántos pacientes se necesitan para realizar el seguimiento, uno podría conocer el periodo medio de tiempo luego de cierto tiempo que la enfermedad reapareció.



II.- DESCRIPCIÓN DE LOS DATOS

La representación gráfica de los datos es un primer paso indispensable en el análisis de los mismos, pero tiene dos limitaciones:

1. En primer lugar, si alguien le pregunta que describa lo fundamental de lo que está describiendo, todo lo que puede hacer es coger un trozo de papel y trazar un gráfico.
2. En segundo lugar, no es mucho lo que podemos hacer con los resultados, salvo mostrarlos. No podemos comparar fácilmente los resultados de dos o más grupos o ver si existen diferencias importantes entre ellos; sin embargo, sirve de gran ayuda poder resumir los resultados en unos pocos números y no es por casualidad que existan. Los dos más importantes que tenemos son las medidas de centralización y las de dispersión.

La representación gráfica de datos juega un papel fundamental en la interpretación y análisis de información. Su importancia radica en su capacidad para traducir información compleja en imágenes comprensibles, facilitando la identificación de patrones, tendencias y relaciones que podrían pasar desapercibidas en formato numérico.

A.- DESCRIPCIÓN DE LOS DATOS CUALITATIVOS

1.- DATOS NOMINALES

La mejor forma de describir los datos nominales es expresando el número de cada grupo como una proporción del total.

EJEMPLO

Se tienen los siguientes datos de altura de un grupo de En la ciudad de Guayaquil, 55 personas fueron examinadas, de ellas, 32 son del Carchi, 20 del Azuay y 3 del Cañar.

La proporción del Carchi es: $32/55 = 0,58$

La proporción normalmente es expresada en porcentaje = 58 %

Interpretación:

La proporción de personas del Carchi es de 0,58, lo que equivale al 58%.

Las personas del Azuay representan el 36,36% (20/55) de la muestra.

Las personas del Cañar representan el 5,45% (3/55) de la muestra

2.- DATOS ORDINALES

Los datos ordinales son similarmente descritos como los nominales como en el ejemplo ya descrito, en el cual se debería respetar el orden jerárquico.

EJEMPLO

Presencia de bacterias en la orina de 90 pacientes comprendidos en las edades de 7 a 15 años en el centro de salud de Azogues.

PRESENCIA DE BACTERIAS	NÚMERO OBSERVADO	%
+	50	56
++	26	29
+++	14	15
TOTAL	90	100

B.- ORGANIZACIÓN DE LOS DATOS CUANTITATIVOS

Los datos son generalmente organizados tomando dos parámetros:

1. Un parámetro de localización o tendencia central (posición central)
2. Un parámetro de dispersión (como los datos están distribuidos)

El término parámetro es usado para referirse a las características de la población. En el caso de una muestra de una población las características son llamadas estadísticas.

Para el mejor análisis y comprensión, seguiremos usando el dato del ejemplo de los latidos del corazón.

1.- PARÁMETROS DE TENDENCIA CENTRAL

a.- LA MEDIA (LA MEDIA ARITMÉTICA)

Descrita con el símbolo \bar{X} , cuando se refiere a una muestra y μ cuando se refiere a la población en general (del cual la muestra fue tomada).

1.- Cuando las observaciones no están agrupadas

La media es la suma de todos los valores divididos para el número de observaciones.

$$\bar{X} = \frac{\sum x}{n}$$

Σ = suma

x = valor de las variables que son estudiadas

n = número de observaciones

En el ejemplo de los latidos cardiacos

$$\Sigma x = 687$$

$$n = 10$$

$$\bar{x} = \frac{687}{10}$$

$$\bar{x} = 68,7$$

2.- Cuando las observaciones están agrupadas

GRUPO DE EDAD	NÚMERO DE OBSERVACIONES
55 - 59	2
60 - 64	2
65 - 69	1
70 - 74	2
75 - 79	2
80 - 84	1

Para calcular la media, tomamos el valor central de cada grupo multiplicado por el número de observaciones de ese grupo.

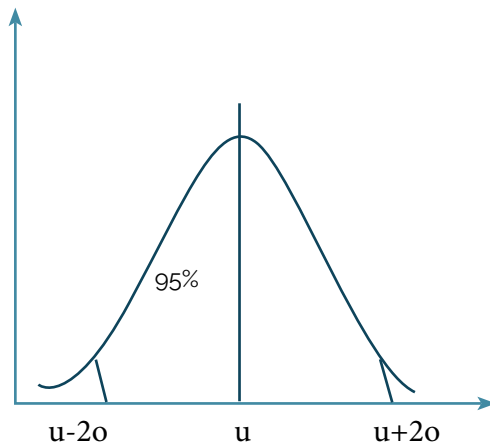
GRUPO DE EDAD	VALOR CENTRAL	NÚMERO DE OBSERVACIONES	TOTAL
55 - 59	57,5	2	115,0
60 - 64	62,5	2	125,0
65 - 69	67,5	1	67,0
70 - 74	72,5	2	145,0
75 - 79	77,5	2	155,0
80 - 84	82,5	1	82,5
		10	690,0

$$\bar{X} = \frac{\sum x}{n}$$

$$\bar{X} = \frac{690}{10} = 69,0$$

Nótese que calculando " \bar{X} " de los datos sueltos el resultado es de 68,7 latidos por minuto.

La media aritmética es fácil de calcular, la mayoría de calculadoras de mano están programadas directamente para hacerlo, sin embargo, la media aritmética sólo tiene sentido cuando la distribución es normal y no hay datos muy extremos dentro de la llamada **Campaña de Gauss**.



En una distribución bimodal, por ejemplo, la media no permite una práctica organización de la información. Por lo tanto, antes de calcular la media, resulta necesario estudiar la distribución de los datos.

b.- LA MEDIANA

La mediana está representada por el símbolo "**Me**". Es el valor central de la distribución en un grupo de datos ordenados, así; hay algunos valores con bajos y altos de acuerdo a la ubicación de la mediana.

Si se tiene un número impar de elementos en tu conjunto de datos, la mediana será el elemento del medio. Si se tiene un número par de elementos, la mediana será el promedio de los dos elementos del medio.

La posición de la mediana (Me):

$$\text{Posición de la mediana} = \frac{n + 1}{2}$$

n = número de datos

Cálculo:**Si n=impar****Me = (n+1) / 2 → Valor de la observación****Ejemplo:**n: 1,2,3,4,5 **Me=3**

Si existen 11 observaciones, aplicando la fórmula tendríamos:

$$(11+1) / 2 = 6$$

Así que, el valor de la sexta observación (ordenada) es la mediana.

Cálculo:**Si n = par.** A menudo, la posición media cae entre dos observaciones:**Me = (n+1) / 2 → Media de las observaciones****Ejemplo:**n: 1,2,3,4,5,6 **Me=3,5**

Esto es el resultado de aplicar la fórmula para un conjunto de datos par (6 es número par) y dividir entre 2

$$(6+1) / 2 = 3,5$$

En nuestro ejemplo de 10 observaciones (n=10) la mediana corresponde a un valor entre la quinta y la sexta unidad de observación. Entonces; si la quinta observación fue de 65 latidos/minuto y el sexto valor fue de 72 latidos/minuto, entonces la mediana sería:

$$(65 + 72) / 2 = 68,5 \text{ latidos/minuto}$$

c.- La moda

La moda en estadística es el valor que aparece con más frecuencia dentro de un conjunto de datos. Es una medida de tendencia central que indica el elemento más común.

1.- Cuando los datos no están agrupados

La moda es igual al valor más frecuente.

Ejemplo:

59, 72, 58, 65, 77, 83, 72, 77, 62, 62 latidos por minuto

En nuestro ejemplo hay tres modas: 62, 72, 77

2.- Cuando los datos están agrupados en grupos con intervalos constantes

La moda es igual al grupo o los grupos con la más alta frecuencia absoluta simple.

La moda de los grupos está determinada por los **datos** y por **los límites de los grupos**. En nuestro ejemplo; el grupo 55 a 59, 60 a 64, 70 a 74, 75 a 79, son todos los grupos de la moda (cada grupo tiene una frecuencia de 2).

Por su parte; si los datos: 58, 59, 62, 62, 65, 72, 72, 77, 77, 83 latidos por minuto estarían agrupados tendríamos un grupo de 55 a 69 y otro grupo de 70 a 79. En este caso, el grupo 70 a 79 tendría una frecuencia de 4 y dos modas: 72 y 77.



Una distribución con una única moda es llamada **unimodal**. Cuando hay varias modas la distribución es **multimodal**.

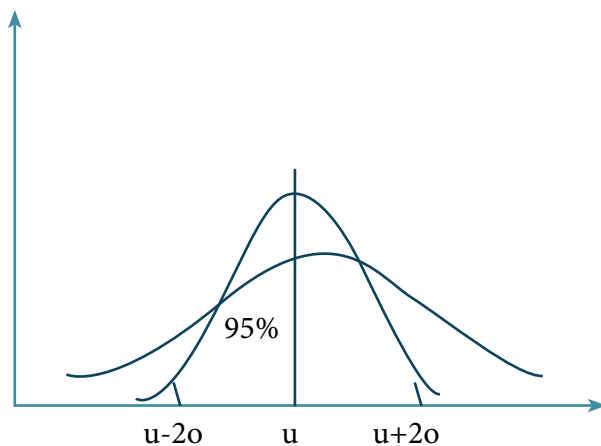
El uso de la moda.

- Da el punto o categoría con el más alto número de casos. Esto puede influenciar en el grupo prioritario o el punto que determina una persona.
- Da información acerca de la complejidad de la distribución:
 - Una distribución unimodal puede ser compatible con una distribución teórica
 - Una distribución multimodal indica que el grupo de datos es complejo y puede necesitar exploración para encontrar a futuro la razón de esta multimodalidad (la población puede consistir de dos distintas subpoblaciones)
- La moda ayuda a seleccionar apropiadamente los parámetros de dispersión (Sokal & Rohlf, 2002).

2.- PARÁMETROS DE DISPERSIÓN

El siguiente gráfico demuestra que los parámetros de localización central no describen completamente la situación dada. La dispersión de las observaciones ha sido considerada.

Campana de Gauss.



Hay parámetros de dispersión que corresponde a cada una de las medidas de localización central.





a.- EL RANGO

Este parámetro (corresponde a la moda), es la diferencia entre el valor más alto y el más bajo.

En el ejemplo de los latidos por minuto: el rango más alto es 83 y el rango más bajo o inferior es 58, por lo tanto:

59, 72, 58, 65, 77, 83, 72, 77, 62, 62

$$83 - 58 = 25 \text{ latidos por minuto}$$

Uso del rango:

Da una idea de la variabilidad de los datos, pero:

- El rango se incrementará con la muestra simple.
- Es extremadamente influenciada por los extremos.
- Únicamente considera los dos valores extremos.
- Las pruebas basadas en los extremos son complicadas (Sokal & Rohlf, 2002).

b.- LOS CUARTILES (y rango del intercuartil)

Los cuartiles son medidas de posición que dividen un conjunto de datos ordenados en cuatro partes iguales. Son útiles para describir la distribución de los datos y comprender cómo se encuentran divididos en diferentes grupos (López Segovia, 2020).

Cuartil 1 (Q1): El valor que separa el 25% inferior del conjunto de datos del 75% restante.

Cuartil 2 (Q2): La mediana del conjunto de datos, que divide los datos en dos mitades iguales (50%)

Cuartil 3 (Q3): El valor que separa el 75% superior del conjunto de datos del 25% inferior.

Para encontrar los cuartiles, **clásicamente**, realizamos las siguientes operaciones:

La posición del cuartil 1: $\frac{n + 1}{4}$

La posición del cuartil 3: $\frac{(n + 1) * 3}{4}$

Pero, para calcular los cuartiles 1 y 3, nosotros seguimos el método modificado de Turkey (USA):

Posición (cuartil 2) = Posición (mediana) = $\frac{n + 1}{2}$

Posición (cuartil 1) = $\frac{\text{Posición (mediana)} + 1}{2}$

Posición (cuartil 3) = El mismo valor, pero contado desde el valor más alto

Cuando la posición de la mediana nos corresponde a un número entero, es redondeado hacia el valor bajo para calcular la posición del cuartil 1 y 3.

EJEMPLO

Si: $n = 11$;	Posición de la mediana	=	$(11+1) / 2 = 6$
	Posición el cuartil 1 (q_1)	=	$(6 + 1) / 2 = 3,5$
Si: $n = 10$;	Posición de la mediana	=	$(10+1) / 2 = 5,5$
	Posición el cuartil 1 (q_1)	=	$(5 + 1) / 2 = 3$

Los cuartiles calculados de esta forma difieren significativamente de los valores calculados en forma clásica. Se prefiere este método por ser simple y porque los cuartiles obtenidos son los que se van a usar en las representaciones estadísticas futuras como el gráfico de caja y en el cálculo del intervalo de confianza alrededor de la mediana.

c.- LA VARIANZA Y LA DESVIACIÓN ESTÁNDAR

Estos son parámetros de dispersión alrededor de la media.

El promedio de la diferencia entre todos los valores y la media parece ser una buena medida de dispersión alrededor de esta media. Pero, por supuesto, este promedio es siempre cero.

$$\frac{\sum_{i=1}^N (X_i - \bar{X})}{N} = 0$$

En el ejemplo del pulso:

Media = 68,7

58 - 68,7	=	-10,7	}	-37,5
59 - 68,7	=	- 9,7		
62 - 68,7	=	- 6,7		
62 - 68,7	=	- 6,7		
65 - 68,7	=	- 3,7		
72 - 68,7	=	+ 3,3	}	+37,5
72 - 68,7	=	+ 3,3		
77 - 68,7	=	+ 8,3		
77 - 68,7	=	+ 8,3		
83 - 68,7	=	+14,3		
0				

Uno puede tomar los valores absolutos de todas las diferencias, pero por razones matemáticas, se prefiere el cuadrado de las diferencias:

$(58 - 68,7)^2$	=	$(-10,7)^2$	=	114,49
$(59 - 68,7)^2$	=	$(-9,7)^2$	=	94,09
$(62 - 68,7)^2$	=	$(-6,7)^2$	=	44,89
$(62 - 68,7)^2$	=	$(-6,7)^2$	=	44,89
$(65 - 68,7)^2$	=	$(-3,7)^2$	=	13,69
$(72 - 68,7)^2$	=	$(+3,3)^2$	=	10,89
$(72 - 68,7)^2$	=	$(+3,3)^2$	=	10,89
$(77 - 68,7)^2$	=	$(+8,3)^2$	=	68,89
$(77 - 68,7)^2$	=	$(+8,3)^2$	=	68,89
$(83 - 68,7)^2$	=	$(+14,3)^2$	=	204,49
				<hr/>
				679,10

Esta suma dividida por $(n - 1)$ es llamada **VARIANZA** de una muestra y se representa con S_2 .

$$S_2 = \frac{\sum (X_i - X)^2}{N - 1} \text{ en nuestro ejemplo } = \frac{679,10}{10 - 1} = 75,12$$

Teóricamente, hay argumentos para dividir por $(n - 1)$. Corresponde a los grados de libertad, el cual es un número independiente en una serie, por ejemplo: el número de valores en una serie menos el número de restricciones impuestas sobre esos valores.

En nuestro ejemplo hay 10 diferentes $(X_i - X) = 10$.

Hay una limitación: La suma de todos los valores = 0

Como consecuencia hay $10 - 1 = 9$ grados de libertad.

La interpretación de la varianza es difícil, pero, las unidades originales son los cuadrados. En nuestro ejemplo del pulso, la media es 68,7 latidos/minuto y la varianza es 75,12 (latidos / minuto) ², un concepto difícil de imaginar.

Una forma de obtener las unidades originales es tomando el cuadrado:

$$\sqrt{\text{Varianza}} = \text{Desviación Estándar} = \sqrt{\frac{E (X_i - X)^2}{N - 1}}$$

$$\sqrt{75,12 \text{ (l/m)}^2} = 8,67 \text{ latidos/minuto}$$

Interpretación de la Desviación Estándar

La desviación estándar es una medida de dispersión alrededor de la media. La desviación estándar únicamente tiene sentido cuando la distribución de los valores es normal.

La distribución estándar depende de las unidades de observación realizadas, así que, la desviación estándar es difícil de comparar cuando tenemos desviaciones estándar de dos diferentes variables. Sin embargo, esta comparación puede ser posible cuando las desviaciones estándar están

estandarizadas, lo cual puede ser hecho dividiendo estos por la media.

$$\frac{\text{Desviación estándar (DE)}}{\bar{X}} * 100 = \text{coeficiente de variación (CV)}$$

- El coeficiente de variación no tiene dimensiones. Esto permite la comparación entre diferentes grupos, aún si estos son medidos en diferentes unidades.
- El coeficiente de variación elimina el efecto de la magnitud de los datos porque es expresado como un porcentaje de la \bar{X} . Si el CV es pequeño, todos los valores encontrados son concentrados alrededor de la media. Si el CV es largo, hay una amplia dispersión.

Ejemplo:

En una población de 566 personas saludables, la media de la desviación estándar del pulso, la presión sistólica y diastólica es:

	Media	DE	CV
Latidos por minuto	77,3 l/m	12,8 l/m	16,6
Presión sistólica	128,8 mmHg	13,1 mmHg	10,2
Presión diastólica	79,7 mmHg	9,4 mmHg	11,8

La variabilidad de la presión sistólica y diastólica es comparable.

III.- TRANSFORMACIÓN DE DATOS CUANTITATIVOS Y CUALITATIVOS

En epidemiología, se requiere con frecuencia transformar los datos numéricos dentro de los datos ordinales, por ejemplo: los valores de hemoglobina son divididos en dos grupos:

$< 10,5$ g Hb = anemia
 $\geq 10,5$ g Hb = normal

O pueden ser divididos en tres grupos:

$<$ de 6 g Hb = muy anémico,
6,0 a 10,4 g Hb = anémico,
 $\geq 10,5$ g Hb = normal.

Una verdadera descripción de los datos cuantitativos comprende:

1. Un resumen de los parámetros de posición y dispersión y el estudio de esos parámetros.
2. Una transformación (de los datos numéricos) a una escala ordinal, por ejemplo:
 - No anemia
 - Anemia moderada
 - Anemia severa

3.- Si desea una transformación a una escala dicotómica, por ejemplo:

Anemia / no anemia

Los puntos de corte (árbol/hojas y límite de grupo) pueden ser definidos basados en:

- a. Valores tomados de la literatura sobre la opinión de expertos, etc.
- b. La información contenida en los datos. En este caso, la moda, la mediana, o los cuartiles pueden guiar la decisión.

Las ventajas de esa transformación son:

- Resumen los datos
- Facilita la aplicación de test estadísticos, por ejemplo, chi cuadrado (χ^2)
- Facilita la comunicación con el público en general

Las desventajas de esta transformación son:

- Los problemas de sensibilidad y especificidad
- La pérdida de información

CONCLUSIONES

Conceptos considerados:

- Gráfico de barras
- Gráfico de sector
- Distribución de frecuencia
- Histograma
- Polígono de frecuencia
- Parámetros de tendencia central y de dispersión
- Media
- Mediana
- Moda
- Varianza
- Desviación estándar
- Coeficiente de variación
- Cuartiles
- Rango
- Transformación de grupo de datos

Después de estudiar este capítulo, Ud. Debería ser capaz de:

- Representar gráficamente los datos e interpretar los mismos
- Resumir los datos en forma correcta e interpretar los parámetros de reducción
- Transformar datos cuantitativos a cualitativos.

EJERCICIOS

1. Procedente de un colegio recomendado por la superioridad (moral y de otros tipos) del ABP-PG (es decir, Aprendizaje Basado en los Problemas, en Pequeños Grupos), hacemos un estudio colocando la mitad de los alumnos de manera aleatoria en las clases ABP-PG, y la otra mitad, en las clases tradicionales. Al final, medimos las siguientes variables.
 - a. Resultados de un examen final.
 - b. Tiempo para realizar el examen final (sin límite de tiempo).
 - c. Basándose en un periodo de 5 años después de finalizar los estudios, el número de artículos rechazados por cada persona por contener análisis de datos incorrectos.
 - d. El tipo de dolencia (jaqueca, claustrofobia o tensión) desarrollada por los estudiantes durante la clase (en este caso, los dos grupos juntos).

Para cada una de estas variables, calcular la *mejor medida de centralización* y la *mejor medida de dispersión*.

2. Sólo para que practiques un poco. Calcula los siguientes parámetros estadísticos para este conjunto de datos (se da números sencillos de forma deliberada, de manera que no necesite calculadora):

4 8 6 3 4

- a. La media es _____
- b. La mediana es _____
- c. La moda es _____
- d. El rango o recorrido es _____
- e. La desviación estándar es _____

3. Un estudio de 100 individuos contiene por desgracia 5 datos desconocidos. Estos se han codificado como '99' en el ordenador. Supongamos que los verdaderos valores de las variables son:

$$\begin{array}{ll} X = 45,0 & \text{Mínimo} = 16 \\ DE = 5,6 & \text{Máximo} = 65 \end{array}$$

Si el estadístico analizó los datos como si los 99 fueran datos reales, ¿serán los parámetros estimados que siguen, *mayores, menores o iguales?*

- a. La moda.
- b. La mediana.
- e. La media.
- d. La desviación estándar.
- e. El recorrido.

ESTADÍSTICA DESCRIPTIVA. ANÁLISIS DE LOS DATOS EXPLORADOS

PARTE III

1.- Definición

El análisis de los datos explorados es un nuevo enfoque de la exploración estadística de los datos propuesto por Tukey (1977). De acuerdo a Tukey, los datos obtenidos en el campo tienen que:

- Ser analizados en función de los objetivos o de la hipótesis del estudio.
- Pero, al mismo tiempo uno tiene que estar atento a los detalles específicos encontrados en los datos recolectados.

Nosotros debemos aprender a examinar el grupo de datos de la misma forma que nosotros examinamos una placa radiográfica de tórax, en el cual se encuentra la forma del tórax, el mediastino, el lugar y tamaño del corazón, etc; sin embargo, resulta importante mirar antes, por ejemplo, el

sistema pulmonar, el cual puede ser de específico interés y para lo cual fue tomado la radiografía en primer lugar.

Al analizar el grupo de datos, uno debe estar atento para la pérdida, valores extremos, valores que aparentemente pueden estar perdidos por irregularidades en el patrón de distribución de los valores.

Para optimizar esta exploración, los datos deben ser presentados gráficamente.

Tukey ha desarrollado técnicas que permiten un rápido ordenamiento de los datos y su representación gráfica que son más simples que los tradicionalmente usados en la estadística descriptiva. Más aún, el autor propone un rápido procedimiento de estadística inferencial basada en la mediana.

Dos de las técnicas gráficas serán explicadas a continuación:

2.- Técnica de Tallo y Hoja

Es una técnica que clasifica, ordena datos y al mismo tiempo da una representación gráfica.

El principio es el siguiente: los datos por sí solos pueden ser usados como un instrumento para clasificar y representar. Los datos tienen que ser divididos en dos partes: la primera parte llamada **tallo**, ayuda en la clasificación y permite ordenar los datos; la segunda parte, llamada **hojas**, permite



individualmente manejar los datos en el tallo y jugar el rol de representación gráfica.

Procedimiento:

1. Seleccione el número de dígitos que formarán el tallo.
2. Dividir cada una de las observaciones en dos partes: la primera parte, donde la "cabeza" de la figura de los valores formarán el tallo y el remanente de la observación formará la hoja u hojas.
3. Realice una columna con todos los posibles valores del tallo en orden ascendente.
4. Anote al costado la parte de cada valor próximo a su correspondiente valor del tallo. Por ejemplo: 1 y 5, la primera figura (1, en este caso corresponde a la décima) sirve como el clasificador de los elementos y la segunda parte (5, unidad), sirve como la parte gráfica.

Aplicando este procedimiento a un ejemplo sobre latidos cardíacos de un grupo de estudiantes tendríamos lo siguiente.

Valores de latidos cardíacos: 59, 72, 58, 65, 77, 83, 72, 77, 62, 62

1. Primero identificamos el valor más bajo y el valor más alto

El valor más bajo = 58

El valor más alto = 83

2. Para el tallo, utilizamos las décimas y determinamos el valor del tallo. Así; encontramos que el valor más bajo tiene que ser 5 y el valor más alto tiene que ser 8.
3. Construimos el tallo

5	
6	
7	
8	

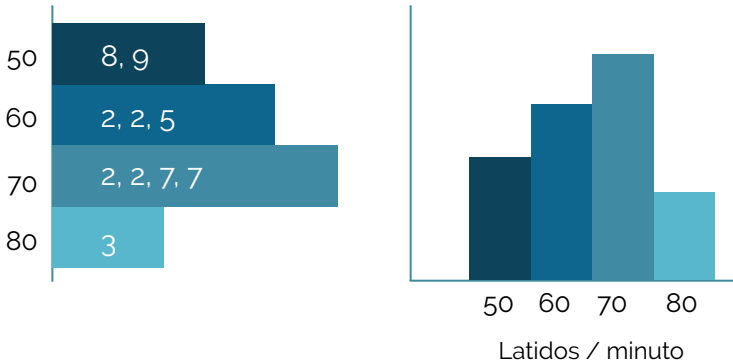
4. De los datos presentados en orden, (59, 72, 58, 65, 77, 83, 72, 77, 62, 62), extraemos la parte de las hojas y escribimos estos de acuerdo al correspondiente valor del tallo.

5	9, 8
6	5, 2, 2
7	2, 7, 2, 7
8	3

5. Posteriormente, ordenamos las hojas en orden ascendente

5	8, 9
6	2, 2, 5
7	2, 2, 7, 7
8	3

6. Utilizamos este orden para estudiar la forma general y la distribución de los datos.



7. Empleamos este orden para identificar los cuartiles

Esta técnica tiene las siguientes ventajas:

1. Fácil de construir y fácil de leer: la técnica permite clasificar en forma rápida.
2. Permite localizar en forma inmediata la mediana y los cuartiles.
3. Permite estudiar los datos originales:
 - a. el rango
 - b. la concentración de los valores
 - c. la simetría de los valores a ambos lados de la mediana
 - d. la continuidad en una serie (valores perdidos)
 - e. la presencia de los valores extremos

3.- Gráfico de Caja

La técnica del tallo y hojas permite conseguir un claro enfoque de todo el grupo de datos, pero hay un gran detalle en esta presentación. Para conseguir una idea general de la distribución, uno tiene que responder a las siguientes preguntas:

- a) ¿La distribución es normal?
- b) ¿Dónde colocamos la concentración de las observaciones?
- c) ¿Cuál es el rango de las observaciones?
- d) ¿Hay alguna pérdida de las observaciones (extremos)?

Tukey (1977) ha propuesto una representación gráfica, "el gráfico de caja" para responder a las interrogantes.

Construcción del gráfico de caja:

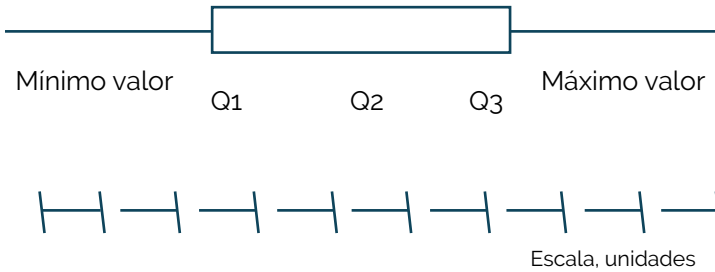
Se construye considerando los siguientes puntos:

- Los dos valores extremos
- Los cuartiles Q_1 y Q_3
- La mediana

Y es graficado de la siguiente forma:

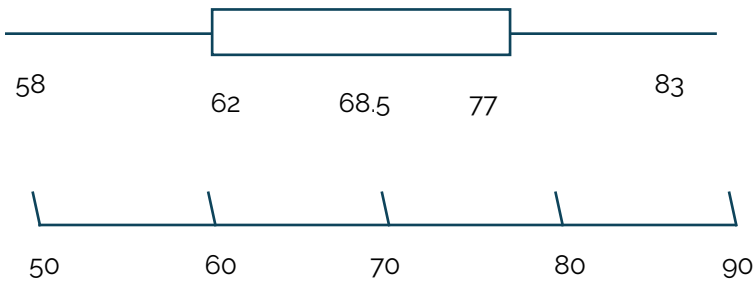
- Realice una escala en la coordenada "x" cubriendo todos los valores observados
- Dibuje una línea paralela a la escala que ayudará a la construcción de los bigotes

- Marque los cinco puntos (mínimo, máximo, cuartil 1, cuartil 3, y mediana), sobre esta línea.
- Dibuje una caja utilizando el Q1 y Q3 como los límites para la mediana dibuje una línea que atraviese la caja. Los extremos están conectados a la caja con una línea denominada bigote.
- No se olvide de poner el título, (tiempo, lugar y persona).



Apliquemos, la técnica al ejemplo de los latidos cardíacos de los 10 estudiantes:

LATIDOS CARDÍACOS DE UN GRUPO DE ESTUDIANTES



Propósito de la representación gráfica:

Organizar un grupo de datos cuando se tienen muchas observaciones situadas alrededor de la mediana y dispersas hacia los extremos, para lo cual se tiene que conocer:

- La posición del valor central
- El patrón de concentración de los valores alrededor del centro, y
- El rango entre los dos extremos de los valores.

En general, los valores alrededor del centro proveen una pequeña información específica, pero los valores cerca de los dos extremos pueden esconder información interesante. Los datos pueden ser normalmente dispersados alrededor del centro o pueden demostrar una mezcla en la distribución, pero también los extremos pueden ser observados.

Esta última posibilidad puede ser debido a:

- Errores de observación, de codificación o de transcripción; para lo cual se sugiere siempre chequear los datos originales.
- Valores reales excepcionales, los cuales pueden esconder información muy interesante.

Los detalles del gráfico de caja

Para analizar los dos extremos del grupo de datos (el 25% bajo Q_2 y el 25 % sobre Q_3), es importante individualizar los valores que están situados lejos del centro de la distribución.

¿En qué casos particulares, un valor llega a ser considerado que se encuentra en el extremo?

Para ello podemos realizar una distribución entre valores que están lejos del centro, lo que se llamarían **valores distantes**, y aquellos que están muy lejos del centro, se llamarían **valores excepcionales**.

Los valores distantes se encuentran entre $1\frac{1}{2}$ y 3 veces de los intervalos intercuartiles ($Q_3 - Q_1$) de los cuartiles.

Los valores excepcionales son encontrados fuera de las 3 veces de los intervalos intercuartiles de los cuartiles.

Cada uno de los valores distantes y excepcionales debe ser individualizado.

Ejemplo:

El peso en Kilogramos de un grupo de niños es el siguiente:

$3\frac{1}{2}$, 5, 9, 10, 10, $10\frac{1}{2}$, 11, 11, 12, 12, 13, 15, 16 Kg.

n = 13

La mediana = 11 Kg.

q_1 = 10

q_2 = 12

Intervalo intercuartil = $(12 - 10) = 2$ Kg.

Construcción del detalle del gráfico de caja:

a.- Colocar Q1, Q2 (mediana), Q3 en el gráfico de caja

b.- Encontrar los límites de los valores distantes y excepcionales

Valores distantes: entre (4 y 7 Kg) y (15 y 18 Kg).

$$10 - (1 \frac{1}{2} \times 2) = 7$$

$$12 - (3 \times 2) = 4$$

Hay un valor distante en esta área (entre 4 y 7) que es de 5 Kg.

$$12 + (1 \frac{1}{2} \times 2) = 15$$

$$12 + (3 \times 2) = 18$$

Hay también en el área un valor distante que es de: 16 Kg.

Los valores distantes se encuentran fuera de las 3 veces la distancia de los intercuartiles; así, por debajo de los 4 Kg. Y por encima de los 18 Kg.

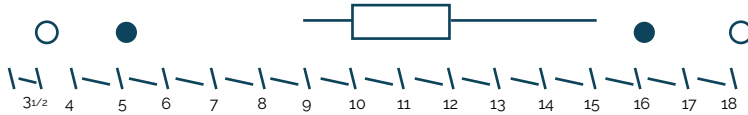
En este grupo de datos hay un valor excepcional: $3\frac{1}{2}$ Kg.

c.- Identificar el máximo y el mínimo valor de los valores no distantes

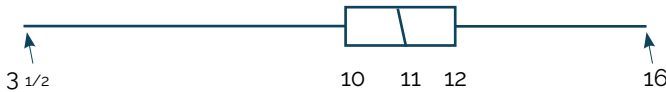
En nuestro caso 9 y 15 Kg.

d.- Graficar los valores distantes sobre el diagrama con un asterisco o círculo

e.- Graficar el valor excepcional con un círculo abierto



Un simple gráfico de caja estaría representado de la siguiente manera:



Uno puede fácilmente en este gráfico observar de mejor forma la información de los valores extremos.

Interpretación del gráfico de caja

Para la interpretación del gráfico de caja, los siguientes puntos deben ser considerados:

- La simetría hacia la mediana (del rango de los cuartiles y del rango total).
- La concentración de $Q_3 - Q_1$.
- Los valores distantes.

- Las zonas de los valores extremos no deberán contener más que el 1% de las observaciones.
- Los valores excepcionales: cada valor excepcional, bajo sospecha, sino constituye un error de medida, de codificación, o transcripción, puede ser encontrado; en este sentido, el valor tiene que ser considerado como no normal y la causa de esta anormalidad tiene que ser estudiada.

Finalmente, el gráfico de caja puede ser fácilmente usado para comparar la distribución de múltiple grupo de datos.

4.- Conclusión

Los siguientes conceptos tienen que ser discutidos:

Técnica del tallo y hojas

Gráfico de caja

Simple gráfico de caja

Valores extremos

Valores distantes

Valores excepcionales

Ud., debería ser capaz de construir un diagrama tallo y hojas e interpretar el gráfico de caja.

ESTADÍSTICA DESCRIPTIVA.

DESCRIPCIÓN ESTADÍSTICA DE DOS VARIABLES

PARTE IV

1.- Relación entre diferentes variables

Hasta ahora se ha llevado a cabo un análisis individual de las variables, sin embargo, en el campo de la epidemiología, el interés principal se centra en las relaciones que se establecen entre ellas.

Por ejemplo: El peso de un niño está en relación a su edad

La tasa de mortalidad infantil está en función de la cobertura de servicio de salud.

Entonces, tenemos que distinguir entre las variables dependientes y las variables independientes.

Una variable dependiente varía en función de la variable independiente. El nivel de la independencia de la variable permite explicar o predecir el nivel de la variable dependiente.

Por ejemplo: En una zona de salud denominada "Bienestar" 252 niños nacieron en el año 2000

Hasta el 31 de diciembre del 2000, 127 de ellos (50,4 %) fueron atendidos en el centro de salud y 125 (49,6 %) no fueron atendidos.

202 (80 %) de los niños viven en la ciudad y 50 (20 %) viven en el área rural.

Es interesante analizar cada variable, pero lo más interesante es examinar la asociación entre las dos variables.

Ejemplo: ¿La atención en el centro de salud depende del lugar de origen?

La respuesta a esta pregunta puede ser proveída por un análisis bivariante.

En un análisis bivalente, cuatro situaciones diferentes pueden ser distinguidas:

VARIABLE INDEPENDIENTE	VARIABLE DEPENDIENTE
CUANTITATIVA	CUANTITATIVA
CUALITATIVA	CUANTITATIVA
CUANTITATIVA	CUALITATIVA
CUALITATIVA	CUALITATIVA

2.- Análisis de dos variables cualitativas

En la simple forma hay dos modalidades por variable (si/no, síntomas presentes/síntomas no presentes, urbano/rural, etc.) y para esto una tabla de contingencia de 2×2 puede ser construido,

En el ejemplo de los 252 niños, encontramos que por cada niño una historia clínica fue abierta (si / no) y su origen (urbano / rural).

La tabla de contingencia 2×2 es construida de la siguiente manera:

	Origen		TOTAL
	Urbano	Rural	
Atención Médica			
Si	114	13	127
No	88	37	125
TOTAL	202	50	252

Generalmente, una tabla 2 x 2, se muestra como sigue:

VARIABLE Y	VARIABLE X		TOTAL
	Modalidad x1	Modalidad x2	
Modalidad Y1	X1 Y1	X2 Y1	X1 Y1 + X2 Y1
Modalidad Y2	X1 Y2	X2 Y2	X1 Y2 + X2 Y2
TOTAL	X1 Y1 + X1 Y2	X2 Y1 + X2 Y2	X1 Y1 + X2 Y1 + X1 Y2 + X2 Y2

X (origen) es la variable independiente

Y (atención médica) es la variable dependiente (depende de X que es el origen)

El objetivo del análisis es establecer el grado de dependencia de Y sobre X.

La consecuencia práctica de encontrar una dependencia es que esto abre las perspectivas de influencia de Y por el cambio de X.

En nuestro ejemplo, si hay una relación entre el origen y la tasa de atención médica en el centro de salud, uno puede subir la tasa de atención médica, abriendo un centro de salud en el área rural.

Es interesante resumir una tabla 2 x 2 con un parámetro. En epidemiología se prefiere utilizar la medida denominada **RIESGO RELATIVO (RR)** como medida de asociación.

$$\text{Riesgo Relativo} = \frac{X_1 Y_1 / (X_1 Y_1 + X_1 Y_2)}{X_2 Y_1 / (X_2 Y_1 + X_2 Y_2)}$$

En nuestro ejemplo:

La proporción de niños del área urbana es = $114 / 202$

La proporción de niños del área rural es = $13 / 50$

$$\text{Riesgo Relativo} = \frac{114 / 202}{13 / 50} = 2,17$$

Interpretación: la proporción de niños registrados que fueron atendidos en el centro de salud es 2,17 veces más alto en el área urbana que en el área rural.

Cuando no hay asociación el riesgo relativo es 1.

Resulta importante mencionar que una asociación también puede ser negativa, ejemplo: La proporción de niños atendidos del área urbana podría ser menor que la proporción de niños atendidos en el área rural.

REGISTRADOS	ORIGEN		TOTAL
	Urbano	Rural	
SI	53	28	81
NO	148	22	171
TOTAL	202	50	252

Entonces:

$$\text{Riesgo Relativo} = \frac{53 / 202}{28 / 50} = 0,47$$

Interpretación: La proporción de niños atendidos en el área urbana es 0,47 veces más alta que los niños atendidos en el área rural.

Como se puede observar, el RR indica tanto la dirección como la fuerza de una asociación observada. Los RR>1 indican una asociación positiva; los RR<1 indican una asociación

negativa. Cuanto más lejos se encuentre el RR de 1, más fuerte será la asociación.

3.- Asociación entre una variable cualitativa y una variable cuantitativa

Presentaremos un ejemplo usando el enfoque del análisis exploratorio de los datos.

Variable independiente: fumadores con dos modalidades:
fumadores / no fumadores

Variable dependiente: presión sistólica en mmHg

Presión arterial de los fumadores:

110, 110, 130, 130, 140, 140, 150, 160, 170, 170 mmHg

$n = 10$

Mediana (fumadores) = 140 mmHg

Q1 = 120 mmHg

Q3 = 160 mmHg

Presión arterial de los no fumadores:

90, 90, 100, 110, 110, 130, 140, 150, 160, 170 mmHg.

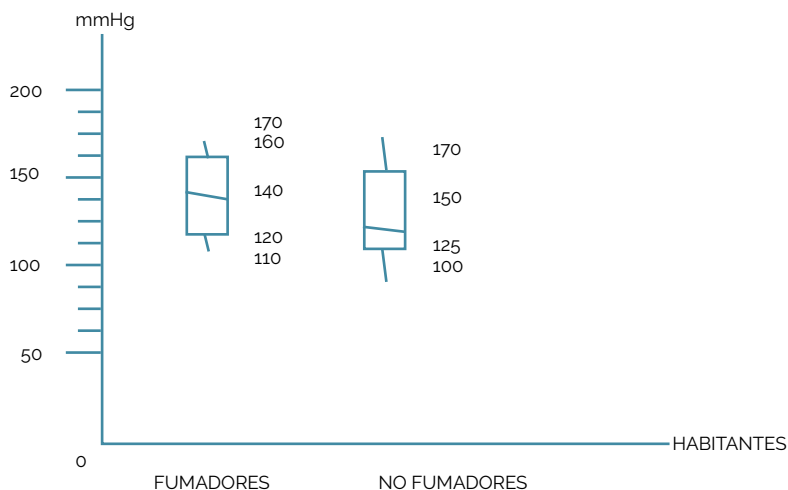
$n = 10$

Mediana (no fumadores) = 125 mmHg

Q1 = 100 mmHg

Q3 = 150 mmHg

REPRESENTACIÓN GRÁFICA (Gráfico de caja)



Interpretación: Esta figura demuestra que la presión arterial de los fumadores tiene una distribución simétrica, y que estos datos están bien concentrados alrededor de la mediana. Para los no fumadores, existe simetría, pero también existe una amplia dispersión.

El intervalo intercuartil de los fumadores es de 40 mmHg, mientras que de los no fumadores es de 50 mmHg.

La mediana de los fumadores es 15 mmHg más alto que de los no fumadores.



Los valores altos son iguales en ambos grupos, pero los valores bajos de los no fumadores es 20 mmHg más bajo que de los fumadores.

4.- Asociación entre variables cuantitativas

Ejemplo para las variables: edad y peso:

En la fase descriptiva del análisis de la relación entre dos variables cuantitativas, primero se empieza con la construcción de un gráfico de dispersión (una representación gráfica de la distribución de dos variables. Dicho diagrama nos permite evaluar:

- a.- El tipo de relación entre la variable dependiente (X) y la variable independiente (Y) (¿lineal? ¿Positiva?).
- b.- La fuerza de la relación (si la relación es lineal, uno puede correlacionar a través de del coeficiente de Pearson y calcular la dependencia de Y sobre X.
- c.- La presencia de los valores extremos.

Obsérvese el siguiente ejemplo donde se requiere estudiar la relación entre el peso de la madre y el peso del recién nacido.

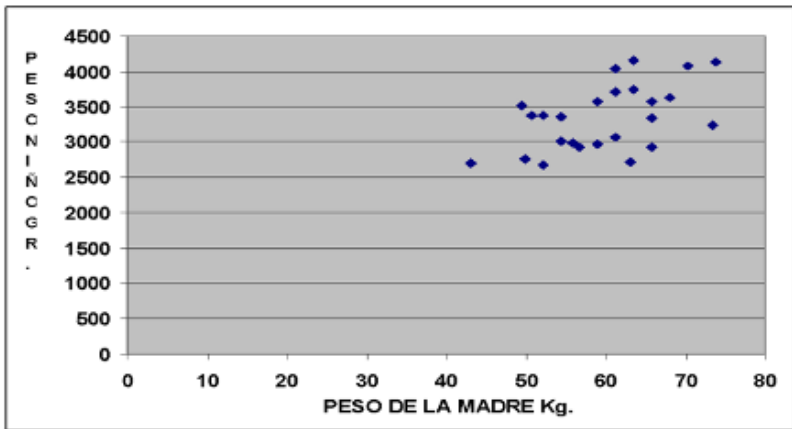
TABLA DE DATOS DE PESO DE LA MADRE Y LOS NIÑOS RECIÉN NACIDOS

#	PESO MADRES	PESO RECIÉN NACIDO
NIÑO	ANTES EMBARAZO Kg.	EN GRAMOS
1	49,5	3515
2	63,5	3742
3	68,0	3629
4	52,2	2680
5	54,4	3006
6	70,3	4068
7	50,8	3373
8	73,9	4124
9	65,8	3572
10	54,4	3359
11	73,5	3230
12	59	3572
13	61,2	3062
14	52,2	3374
15	63,1	2722
16	65,8	3345
17	61,2	3714
18	55,8	2991
19	61,2	4026
20	56,7	2920
21	63,5	4152
22	59	2977
23	49,9	2764
24	65,8	2920
25	43,1	2693

Por cada niño, existe un par de valores (peso de la madre / peso propio del niño).

Obsérvese la siguiente representación gráfica (gráfico de dispersión), donde cada par de valores es representado por un punto.

PESO DEL NIÑO EN FUNCIÓN DEL PESO DE LA MADRE



El gráfico demuestra que existe una moderada relación positiva lineal entre las variables.

Para cuantificar la fuerza de la relación lineal, se puede calcular el coeficiente de correlación de Pearson (r). Sus límites son -1 y 1 .

- -1: perfecta relación negativa lineal
- 0: ausencia de relación lineal
- 1: perfecta relación positiva lineal
- En nuestro ejemplo: $r = 0,52$

5.- Transformación

Como se ha discutido en la parte II (organización de los datos), las variables numéricas pueden ser transformadas a ordinal y dicotómicas.

Esto siempre permite a un estudio eventual, ver la existencia de asociación entre variables en una tabla de 2×2 .

6.- Conclusiones

Los siguientes conceptos tienen que ser discutidos:

1. Variable dependiente e independiente.
2. Asociación entre variables y las medidas de asociación.

Debería ser capaz de seleccionar un gráfico apropiado para demostrar la asociación entre variables, para interpretar medidas de asociación y conducir un análisis de asociación en una tabla de 2×2 .

EJERCICIO

En un centro de salud, la malnutrición es reportada como un serio problema de salud en los niños menores de tres años.

El gobierno lanza un programa de educación nutricional, que ha sido integrado a las unidades operativas de tipo ambulatorio. Este problema de malnutrición nunca ha sido evaluado en el pasado, no hay una línea de base disponible para hacer el diagnóstico previo y ningún plan fue realizado para evaluar el programa a nivel nacional.

El médico encargado del centro de salud tiene sus dudas sobre la eficacia del programa. Para documentar el problema, el médico decide organizar una pequeña investigación. Desafortunadamente, no hay censo disponible y los recursos humanos de su unidad operativa son más bien limitados.

El se ve forzado a tomar una muestra de los niños registrados en el centro de salud. Los resultados obtenidos son mostrados en la tabla 1.

Sin organización de los datos, es difícil conseguir una buena información sobre el contenido de esta tabla, más aún cuando las observaciones fueron hechas únicamente sobre 29 niños.

Preguntas:

1. Identificar la escala de medida de todas las variables.
2. Realice las representaciones gráficas.
3. Organice los datos.
4. Reduzca (dicotomíce) cuando sea apropiado.
5. Sería relevante el estudio de asociación entre variables y cuáles serían las recomendadas.
6. Si es necesario, utilice las tablas de contingencia.
7. Pruebe la asociación entre las variables.

Tabla # 1

Datos recolectados de 29 niños de edades comprendidas entre 0 a 59 meses y que fueron registrados en el centro de salud (muestra al azar)

No	Género	Edad meses	Nutrición A	Lección B	Diarrea C	Altura D	Grupo socio económico E	Distancia F	Procedencia G	Consultas H
1	M	28	N	0	-	85,3	S	10	V	4
2	M	15	M	2	+	74,7	C	3	R	10
3	F	6	S	2	+	66,2	W	2	V	7
4	M	5	N	1	-	70,0	S	4	R	10
5	M	8	M	1	+	72,8	W	9	V	5
6	F	35	S	1	-	87,4	W	7	V	3
7	F	11	N	2	+	69,7	S	0	V	12



8	M	8	M	0	+	71,9	W	5	R	11
9	M	49	N	0	-	102,9	C	8	V	6
10	F	42	N	0	-	98,2	C	4	R	9
11	F	51	N	0	+	98,1	W	6	R	3
12	M	50	S	0	-	94,9	W	0	V	7
13	M	51	N	0	-	97,3	W	8	V	5
14	F	26	M	1	+	84,6	C	4	V	5
15	M	45	N	0	-	90,8	W	7	V	6
16	M	30	S	1	+	85,1	W	3	R	7
17	F	42	N	0	-	92,1	W	9	V	2
18	M	42	N	0	+	93,5	C	11	R	3
19	F	23	M	1	+	86,9	W	1	R	6
20	F	49	N	0	-	92,2	C	7	V	3
21	M	33	N	2	-	88,4	W	4	R	12
22	M	37	N	2	-	90,8	C	2	R	13
23	F	34	S	3	+	89,5	W	2	V	4
24	M	9	M	3	+	69,3	W	3	V	10
25	F	19	S	2	-	75,6	C	1	N	5
26	M	12	N	2	-	75,5	C	1	N	15
27	M	28	N	3	-	95,3	S	4	R	10
28	F	59	N	0	-	107,0	W	5	R	8
29	M	40	N	0	+	97,9	C	6	V	7

CLAVES DE LA TABLA # 1

A = ESTADO NUTRICIONAL: N: normal
M: desnutrición moderada
S: desnutrición severa

B = NÚMERO DE CLASES SOBRE EDUCACIÓN NUTRICIONAL:
asistidas por la madre durante los previos tres meses

**C = EPISODIOS DE DIARREA Y/O INFECCIÓN RESPIRATORIA
AGUDA DURANTE EL ÚLTIMO MES**

+ = SI

- = NO

D = ALTURA EN CENTÍMETROS: acostado para niños de
0 – 36 meses y parados para niños mayores de 37 meses.

E = GRUPO SOCIO ECONÓMICO

G: trabajador del estado

C: pequeño productor

S: pequeño productor con subsidios del estado

W: agricultor contratado

F = DISTANCIA: entre la comunidad y el centro de salud, en
kilómetros

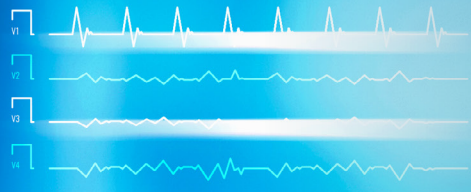
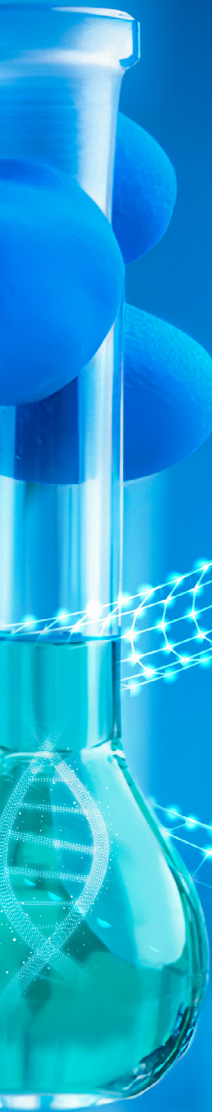
G = PROCEDENCIA:

V: la familia vive en la comunidad

R: la familia vive aislada, lejos de la comunidad

N: no vive dentro de la comunidad

H = CONSULTAS: número de consultas curativas realizadas por la madre durante un año



10 mm/s 10 mm/mV 0.05-35



Capítulo V

ESTADÍSTICA ANALÍTICA. INFERENCIA ESTADÍSTICA

1.- PRINCIPIOS DE INFERENCIA ESTADÍSTICA

1.1.- Inferencia estadística y Validación

Cuando se estudia una muestra, no siempre se está satisfecho con los resultados. De hecho, se aprende algo de la muestra; sin embargo, siempre se desea conocer más sobre la población total.

En este contexto, tres niveles pueden ser distinguidos en la extrapolación de la información obtenida por los resultados de la muestra:

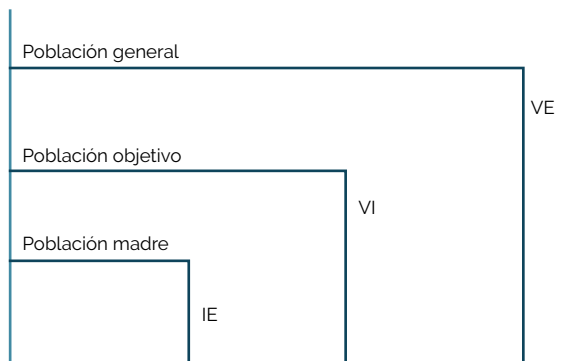
1. Inferencia de la muestra a la población madre = Inferencia Estadística (IE)

Población madre: población en la cual la muestra examinada fue obtenida.

2. Inferencia de la muestra hacia la población objetivo = Validación Interna (VI) Población objetivo: la población a quién uno le gustaría inferir los resultados de la muestra.

3. Inferencia de la muestra a la población general = Validación Externa (VE)

Estos tres niveles pueden ser representados por una figura, tomado del libro de Kleinbaum et al., 1982, Epidemiologic Research: Principles and quantitative methods, p.187.



- ESTUDIOS DE LA POBLACIÓN

La VI y VE fueron tratados en el capítulo de Validez. En este capítulo, resulta importante entender que, la inferencia estadística es solo una parte de todo el procedimiento que tiene que ser integrado por la técnica de bioestadística epidemiológica.

1.2.- Punto de estimación y estimación del intervalo de confianza

Para predecir algo acerca de un parámetro desconocido de la población madre, dos enfoques pueden ser usados:

a.- Un enfoque que responda a las siguientes preguntas:

¿Qué único valor representa el mejor parámetro para una población desconocida?

Un solo valor es suficiente para responder a esta pregunta. Si el problema es, por ejemplo, ver cuál es la diferencia de dos tasas de prevalencia de dos muestras. La respuesta puede ser respondida como sigue:

¿Cuál de las diferentes tasas de prevalencia representa la mejor para la población(es) del que fueron tomadas las dos muestras?

Esta diferencia es dada por un número. En el procedimiento de prueba, nosotros proponemos una figura bajo la hipótesis nula (H_0). Si el resultado de la prueba no es compatible con la figura propuesta bajo H_0 , proponemos la hipótesis alterna (H_a).

Este enfoque es llamado **PUNTO DE ESTIMACIÓN**

b.- Un enfoque que responda la siguiente pregunta: ¿Cuál es el rango de los posibles valores para un parámetro desconocido?

Para responder esta pregunta, se tiene que proponer un grupo de números.

En el contexto de las diferentes tasas de prevalencia, la pregunta sería: ¿Cuáles son los valores más probables para las diferentes tasas de prevalencia de las poblaciones?

Aquí uno propone un grupo de valores y también el grado de confianza que uno tiene en la estimación de la diferencia real.

Este enfoque es llamado: **ESTIMACIÓN DEL INTERVALO.**

1.3.- Estructura de la hipótesis

1.3.1.- ESTRUCTURA LÓGICA

Empezaremos el análisis de la estructura lógica con dos ejemplos:

Ejemplo A:

Una persona está en buen estado de salud o está con un mal estado de salud.

Si la persona está en buen estado de salud, entonces la persona no está en mal estado de salud. Pero, si la persona está



en mal estado de salud, entonces se **rechaza la hipótesis** de que la persona está en buen estado de salud, es decir, intrínsecamente existe un efecto o asociación.

Ejemplo B:

En una corte, una persona acusada es inocente o culpable.

Si el jurado acepta que la persona es inocente, entonces, no es culpable. Pero, si el jurado le declara culpable, entonces, no es inocente.

¿Cómo el jurado da el veredicto de culpable?

El jurado llega a tal veredicto, si hay suficiente evidencia para **rechazar la hipótesis de inocente**.

Si hay poca evidencia el jurado tiene que pesar esta evidencia: si el jurado considera que el volumen de la evidencia no es suficiente para convencerlos, **entonces la hipótesis de no culpable no es rechazada**.

Lo contrario, si el jurado considera que la masa de evidencia es suficientemente grande, entonces, **la hipótesis de no culpable es rechazada**, y la persona es declarada culpable.

Como el sistema es muy conservador a favor de la persona acusada, entonces podemos hacer la diferencia de dos tipos de errores:

a.- No rechaza la hipótesis de inocente, cuando la persona es no culpable.

(En el caso planteado, si la hipótesis de inocencia no se rechaza, significa que no hay pruebas suficientes para demostrar la culpabilidad del acusado. Por lo tanto, el tribunal debe declararlo inocente).

b.- Rechaza la hipótesis de inocente, (declarando a la persona culpable) cuando la persona es en verdad culpable.

(En este caso, la persona es considerada culpable hasta que se demuestre lo contrario. El acusado podría apelar la sentencia y presentar nuevas pruebas que respalden su inocencia, pero si el tribunal encuentra que las pruebas son suficientes para demostrar su culpabilidad, la sentencia se mantendrá).

La razón para discutir los detalles de los procedimientos de la corte con la lógica de las pruebas estadísticas es muy similar:

Un investigador sospecha la existencia de un fenómeno X en una población (por ejemplo: la existencia de una tasa de prevalencia diferente entre las personas de género masculino y femenino en una población dada).

Formula una hipótesis a probar y procede de la siguiente manera:

Toma una muestra simple de la población y determina la tasa de prevalencia en los géneros masculino y femenino. Pero el investigador no puede probar su hipótesis (de una diferencia) por dos razones:

a.- Matemática:

- Supongamos que la diferencia es $= 0,02$
- Aun cuando la diferencia observada sea de $0,02000001$ se rechazaría su hipótesis.

b.- Distribución simple:

Hay algunas posibles muestras, la diferencia de las tasas de prevalencia observada es una entre algunas muestras posibles.

El investigador plantea el problema de una forma diferente. Formula la hipótesis nula (H_0): la hipótesis bajo la cual no hay diferencia entre ambas poblaciones, o no hay asociación entre las dos variables.

La investigación no puede probar H_0 , entonces el investigador puede únicamente falsificar H_0 .

H_0 corresponde a la hipótesis de "no culpable". Si no hay suficiente evidencia para rechazar H_0 , entonces se concluye que la diferencia es no significativa.

Si, por el contrario, hay suficiente evidencia para rechazar H_0 , entonces el investigador rechaza H_0 y acepta la hipótesis alterna H_a , concluyendo que la diferencia es significativa

El investigador puede cometer dos tipos de errores:

Alfa error: el error de rechazar H_0 y aceptar H_a , cuando H_0 es verdadero (= declara a alguien culpable cuando es inocente)

Beta error: el error de no rechazar H_0 , cuando H_a es verdadero (= declara a alguien inocente, cuando es culpable).

1.3.2.- DISTRIBUCIÓN SIMPLE EN FORMA TEÓRICA

¿Cómo se determina el punto desde el cual, se piensa que se tiene suficientes argumentos para rechazar H_0 ?

Para entender la lógica de la determinación de este punto, tenemos que revisar las características de la distribución normal. Cuando nosotros tomamos toda la muestra simple del tamaño n de una población dada, del tamaño de la población madre N , y tomamos los parámetros de descripción de tendencia central como la media (\bar{X}), entonces obtendríamos una distribución de esas medias, con las siguientes tres características:

1. La distribución es normal (cuando la población n dada es lo suficientemente grande), aun cuando la distribución del total de la población no tiene una distribución normal.

2. La media de esta distribución es igual a la media de la población madre
3. La desviación estándar de esta distribución de la media es igual:

$$= \frac{\sigma}{\sqrt{n}}$$

Llamado el error estándar de la media.

Entonces, si uno conoce n (tamaño de la muestra), μ (la media de la población) y (error σ estándar de la media), podemos determinar la distribución de la media de todas las muestras simples y así también la localización de la media en la muestra simple.

En la práctica:

- a. Nosotros no conocemos la media de la población μ , pero proponemos un valor para este (μ_0) bajo las condiciones de H_0 (hipótesis nula).
- b. Nosotros raramente la conocemos σ y reemplazamos esta por la desviación estándar de la muestra.

En conclusión:

Para la inferencia del punto de estimación necesitamos:

1. El tamaño de la muestra, n .
2. Formular las hipótesis: H_0 y H_a

3. Proponer un valor para la media de la población, μ , bajo H_0 .
4. Seleccionar el grado de confianza en el rechazo de H_0 . Se considera, error cuando rechazamos H_0 que es igual a alfa error ($H_0 = \text{alfa-error}$). Convencionalmente se acepta un error de 0,05 o de 0,01. Los grados de confianza es $(1 - \text{alfa})$ entonces $1 - 0,05 = 0,95$ o $1 - 0,01 = 0,99$.
5. n tiene que ser lo suficientemente grande, caso contrario, la distribución normal de la media no estaría garantizada.
6. La media tiene que ser una buena descripción estadística de la muestra. Así, para usar la media como un parámetro de inferencia, la muestra tiene que ser normal, si no es así; tendríamos que usar otro parámetro de inferencia. (ejemplo: la mediana).

1.4.- INTERVALO DE CONFIANZA

En lugar de formular H_0 y H_a , nosotros podríamos tratar de predecir donde se ubica el parámetro de la población desconocida.

En la práctica se determinan los posibles valores del rango de la muestra estadística.

$$\text{———} \left(\text{———} \overset{\text{I}}{\text{———}} \text{———} \right) \text{———}$$

$$\underline{\hspace{10em}}$$

$$\bar{X}$$



El rango es llamado **INTERVALO DE CONFIANZA**.

Los límites de este intervalo de confianza son llamados **LÍMITES DE CONFIANZA**.

EL GRADO DE CONFIANZA = la oportunidad de encontrar los parámetros en el intervalo = $(1 - \alpha) \times 100 \%$

La longitud del intervalo de confianza depende de n , σ y α

- a. Si n tiene un valor alto, entonces la longitud será baja
- b. Si σ tiene un valor alto, entonces la longitud será alta
- c. Si α tiene un valor alto, entonces la longitud será alta

INTERPRETACIÓN DEL INTERVALO DE CONFIANZA DE 95 %

Uno está 95% seguro que el parámetro de la población se encuentra dentro de este intervalo de confianza recomendado.

Nota: Resulta un error decir que el 95 % de las observaciones están dentro de los límites de confianza.

Los límites de confianza son bilaterales: - está localizado a los lados de la estadística descriptiva de la muestra, por lo tanto, hay un límite de confianza bajo y un límite de confianza alto.

Es de común uso, indicar las estadísticas y los límites de confianza de la siguiente forma:

- Parámetro estimado de la muestra
- $\hat{\theta}$ Límite de confianza superior
- $\underline{\theta}$ Límite de confianza menor

Apliquemos esto con un ejemplo de la tasa de diferencia:

Supongamos \hat{TD} (media de las medias): 0,30
 Y el 95 % de límite de confianza es $(\underline{0,20}; \overline{0,40})$.

Entonces, la tasa de diferencia en la muestra es = 30 % y estamos 95 % seguros que la tasa de diferencia real cae entre el 20 % y el 40 %.

Hay una similitud entre el punto de estimación y el intervalo de estimación:

En efecto, si ponemos el 95 % de intervalo de confianza corresponde probar si:

$H_0: TD = 0;$

$H_a: TD \neq 0.$

Entonces, 0 está incluido en el intervalo de confianza en donde H_0 no es rechazado; en el ejemplo dado 0 no está incluido en el intervalo $(\underline{0,20}; \overline{0,40})$, por lo que H_0 es rechazado y H_a es aceptado.

La estimación del intervalo, tiene una ventaja sobre el punto de estimación: la estimación del intervalo demuestra donde

el parámetro real cae. En el ejemplo dado, nosotros rechazamos H_0 y aceptamos H_a por lo que nosotros estamos 95 % seguros que la tasa de diferencia real es al menos del 20 % y el mayor 40 %.

1. LÍMITE DE CONFIANZA UNILATERAL

Es posible que únicamente se esté interesado en un límite de confianza, esto corresponde a una prueba unilateral de la hipótesis. En el ejemplo anterior:

H_0 : $TD = 0$

H_a : $TD > 0$

Consideremos el siguiente ejemplo:

En un tratamiento experimental de cáncer, la principal pregunta es: si este tratamiento es mejor que el clásico tratamiento.

Una hipótesis bilateral debería ser: El tratamiento experimental es significativamente mejor o peor que el tratamiento clásico.

Con el clásico tratamiento de cáncer, los pacientes tienen un 30 % de sobrevida de 5 años.

Con el tratamiento experimental, nosotros planteamos lo siguiente:

n : 100 pacientes

Resultado: 40 pacientes con sobrevida luego de 5 años

Pregunta: ¿El tratamiento experimental es mejor que el tratamiento clásico?

Para $\alpha = 0,05$

$Z_{1-\alpha} = 1,645$ (ver la tabla Z unilateral)

Construyamos entonces el límite menor de confianza alrededor de la muestra:

$$\Pr [p > .40 - 1,645 \times \sqrt{\frac{0,4 \times 0,6}{100}}] = 0,95$$

$$\Pr [p > .40 - 1,645 \times 0,049] = 0,95$$

$$\Pr [p > .319] = 0,95$$

Dado que 31,9 es más alto que la línea de base del 30 %, nosotros podemos concluir (y estamos 95% seguros) que el tratamiento experimental es mejor que el clásico tratamiento y que los cinco años de sobrevida es al menos 1,9 % más alto.

2. ¿CÓMO ENFOCAR UN ANÁLISIS ESTADÍSTICO?

Para entender mejor los diferentes pasos del punto o intervalo de estimación proponemos el siguiente esquema:

PASO 1: Valorar el grupo de datos

- a. El número de datos y su eventual independencia
- b. La escala de medida de cada variable

- c. La forma en que los datos están disponibles: fila, datos, transformados, agrupados, etc.
- d. ¿Son los datos independientes, iguales o duplicados?

PASO 2: Especificar la propiedad pertinente de la distribución de las variables en el grupo de datos: univariante, bivariante, multivariante.

PASO 3: Determinar el asunto de la inferencia

La inferencia sería un punto de estimación (una prueba hipotética) o un intervalo de estimación (cálculo de los límites de confianza alrededor de la muestra estadística).

PASO 4: PUNTO DE ESTIMACIÓN

- a. Describa la hipótesis que ha sido probada y específica H_0 y H_a (=describa H_0 y H_a en palabras) Si uno conoce algo acerca del problema donde la hipótesis tiene que ser unilateral.
- b. Seleccione el parámetro de inferencia en función de la escala de medida, la normalidad de la distribución de la muestra, el diseño de estudio, etc.
- c. Especifique H_0 y H_a en términos de parámetros de inferencia.
- d. Decida sobre el alfa – error
- e. Ejecute el apropiado test estadístico.
- f. Interprete los resultados del test.

PASO 5: ESTIMACIÓN DEL INTERVALO

- a. Seleccione el parámetro de inferencia = igual al paso 4b
- b. Decida sobre los límites de confianza, unilateral o bilateral, en función de priorizar la información y del problema básico de la hipótesis.
- c. Decida sobre el alfa error y el intervalo de confianza (1- alfa error)
- d. Calcule los límites de confianza
- e. Interprete los límites de confianza en función de los problemas básicos de la hipótesis.

3. PUNTO DE ESTIMACIÓN = TEST ESTADÍSTICOS

3.1.- Para variables con escala numérica

Explicaremos esto con la ayuda del ejemplo de la frecuencia del pulso.

Como se recuerda una muestra simple del tamaño de $n = 10$, da el siguiente resultado:

59, 72, 58, 65, 77, 83, 72, 77, 62, 62

Aplicaremos los pasos de la guía analítica en este ejemplo:



Paso 1:

- a. la variable
- b. Escala numérica
- c. Datos en columna
- d. Los datos son independientes

Paso 2: es una distribución univariante

Paso 3: el problema de la inferencia es el punto de estimación

Paso 4:

- a. La muestra es una muestra simple obtenida de una población con características conocidas. Aquí la hipótesis puede ser unilateral.
- b. Los parámetros de inferencia es el parámetro de tendencia central. Usaremos la media debido a que los datos se ven distribuidos normalmente.
- c. $H_0: u = u_0$, $H_a = u > u_0$
 u = la media de la población de la cual la media fue tomada
 u_0 = la media de una población conocida con un valor de 60 pulsaciones/minuto
- d. Alfa error = 0,05
- e. t-test

$$t_{n-1} = \frac{\bar{x} - \mu_0}{\sqrt{\frac{DS}{n-1}}}$$

$$\bar{x} = 68,7$$

$$\mu_0 = 60,0$$

$$DS = 8,67$$

$$n = 10$$

$$t_9 = \frac{68,7 - 60,0}{\sqrt{\frac{8,67}{9}}}$$

$$= 3,01$$

Interpretación:

El resultado obtenido de la prueba de 3,01 es más alto que el valor crítico de 1,833 para un alfa error de 0,05 (una hipótesis unilateral y 9 grados de libertad)

En efecto, este resultado corresponde con un valor hipotético de p:

$$0,02 < p < 0,01$$

La oportunidad de encontrar un resultado idéntico como el obtenido o un resultado más extremo dado por la hipótesis nula, es verdadero; en efecto, el resultado es menor que el 2% pero más alto que el 1%

De todas formas, esta oportunidad es menor que la propuesta de alfa error, por lo tanto, se rechaza H_0 y se acepta H_a .

Esta muestra no viene de una población con una media de 60, pero viene de una población con una media mayor que 60 pulsaciones/ minuto.

NOTA:

- a. La media tiene que ser muy bien descrita en la distribución de la muestra, por lo que siempre tenemos que examinar si la distribución de la muestra es compatible con una distribución normal.
- b. Si uno prueba dos muestras independientes, el t-test, asume ambas varianzas como iguales, por lo que, las igualdades de las varianzas de las muestras tienen que ser comprobadas.

3.1.2.- DISTRIBUCIÓN BIVARIANTE

En la distribución bivalente hay diferentes muestras posibles, basadas en, por ejemplo:

- Correlación del coeficiente lineal
- Regresión lineal

3.2.- DATOS ORDINALES/NOMINALES

Estudiaremos primero la escala nominal dicotómica y la escala de medidas ordinal con dos niveles únicamente.

Ejemplo:

En un subcentro de salud, parece que la cobertura de vacunación es muy baja.

Un estudio de este problema está siendo llevado a cabo, tomando una muestra representativa de 200 niños de 2 a 4 años de edad. 60 de ellos tenían el esquema de vacunación completa (=estudio A).

Este resultado es considerado insuficiente y una intensiva campaña de vacunación ha sido instaurada.

Un año después, los resultados han sido evaluados por medio de un estudio, tomando una muestra representativa de 100 niños de 2 a 4 años, 50 fueron encontrados con el esquema completo de vacunación (=estudio B)

Pregunta: ¿la cobertura de vacunación mejoró en la población?

3.2.1.- Z-TEST

Para analizar los datos seguiremos el siguiente esquema analítico:

PASO 1:

- a. Hay dos variables: (grupo de estudio, esquema de vacunación), siendo en este caso el esquema de vacunación la variable dependiente

- b. Son datos nominales
- c. Los datos son independientes (esto significa que los niños de la muestra B pueden ser niños de la muestra A u otros, pero en el estudio B no tuvieron un niño vacunado del estudio A debido a que el seguimiento fue realizado sistemáticamente un año más tarde).

PASO 2: Es un estudio bivalente

Puede ser representado en una tabla de 2x2

ESTUDIO	ESQUEMA DE VACUNACIÓN		TOTAL
	+	-	
A	60	140	200
B	50	50	100
TOTAL	110	190	300

PASO 3: el problema de estudio es el punto de estimación

PASO 4:

- a. La hipótesis es que el programa tiene un impacto positivo sobre el esquema de vacunación.
 - H_0 : el esquema de vacunación de la población A es igual que de la población B (o las dos muestras simples son tomadas de dos poblaciones con una cobertura común).

- Ha: El esquema de vacunación de la población B es mejor que de la población A (o las dos muestras simples son tomadas de dos poblaciones del cual la cobertura de vacunación de la población B es mejor que de la población A)
- b. El parámetro de mayor interés es la proporción de niños vacunados
- c. Ho
Ha
- d. Error = 0,05
El radio crítico para 0,05 = +1,645
- e. La aplicación del Z-test es el más apropiado método de análisis:

$$\text{Fórmula Z} = \frac{(I_b - p_a) - (\widetilde{I}_{ribú} - II_{ia})}{p_c - q_c \left(\frac{1}{n_a} + \frac{1}{n_b} \right)}$$

p_c = la proporción común, es calculada de la siguiente manera:

$$p_c = \frac{n_a p_a + n_b p_b}{n_a + n_b}$$

$$= \frac{60 + 50}{200 + 100} = \frac{110}{300}$$

$$q_c = 1 - p_c$$

$$Z = \frac{(0,50 - 0,30) - 0}{\sqrt{\frac{110}{300} \times \frac{190}{300} \times \left(\frac{1}{200} + \frac{1}{100} \right)}}$$

$$Z = \frac{0,20}{\sqrt{0,232 \times 0,015}} .$$

$$Z = \frac{0,20}{0,059} = 3,389$$

f.- Interpretación:

La oportunidad de encontrar un resultado de 3,389 (o un valor más extremo todavía), dado que H_0 sería verdadero, es menos que $1/1000$. Consecuentemente rechazamos H_0 y aceptamos H_a . Hay entonces un significativo incremento de la cobertura de vacunación.

3.2.2 ALTERNATIVA DEL Z-TEST: PRUEBA CHI CUADRADO DE PEARSON χ^2

Principios del test:

Se organiza a través de una sola figura de las diferencias entre los valores observados y los valores calculados bajo el H_0 de la variable independiente.

Bajo H_0 los valores esperados de cada celda son calculados como sigue:

$$\frac{\text{Total, de columnas por total de filas}}{\text{Total, general}}$$

Esta determinación racional es de la siguiente forma:

$$p(AB) = p(A) \cdot p(B)$$

Para el esquema de vacunación de los niños en el estudio A:

$$\begin{aligned} & p[(\text{vacunación } +) (\text{estudio A})] \\ &= p(\text{vacunación } +) (\text{estudio A}) \\ &= 110/300 \times 200/300 \end{aligned}$$

Para obtener el número n de niños esperados, tenemos que multiplicar por 300:

$$= 110/300 \times 200/300 \times 300$$

Aplicando al ejemplo, tenemos:

ESTUDIO	ESQUEMA DE VACUNACIÓN		TOTAL
	vacunados +	vacunados-	
A	60	140	200
B	50	50	100
TOTAL	110	190	300

El esperado valor para la celda correspondiente para los niños vacunados en el estudio A:

$$\frac{110 \times 200}{300} = 73.3$$

Los valores esperados para las otras celdas pueden entonces ser calculadas de la misma forma:

ESTUDIO	ESQUEMA DE VACUNACIÓN		TOTAL
	vacunados +	vacunados-	
A	$\frac{200 \times 110}{300} = 73,3$	$\frac{200 \times 190}{300} = 126,7$	200
B	$\frac{100 \times 100}{300} = 36,7$	$\frac{100 \times 190}{300} = 63,3$	100
TOTAL	110	190	300

La prueba de Pearson X^2 , resume en una simple figura la suma de las diferencias entre el valor del parámetro observado y el esperado, con la siguiente fórmula:

$$X_p^2 = \frac{(60 - 73,3)^2}{73,3} + \frac{(140 - 126,7)^2}{12,7} + \frac{(50 - 36,7)^2}{36,7} + \frac{(50 - 63,3)^2}{63,3}$$

$$= 11,42$$

La región crítica de $X_{(1)}^2 = 3,84$ $p < 0.001$

Nota:

1.- Los grados de libertad son calculados por la aplicación de la siguiente fórmula:

$$gl = (r-1)(c-1)$$

r = rango o fila
c = columna

$$= (2-1)(2-1)$$

$$= 1$$

2.- El test X^2 permite probar una hipótesis bilateral:

$$3.- X^2 = Z^2$$

11,42 \approx (3,39) ², las pequeñas diferencias son debidas al recorrido.

4.- No use el X^2 si el valor esperado para una celda es menos de 5.

5.- El X_p^2 es difícil calcular:

Hay una fórmula más simple basada en otros principios:

la PRUEBA MANTEL-HAENSZEL X^2

3.2.3. TEST DE MANTEL-HAENSZEL X^2

Fórmula:

$$X_{MH}^2 := \frac{(O - E)^2}{\frac{(a+b)(c+d)(a+c)(b+d)}{T^2(T-1)}}$$

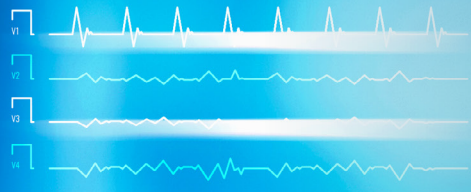
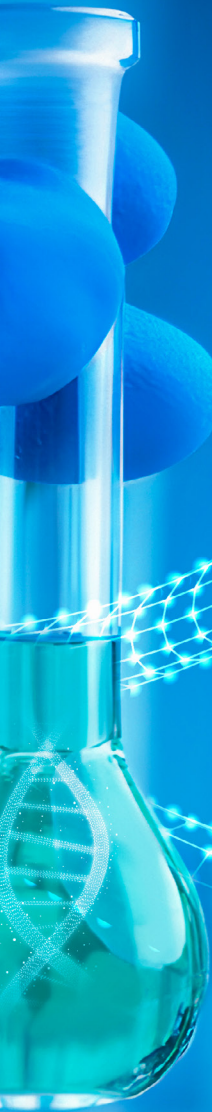
ESTUDIO	VARIABLE A		TOTAL
	+	-	
+	a	c	a+b
-	b	d	b+d
TOTAL	a+b	c+d	T

Aplicado al ejemplo:

$$\begin{aligned}
 & \frac{E = 110 \times 200}{300} = 73,3 \\
 & X_{MH}^2 : \frac{\left(60 - \frac{110 \times 200}{300} \right)^2}{\frac{110 \times 190 \times 200 \times 100}{300 \times 300 \times 299}} \\
 & = 11,45
 \end{aligned}$$

El resultado es el mismo obtenido con el anterior método, con pequeñas diferencias que pueden ocurrir.

X_{MH}^2 Es la fórmula para usar en los programas de calculadora.



10 mm/s 10 mm/mV 0.05-35



Capítulo VI

INTRODUCCIÓN A LA PSICOMETRÍA

1.- FIABILIDAD DE LA ESCALA

La medición precisa de propiedades de objetos es un componente esencial en diversas disciplinas, y en este proceso, los errores de medida son inevitables. Ya sea en mediciones físicas, médicas, químicas o sociales, la presencia de errores es una realidad que debe ser reconocida y gestionada. En el ámbito de la psicometría, la teoría clásica de test aborda esta cuestión mediante el concepto de fiabilidad, que se centra en la precisión de las puntuaciones obtenidas.

La fiabilidad, según De Boeck y Elosua (2016), se define como la medida de la precisión de las puntuaciones y busca estimar la magnitud en la que los errores afectan el proceso de medición. Dado que estos errores son inherentemente aleatorios y no controlados, la estimación de su impacto proporciona una medida de la confiabilidad de las puntuaciones del test.

En la búsqueda de estimar la fiabilidad, se emplean diversos diseños y procedimientos, siendo tres los más destacados: formas paralelas, test-retest y método de las mitades. Las formas paralelas implican la construcción de dos versiones equivalentes del mismo test, cuya aplicación a la misma muestra permite estimar la correlación entre las puntuaciones y, por ende, la fiabilidad. El test-retest, por otro lado, requiere administrar el mismo test en dos ocasiones diferentes a la misma muestra, evaluando la estabilidad temporal de las puntuaciones. Mientras tanto, el método de las mitades, el más simple de todos, estima la fiabilidad a partir de una única administración del test, interpretándose en términos de consistencia interna.

Es crucial reconocer que los valores del coeficiente de fiabilidad oscilan entre 0 y 1, donde valores superiores a 0,8 se consideran aceptables en contextos de investigación, y valores mayores a 0,9 son preferibles cuando las puntuaciones se utilizan como base para inferencias o decisiones individuales (Nunnally, 1978). Este enfoque en la fiabilidad proporciona un marco sólido para comprender y gestionar los

errores de medida en el proceso de evaluación y, por ende, contribuye a mejorar la validez de las inferencias derivadas de las puntuaciones obtenidas.

Coefficientes alfa y omega

Dentro de las diversas metodologías para calcular el coeficiente de fiabilidad a partir de una única aplicación del test, ha ganado prominencia el coeficiente alfa de Cronbach (Cronbach, 1951). Este coeficiente sirve como indicador de la covariación entre los ítems de una escala o la correlación intra-clase, siendo interpretado como un reflejo de la consistencia interna de la escala. Su amplio uso ha suscitado un debate continuo que persiste hasta el día de hoy (Viladrich et al., 2017).

Una alternativa que ofrece Jamovi, es el coeficiente omega de McDonald (1985), cuya estimación se basa en una solución factorial jerárquica implementada sobre un modelo factorial exploratorio unidimensional, según lo descrito por Trizano-Hermosilla y Alvarado (2016) y Revelle (2019). Cuando los datos se ajustan a un modelo unidimensional y las cargas factoriales de los ítems son idénticas, es decir, en situaciones de tau-equivalencia donde no existen errores correlacionados, ambos índices (alfa de Cronbach y omega de McDonald) muestran similitudes.

Una interpretación apropiada de los coeficientes alfa y omega debe considerar que la estimación de estos indicadores no aporta información sobre la estabilidad temporal de la

medida, ya que ambos se centran exclusivamente en la consistencia interna. Además, es importante tener en cuenta que tanto alfa como omega tienen una definición formal como indicadores de consistencia destinados a variables continuas, aunque existen versiones adaptadas para variables ordinales (Gadermann et al., 2019). En la práctica, se podría asumir que las escalas ordinales, como las escalas Likert, se traten como continuas cuando tienen al menos 5 categorías de respuesta y exhiben distribuciones aproximadamente normales (Rhemtulla et al., 2012).

Para obtener los indicadores de fiabilidad de una escala, se deben seguir los siguientes pasos en el programa de Jamovi:

Figura 1. Análisis de fiabilidad (Paso 1).

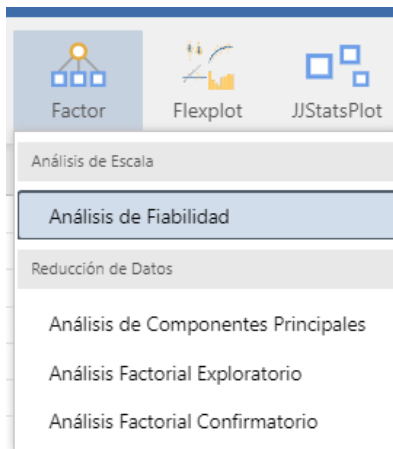


Figura 2. Seleccionar variables (Paso 2). Mover a la ventana Ítems los ítems que constituyen la escala que se desea analizar.

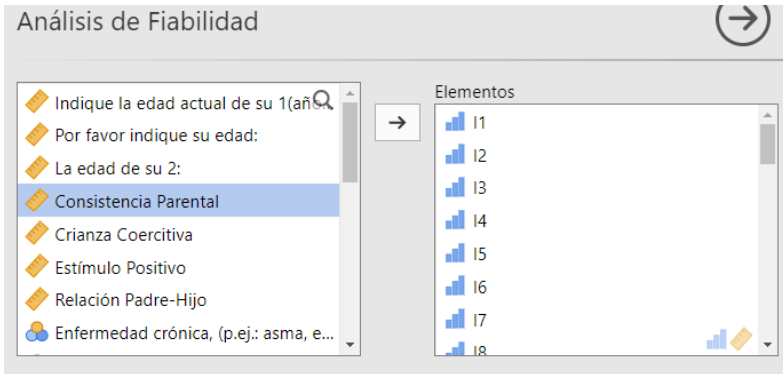


Figura 3. Seleccione el estadístico adecuado (Paso 3).

<input checked="" type="checkbox"/> Alfa de Cronbach	<input checked="" type="checkbox"/> α de Cronbach (si se elimina un elemento)
<input checked="" type="checkbox"/> ω de McDonald	<input checked="" type="checkbox"/> ω de McDonald (si se elimina el elemento)
<input type="checkbox"/> Media	<input checked="" type="checkbox"/> Media
<input type="checkbox"/> Desviación estándar	<input type="checkbox"/> Desviación estándar
Opciones Adicionales	<input type="checkbox"/> Correlación del elemento con otros

Para estimar los coeficientes de fiabilidad por los métodos alfa de Cronbach y omega de McDonald comprobaremos las opciones correspondientes, tanto para la escala global (α de Cronbach y ω de McDonald) como para cada uno de los ítems; α y (α de Cronbach [si se elimina el ítem] y ω de McDonald [si se elimina el ítem]). Véase en la figura 3.

Figura 4. Salida de los resultados en el Jamovi (Paso 4).

Análisis de Fiabilidad

Estadísticas de Fiabilidad de Escala

	Alfa de Cronbach	ω de McDonald
escala	0.827	0.852

[3]

Estadísticas de Fiabilidad de Elemento

	Media	Si se descarta el elemento	
		Alfa de Cronbach	ω de McDonald
11	2.51	0.823	0.849
12	2.51	0.821	0.847
13 *	1.98	0.822	0.849
14 *	1.77	0.826	0.851
15	1.94	0.832	0.855
16	2.04	0.822	0.848

El coeficiente alfa de Cronbach (α de Cronbach) en este contexto es de 0,827 y el valor del coeficiente ω de McDonald es 0,852. Además de estos indicadores globales que se aplican a la totalidad de la escala, resulta valioso evaluar la contribución individual de cada ítem a la confiabilidad global (Figura 4). Ejemplo, ítem 6 de la escala, el coeficiente alfa de Cronbach es 0,822, y el valor de ω es 0,848.



Error típico de medida

El error típico de medida se refiere a la variabilidad esperada en las puntuaciones de los individuos alrededor de su verdadera capacidad o atributo. Es un indicador cuantitativo de la dispersión de las mediciones y aborda la inevitabilidad de las fuentes de error en cualquier proceso de evaluación.

El error típico de medida, en su esencia, representa la variabilidad esperada en las puntuaciones de los individuos alrededor de su verdadero rendimiento. Esta variabilidad, a menudo inevitable debido a diversas fuentes de error, impacta directamente en la fiabilidad del test. La fiabilidad, por otro lado, se refiere a la consistencia y estabilidad de las puntuaciones a lo largo de mediciones repetidas. Es crucial comprender que a medida que la fiabilidad de un test disminuye, el error típico de medida tiende a aumentar, lo que indica una mayor fluctuación en las mediciones.

La relación entre el error típico de medida y la fiabilidad se torna aún más significativa al considerar la estimación del intervalo de confianza. Al realizar mediciones, no podemos asegurar con certeza absoluta la precisión de una puntuación específica. Aquí es donde entra en juego el error típico de medida. La conexión entre este error y la fiabilidad del test se traduce en la capacidad de establecer un intervalo de confianza alrededor de la puntuación observada. Cuanto menor sea el error típico de medida y mayor la fiabilidad, más estrecho será ese intervalo, proporcionando así una

estimación más precisa y confiable de la verdadera capacidad o atributo del individuo.

En conclusión, el error típico de medida y la fiabilidad del test son elementos inseparables en el proceso de evaluación psicométrica. Comprender la relación intrínseca entre ambos es esencial para contextualizar adecuadamente las puntuaciones obtenidas. El error típico de medida, al aumentar con la disminución de la fiabilidad, subraya la importancia de buscar pruebas y evaluaciones con altos niveles de consistencia y estabilidad.

La aplicación práctica de esta comprensión se evidencia en la estimación del intervalo de confianza, donde el error típico de medida actúa como un indicador clave para determinar la amplitud de la zona de confianza alrededor de la puntuación observada. En última instancia, este conocimiento contribuye a una interpretación más matizada y precisa de las mediciones, respaldando decisiones informadas y proporcionando una base sólida para la toma de decisiones en diversos contextos, desde la investigación académica hasta la evaluación clínica.

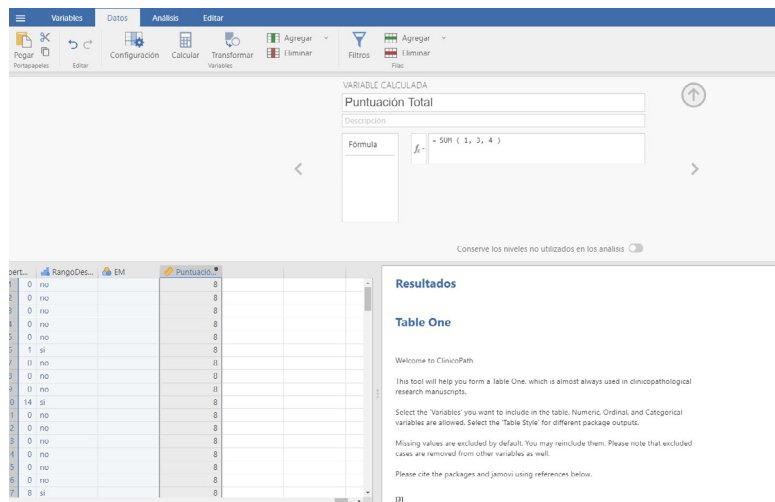
Calcular la puntuación observada

Calcular la puntuación observada en Jamovi implica determinar la suma de las puntuaciones individuales de una persona en cada uno de los ítems que integran la escala. En términos operativos, este procedimiento implica la inclusión de una nueva variable en la hoja de datos para registrar

este valor acumulado, facilitando así la evaluación global del desempeño del individuo en la escala.

Calcular la puntuación total de una persona en Jamovi implica sumar las puntuaciones obtenidas en cada uno de los ítems que conforman la escala. Desde una perspectiva práctica, este proceso implica la creación de una nueva variable en la hoja de datos que refleje la suma de estos valores individuales. Para calcular la puntuación total en Jamovi, primero, abre tu conjunto de datos en la plataforma. Luego, identifica las columnas que representan las puntuaciones individuales de cada ítem en la escala.

Posteriormente, dirígete a la pestaña "Variables" en Jamovi, selecciona "Transformar" y elige "Crear variable". Asigna un nombre a la nueva variable, como "Puntuación Total". En la ventana de creación de la nueva variable, selecciona la función de suma e ingresa las columnas que contienen las puntuaciones de los ítems, por ejemplo, $A + B + C$ si los ítems están en las columnas A, B y C. Después de ingresar la fórmula, haz clic en "Aceptar" para crear la nueva variable y asegúrate de guardar los cambios en tu conjunto de datos. Posteriormente, podrás verificar la puntuación total para cada persona, observando la nueva variable "Puntuación Total" (Figura 5), que reflejará la suma de las puntuaciones individuales en los ítems de la escala.

Figura 5. Creación de la Variable "Puntuación Total" en Jamovi.

Fuente: Los autores

Este proceso simplificado en Jamovi permite una fácil obtención de la puntuación total, brindando una herramienta práctica para analizar y entender mejor los resultados de la escala en el contexto de tu estudio o investigación.

Intervalo de confianza en torno a la puntuación verdadera

Una vez que se tiene información sobre el valor del error típico de medida, es posible calcular, con un nivel de confianza específico, el intervalo de confianza para la puntuación real de los individuos a partir de cualquier puntuación observada.

El error típico de medida, al representar la variabilidad esperada en las mediciones individuales, se convierte en un indicador clave para evaluar la consistencia y fiabilidad de un test. No obstante, su importancia trasciende la simple evaluación de la precisión, ya que su conocimiento abre la puerta a una estimación más precisa y confiable del intervalo de confianza para la puntuación real de los sujetos.

La relación fundamental entre el error típico de medida y la estimación del intervalo de confianza radica en la capacidad de cuantificar la incertidumbre inherente a las mediciones. Conocer el valor de este error permite establecer con un grado de confiabilidad determinado el rango dentro del cual se espera que la puntuación verdadera de un individuo se sitúe. Este enfoque es esencial en la interpretación de las mediciones, ya que reconoce y aborda la variabilidad natural en el desempeño individual.

La estimación del intervalo de confianza se convierte así en una herramienta valiosa para los profesionales en psicometría y evaluación. Al entender que las puntuaciones observadas están sujetas a variaciones normales y esperadas, se fomenta una interpretación más robusta y equilibrada de los resultados. Este enfoque no solo ofrece una medida de la precisión de las puntuaciones, sino que también proporciona una salvaguarda contra interpretaciones excesivamente rígidas o sesgadas.

En conclusión, el conocimiento del error típico de medida despierta una perspicacia fundamental en la evaluación

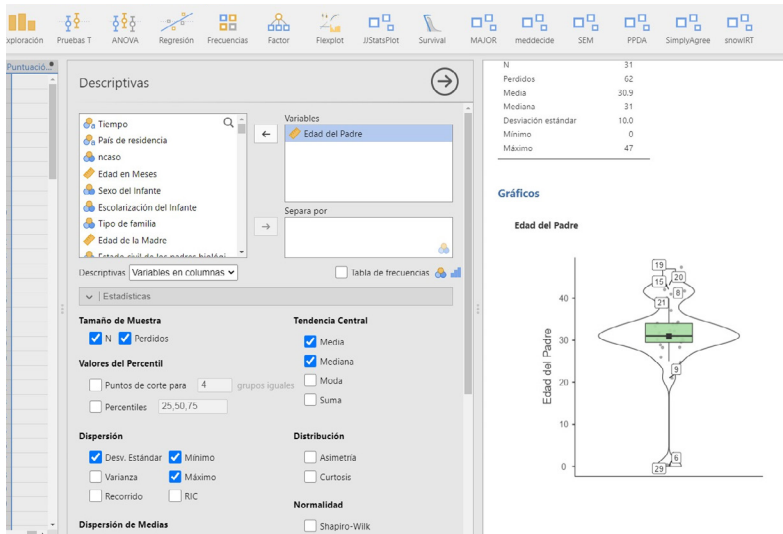
psicométrica. Su papel en la estimación del intervalo de confianza ofrece una metodología sólida para cuantificar la incertidumbre y promover una interpretación más informada y contextualizada de las mediciones. Al reconocer la variabilidad inherente a la medición psicométrica, se establece un fundamento robusto para decisiones informadas y una comprensión más completa de las capacidades y características individuales. En última instancia, este enfoque respalda la integridad y validez de las evaluaciones psicométricas en diversos campos de aplicación.

En Jamovi, el proceso para calcular un intervalo de confianza en torno a la puntuación verdadera generalmente se realiza mediante análisis estadísticos específicos, como regresiones o análisis de medidas repetidas.

Para calcular un intervalo de confianza en Jamovi, primero abre tu conjunto de datos y dirígete a la pestaña "Analizar". Selecciona el tipo de análisis adecuado según la naturaleza de tus datos, como "Regresión" si es pertinente. Configura las variables relevantes, incluyendo la variable dependiente y las variables independientes en la ventana del análisis, y ajustes opciones avanzadas según sea necesario. Posteriormente, ejecuta el análisis y examina la salida de resultados, centrar en la sección relacionada con los intervalos de confianza. Identifica los límites inferior y superior dentro de los resultados, los cuales indicarán la variabilidad alrededor de la puntuación verdadera con cierto nivel de certeza (Figura 6). Finalmente, interpreta los resultados, prestando espe-

cial atención a estos límites, y exporta la información si es requerido para análisis adicionales o presentación.

Figura 6. Análisis de variables e intervalos de confianza en Jamovi



Fuente: Los autores

2.- DIMENSIONALIDAD

Análisis factorial exploratorio

El examen de la validez a través del análisis de la estructura interna busca determinar en qué medida las interacciones entre los elementos y los componentes del examen reflejan con precisión el constructo que se pretende medir, y

que servirá de base para las interpretaciones. Siguiendo los estándares actuales de la APA, se enfoca en explorar la dimensionalidad del test y en analizar el funcionamiento diferencial de los ítems. El análisis de la dimensionalidad aborda principalmente dos aspectos: primero, la confirmación o refutación de la hipótesis de unidimensionalidad; segundo, la evaluación de la estructura multidimensional de los datos.

Estos objetivos implican la aplicación de una metodología destinada a determinar cuántos factores explican la matriz de correlaciones o influyen en la respuesta de un sujeto. En términos generales, el objetivo es definir y demostrar la existencia de una "estructura simple" o "simplicidad factorial", caracterizada por la agrupación de ítems en conjuntos dimensionalmente homogéneos entre sí, pero dimensionalmente distintos entre grupos. El análisis factorial, una técnica multivariada, se emplea para resumir información de un conjunto de variables observadas en un número reducido de variables hipotéticas llamadas factores.

Este procedimiento puede utilizarse de manera exploratoria para describir la estructura interna de los datos o de manera confirmatoria para contrastar hipótesis relacionadas con la dimensionalidad. La tendencia actual considera que el estudio de la dimensionalidad puede adoptar enfoques exploratorios o confirmatorios, y se ha desarrollado una tercera opción conocida como ESEM (Exploratory Structural Equation Modeling) (Ferrando y Lorenzo-Seva, 2014; Morin et al., 2013).

Para llevar a cabo un Análisis Factorial Exploratorio (AFE) en Jamovi, primero, abre tu conjunto de datos y navega hasta la pestaña “Analizar”. Luego, selecciona “Análisis Factorial” de las opciones disponibles. En la ventana del análisis, elige las variables que deseas incluir en el AFE. Ajusta las configuraciones según tus necesidades, como el método de extracción de factores, la matriz de correlación a utilizar y la rotación de factores. Haz clic en “Aceptar” para ejecutar el análisis. Analiza los resultados, prestando especial atención a las cargas factoriales, la varianza explicada y los gráficos de escalamiento multidimensional (Figura 7).

Figura 7. Análisis Factorial Exploratorio (AFE).

Análisis Factorial Exploratorio

Cargas de los Factores

	Factor		Unicidad
	1	2	
28. ¿Cuánto tarda en dormirse?	0.531	0.908	0.792
35. Se va a la cama de mal humor	0.756	0.556	0.686
48. Tiene dificultad para conciliar el sueño por la noche	0.432	0.686	0.661
58. Pesece pronto o me quedo cuando se queda dormido	0.321	0.860	0.533
68. Se levanta o agita partes del cuerpo al dormirse	0.719	0.671	0.728
78. Realiza acciones repetitivas tales como "bucear" al dormirse	0.461	0.585	0.585
88. Sueña excesivamente al dormirse	0.461	0.585	0.615
100. Se despierta más de 2 veces cada noche	0.461	0.585	0.615
110. Después de despertarse por la noche tiene dificultades para dormirse	0.461	0.585	0.615
120. Tiene troncos o sacudidas de las piernas mientras duerme, cambia a menudo de posición o da "patadas" a la ropa de cama	0.677	0.615	0.615

Comprobaciones de Supuestos

Prueba de Esfericidad de Bartlett

χ^2	gl	p
230	55	0.0000+0.0000

Fuente: Los autores

Utiliza esta información para interpretar la estructura subyacente de las variables y la presencia de posibles factores latentes. Además, ten en cuenta las estadísticas descriptivas

y la interpretación gráfica proporcionadas por Jamovi para respaldar tu análisis e interpretación del Análisis Factorial Exploratorio. El procedimiento de análisis factorial implica varias fases, cuya elección precisa depende de la situación específica. En primer lugar, se deben verificar los supuestos necesarios antes de la estimación del modelo estadístico.

Posteriormente, el análisis debe estar guiado por un entendimiento profundo de los datos, permitiendo la formulación de hipótesis sobre su estructura interna. La etapa de extracción factorial se determina mediante la elección de un método adecuado, ajustándose a las características particulares de los datos en cuestión. En muchas ocasiones, la solución inicial puede ser difícil de interpretar, lo que lleva a la aplicación de técnicas de rotación factorial para buscar una solución formalmente equivalente pero más clara desde una perspectiva sustantiva.

La adecuación de la solución depende en parte de cuánto se ajusten los datos a los supuestos del modelo factorial; a medida que aumenta la discrepancia entre estos, la solución obtenida se debilita. Finalmente, la interpretación de la solución factorial debe considerar tanto aspectos formales como sustantivos, y las decisiones sobre el número de factores a retener, su interpretación psicológica y la adecuación del modelo no deben basarse exclusivamente en criterios numéricos, sino también en consideraciones teóricas.

Supuestos

Los supuestos en el análisis factorial constituyen la piedra angular para la validez de las inferencias derivadas de este método estadístico (Forsythe et al., 1986; MacCallum et al., 2007; Muñiz, 2018). El primer paso crítico implica una evaluación exhaustiva para garantizar que los datos cumplan con los requisitos fundamentales del modelo. Antes de aventurarse en la complejidad del análisis factorial, se debe confirmar que la matriz de datos exhiba propiedades esenciales, como la adecuada correlación entre las variables y la suficiente adecuación para la factorización (Alavi et al., 2020). Este paso precautorio no solo respalda la validez del análisis, sino que también establece las bases para la interpretación significativa de los resultados subsiguientes. La comprensión y el respeto de estos supuestos se erigen como cimientos esenciales para desentrañar la estructura subyacente de los datos y garantizar la solidez de las conclusiones extraídas del análisis factorial.

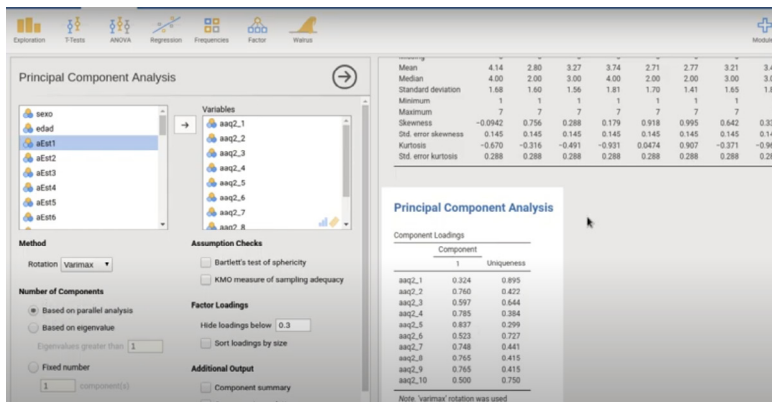
En esencia, se emplean dos medidas para analizar la matriz de correlaciones entre ítems y determinar si es apropiado realizar un análisis factorial. La prueba de esfericidad de Bartlett: Evalúa la hipótesis nula de que la matriz de correlaciones es una matriz identidad. Si la prueba no rechaza esta hipótesis, se podría poner en duda la adecuación de llevar a cabo un análisis factorial (Shrestha, 2021). Además, la medida KMO (Kaiser-Meyer-Olkin) de adecuación de muestreo: Este indicador compara los coeficientes de correlación

parcial con los coeficientes de correlación observados. Si los valores son bajos, no se debe realizar el análisis factorial (Effendi et al., 2019). En contextos de investigación aplicada, se consideran aceptables los valores por encima de 0,70.

Rotación factorial

Después de la extracción de los factores, es posible realizar rotaciones para acercar la solución a una estructura más simple (Meredith, 1964; Ricolfi y Testa, 2021). Jamovi ofrece diversos métodos de rotación, ya sean ortogonales y oblicuos. El primer conjunto genera factores independientes, mientras que el segundo presupone la existencia de factores correlacionados. Varimax (Kaiser, 1958) y Quartimax (Wrigley et al., 1958) son ejemplos de rotaciones ortogonales, mientras que las demás son consideradas rotaciones oblicuas.

Figura 8. (Varimax). Método de Rotación ortogonal.



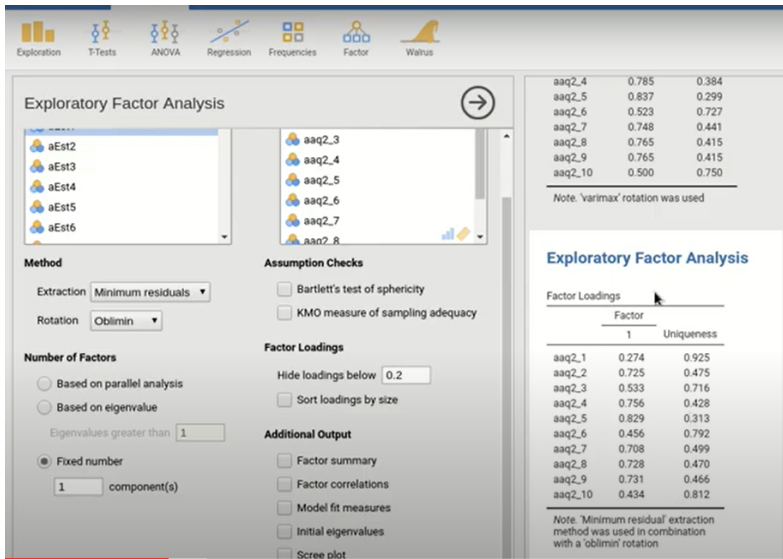
Fuente: Los autores



La rotación factorial emerge como una fase crucial en el análisis de factores, una ventana a la comprensión más profunda de la complejidad de los datos psicométricos. Este proceso, que sigue a la extracción de factores, busca simplificar la estructura subyacente, haciéndola más interpretable y reveladora.

En el vasto paisaje de herramientas estadísticas, la rotación ofrece una paleta de opciones en Jamovi, cada una con su propio matiz y propósito. La bifurcación entre rotaciones ortogonales y oblicuas revela la naturaleza de la relación entre los factores extraídos. En la danza de Varimax y Quartimax, experimentamos la elegancia de rotaciones ortogonales, generando factores independientes que simplifican la interpretación. Mientras tanto, las rotaciones oblicuas, como Promax y Oblimin, introducen la noción de factores correlacionados, capturando las complejidades intrínsecas de la estructura subyacente.

Figura 9. (Oblimin). Método de Rotación oblicua.



Fuente: Los autores

La elección entre estas técnicas no es meramente técnica; es una decisión estratégica que influye en la claridad y la profundidad de las percepciones extraídas. Varimax, con su énfasis en la independencia entre factores, es como destilar la esencia, separando claramente los roles de cada factor. Por otro lado, Promax, al permitir la correlación entre factores, se asemeja más a un diálogo fluido, donde las dimensiones interactúan y se influyen mutuamente.

La aplicación pragmática de estas rotaciones implica no sólo una comprensión matemática sino también una apreciación de la naturaleza del fenómeno subyacente. ¿Buscamos

simplicidad estructural o estamos dispuestos a explorar las interconexiones entre dimensiones? Este dilema refleja la dualidad inherente al análisis factorial, entre la necesidad de claridad conceptual y la aceptación de la complejidad psicológica.

El papel de Jamovi en este escenario no puede ser subestimado. Como facilitador de estas maniobras estadísticas, ofrece una interfaz intuitiva que democratiza el acceso a estas herramientas poderosas. La rotación factorial en Jamovi no es simplemente un proceso técnico; es un viaje hacia la revelación de patrones ocultos y la destilación de la esencia de la estructura subyacente.

En conclusión, la rotación factorial se revela como una danza matemática que desentraña la complejidad de los datos, una herramienta indispensable para los psicólogos y científicos sociales que buscan descifrar los misterios de la mente. A través de la elección sabia de técnicas de rotación y la comprensión de sus implicaciones, se puede desvelar la riqueza subyacente en la matriz de correlaciones, guiando a los investigadores hacia interpretaciones más enriquecedoras y perspicaces.

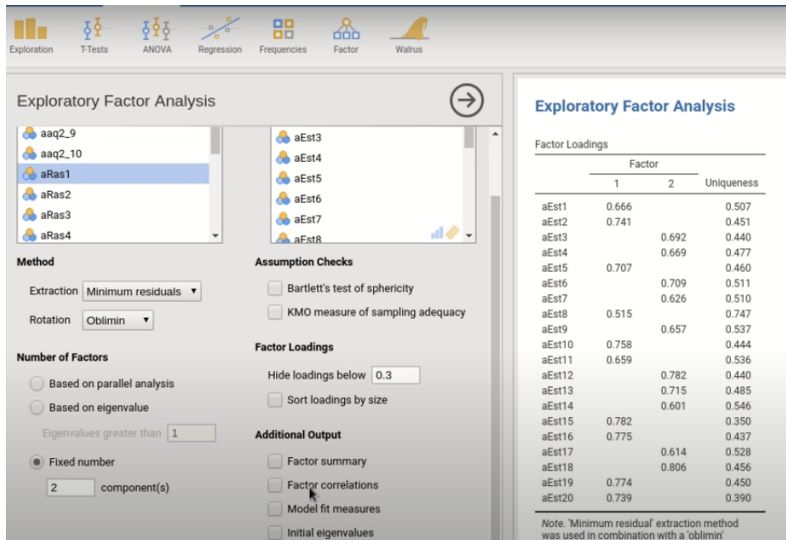
Formato de salida de resultados en Jamovi.

Aparte de detallar las particularidades del proceso de extracción y rotación, Jamovi presenta la capacidad de personalizar el formato de los resultados mediante las siguientes opciones. La opción "Factor Loadings" posibilita la definición del

aspecto de la matriz factorial. En caso de no querer mostrar las saturaciones inferiores a un valor específico, es necesario indicarlo en el campo "Hide loadings below". Asimismo, si se prefiere que las saturaciones aparezcan organizadas, se seleccionará la casilla "Sort loadings by size".

En el contexto del análisis factorial en Jamovi, se proporciona información adicional a través de diversas salidas esenciales. El apartado de "Factor summary" brinda detalles reveladores sobre los factores extraídos, ofreciendo una visión comprensiva de sus características y contribuciones. La sección de "Factor correlations" presenta las correlaciones entre los factores, arrojando luz sobre las posibles interrelaciones y patrones emergentes.

Los "Model fit measures" ofrecen índices que evalúan la adecuación global del modelo a los datos observados, proporcionando una herramienta crítica para la evaluación de la calidad del ajuste. Los "Initial eigenvalues" destacan los valores propios iniciales, ofreciendo perspectivas sobre la varianza explicada por cada factor. Por último, el "Scree plot" representa gráficamente los valores propios en un gráfico de sedimentación de Catell, facilitando la identificación visual de los factores significativos. Estas salidas complementarias enriquecen la interpretación del análisis factorial, brindando una panorámica detallada que guía la comprensión profunda de la estructura subyacente de los datos.

Figura 10. Salida de resultados en Jamovi.

Fuente: Los autores

Unidimensional y Bidimensional

En el ámbito del análisis factorial, la caracterización como unidimensional se refiere a la capacidad de representar un fenómeno o medida en una sola dimensión (Beierl et al., 2018). En este contexto, una solución unidimensional implica que todas las variables o ítems en un conjunto de datos comparten una única fuente subyacente, simplificando la interpretación al sugerir que todas las variaciones observadas están intrínsecamente conectadas. En contraste, la bidimensionalidad implica que el fenómeno o medida se compone de dos dimensiones distintas (Schweizer et al., 2009).

En el análisis factorial bidimensional, las variables se agrupan en dos factores subyacentes únicos, cada uno representando una dimensión única del fenómeno. La variación en las variables se atribuye a la influencia conjunta de estos dos factores independientes, reconociendo la complejidad intrínseca del constructo en cuestión y capturando sus diferentes aspectos. Este enfoque bidimensional permite una representación más completa y detallada del fenómeno que no se limita a una única fuente subyacente.

Imaginemos que estamos investigando la estructura subyacente de un cuestionario diseñado para evaluar la satisfacción laboral. En nuestro análisis factorial, nos enfrentamos a la decisión crucial entre dos enfoques: uno unidimensional y otro bidimensional. En la solución unidimensional, asumimos que todas las preguntas del cuestionario miden una única dimensión, como la satisfacción general en el entorno laboral. Por otro lado, en la solución bidimensional, postulamos que las preguntas se agrupan en dos dimensiones distintas, como la satisfacción con las relaciones laborales y la satisfacción con las oportunidades de desarrollo profesional.

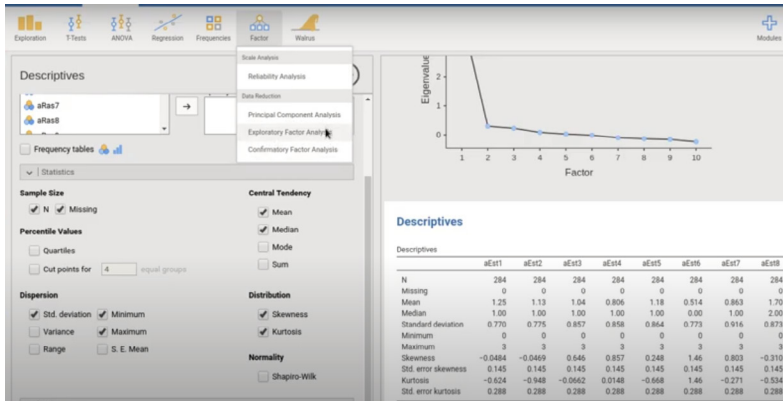
La elección entre estas soluciones no solo afectará la interpretación de los resultados, sino que también influirá en las decisiones prácticas basadas en la evaluación de la satisfacción laboral en un entorno específico. Este ejemplo ilustra cómo el análisis factorial puede ofrecer perspectivas valiosas, pero también destaca la importancia de tomar deci-

siones informadas sobre la estructura subyacente en función de la teoría y los objetivos específicos de la investigación.

Para llevar a cabo un análisis factorial unidimensional en Jamovi, comienza abriendo tu conjunto de datos y accediendo a la pestaña "Analizar". Desde el menú desplegable de análisis, elige "Análisis Factorial" y configura las variables pertinentes en la ventana de análisis, indicando tu interés en una solución unidimensional. Posteriormente, ejecuta el análisis y examina los resultados, prestando especial atención a las cargas factoriales y otros indicadores que respaldan la solución unidimensional. En la interpretación de los resultados, considera que todas las variables se agrupan bajo un solo factor subyacente.

Por otra parte, para realizar un análisis factorial bidimensional en Jamovi, se sigue el mismo procedimiento inicial para abrir datos y seleccionar "Análisis Factorial". En la configuración de variables, especifica tu interés en una solución bidimensional. Posteriormente, ejecuta el análisis y examina los resultados proporcionados por Jamovi, observando las cargas factoriales para ambas dimensiones. En la interpretación de los resultados, considera que las variables se agrupan en dos factores subyacentes distintos, cada uno representando una dimensión única del fenómeno. Estos pasos te permitirán explorar y comprender la estructura subyacente de tus datos tanto en un contexto unidimensional como bidimensional.

Figura 11. Análisis Factorial Exploratorio (AFE) en Jamovi.



Fuente: Los autores

Análisis de contenido

El análisis factorial, una poderosa herramienta estadística, desentraña patrones y relaciones subyacentes en conjuntos de datos, revelando la estructura interna de las variables observadas. Sin embargo, este análisis cuantitativo por sí solo puede carecer de profundidad si no se contextualiza con la sustancia teórica subyacente. Aquí entra en juego el análisis de contenido, un componente esencial para enriquecer la comprensión de los datos y garantizar la validez del análisis factorial (Morin et al., 2020).

El análisis de contenido implica la exploración y evaluación de las variables y sus relaciones a la luz del modelo teórico subyacente. Este enfoque cualitativo permite examinar el sig-

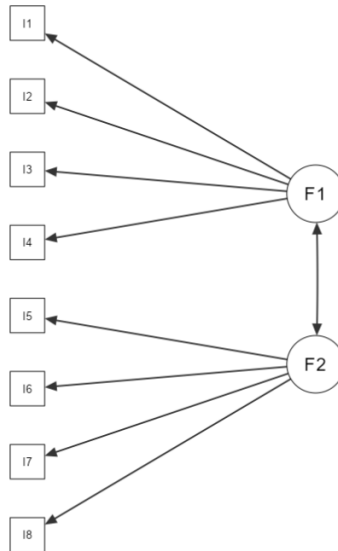
nificado de las variables, sus interconexiones y su alineación con las premisas teóricas. Así, mientras el análisis factorial revela patrones cuantitativos y estructuras subyacentes, el análisis de contenido aporta una dimensión cualitativa, destacando la relevancia teórica de las variables en el contexto del fenómeno investigado.

Para llevar a cabo un análisis de contenido efectivo en conjunto con el análisis factorial, es esencial definir claramente el modelo teórico. Esto implica identificar las relaciones esperadas entre las variables, comprender su significado en el contexto teórico y explorar cualquier desviación significativa de estas expectativas.

En síntesis, el análisis de contenido y el análisis factorial forman un dúo dinámico, donde la rigurosidad estadística del último se equilibra con la profundidad interpretativa del primero. Juntos, no sólo desentrañan la complejidad de los datos, sino que también proporcionan una comprensión holística y enriquecida de los fenómenos estudiados. En la intersección de la cantidad y la calidad, esta combinación metodológica potencia la validez y la riqueza interpretativa de la investigación. Las inferencias derivadas de las soluciones factoriales se respaldan con un análisis teórico, el cual puede enriquecerse mediante métodos confirmatorios de análisis factorial para asistir al investigador en la definición de la dimensionalidad subyacente en los datos. Es imperativo que todas las consideraciones estén fundamentadas en un sólido conocimiento teórico de las variables examinadas; en

ningún caso se debe priorizar la solución estadística sobre la solución teórica.

Figura 12. Estructuras factoriales.



Análisis Factorial Confirmatorio (AFC)

El análisis factorial confirmatorio (AFC) es una técnica estadística utilizada en investigación para evaluar la adecuación de un modelo factorial propuesto a los datos observados. A diferencia del análisis factorial exploratorio, donde se exploran las relaciones entre variables sin un modelo previo, el AFC se basa en una estructura teóricamente estable-

cida que se busca confirmar (Alavi et al., 2020; Hox, 2021; Sureshchandar, 2021).

En el AFC, se especifica un modelo teórico que describe las relaciones entre las variables y los factores subyacentes. Luego, se utiliza un conjunto de datos para evaluar cuán bien se ajustan los datos observados al modelo propuesto. Se emplean diversas medidas de ajuste, como el índice de ajuste comparativo (CFI), el error cuadrático medio de aproximación (RMSEA) y el test de razón de verosimilitud, para determinar la validez del modelo.

El AFC es valioso en la validación de teorías y modelos específicos, ya que permite a los investigadores probar hipótesis sobre la estructura subyacente de sus datos. Además, facilita la confirmación o refutación de relaciones teóricas establecidas previamente. Este enfoque es particularmente útil en disciplinas como la psicología, la sociología y la educación, donde se busca evaluar la validez de constructos teóricos a través de la confirmación estadística de sus estructuras subyacentes.

En el ámbito de la investigación, el análisis factorial confirmatorio (AFC) se erige como un faro que guía la validación de teorías y la confirmación de estructuras subyacentes. Este ensayo se sumerge en el mundo de las pruebas y puntos de corte en el AFC, explorando su importancia en la evaluación de modelos teóricos y la interpretación de resultados.

Uno de los pilares fundamentales del AFC es la especificación de un modelo teórico sólido. Sin embargo, la validación de este modelo va más allá de la formulación teórica; implica someterlo a pruebas estadísticas exhaustivas. Entre estas pruebas, los índices de ajuste como el Comparative Fit Index (CFI) y el Root Mean Square Error of Approximation (RMSEA) destacan como faros que indican la congruencia entre el modelo y los datos observados. Estos índices proporcionan puntos de corte cruciales, donde valores superiores al umbral establecido sugieren un ajuste aceptable (Tennant, 2012).

El punto de corte del CFI, generalmente buscando valores cercanos a 1, indica cuán bien el modelo reproduce las relaciones teóricas subyacentes. Por otro lado, el punto de corte del RMSEA, idealmente por debajo de 0,08, evalúa la calidad del ajuste del modelo en relación con la población. Estos puntos de corte, como faros en la oscuridad, iluminan el camino hacia la aceptación o rechazo del modelo propuesto.

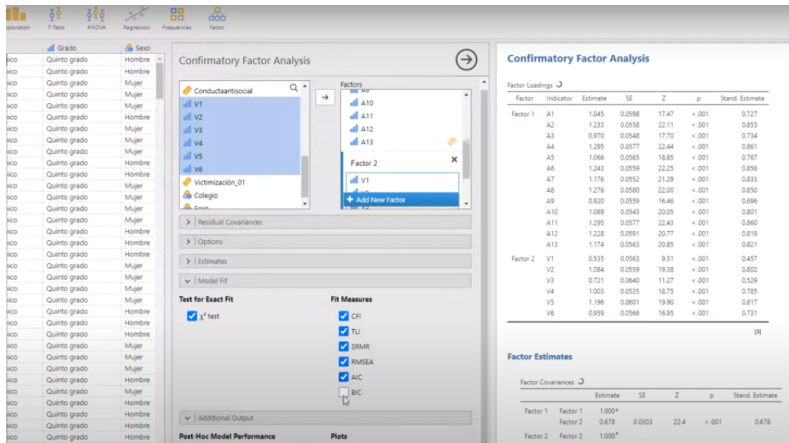
La interpretación de los resultados del AFC también involucra la evaluación de la significancia de los parámetros del modelo. Puntos de corte, como valores de tasa de error estándar (SE) y valores *p* asociados, proporcionan señales vitales sobre la robustez y fiabilidad de las relaciones entre variables. La atención cuidadosa a estos puntos de corte permite a los investigadores discernir la fuerza de las conexiones teóricas propuestas.

El AFC no solo se detiene en la confirmación estática de modelos; también permite la comparación entre modelos

alternativos. Las pruebas de diferencia de ajuste, con sus propios puntos de corte específicos, facilitan la elección del modelo más adecuado. Este proceso comparativo valida la teoría original y fomenta la mejora continua y la adaptación a complejidades inesperadas en los datos.

Además, las pruebas y puntos de corte en el AFC sirven como herramientas vitales en la evaluación de la calidad y validez de modelos teóricos. Como navegantes intrépidos, los investigadores confían en estos faros para orientarse en el vasto mar de datos y teorías, asegurándose de que sus descubrimientos sean robustos, confiables y significativos en la comprensión de la complejidad de los fenómenos estudiados.

Para llevar a cabo un Análisis Factorial Confirmatorio (AFC) en Jamovi, comienza abriendo tu conjunto de datos y dirígete a la pestaña "Analizar". Selecciona "Análisis Factorial Confirmatorio" en el menú desplegable y configura las variables relevantes, especificando la estructura del modelo teórico. Define restricciones y elige un método de estimación. Al ejecutar el análisis, revisa los índices de ajuste, como CFI y RMSEA, para evaluar la bondad de ajuste del modelo. Examina las cargas factoriales y otros parámetros para interpretar la relación entre variables. Si es necesario, ajusta el modelo y compáralo con alternativas usando pruebas de diferencia de ajuste. Guarda los resultados y, si es preciso, itera ajustando y refinando el modelo hasta obtener una representación adecuada de la estructura subyacente a tus datos.

Figura 13. Análisis factorial confirmatorio (AFC) en Jamovi.

Fuente: Los autores

En conclusión, este capítulo ha abordado aspectos fundamentales en la evaluación y comprensión de las medidas psicométricas a través del análisis de fiabilidad y los análisis factoriales, tanto exploratorios como confirmatorios. El análisis de fiabilidad ha sido crucial para entender la consistencia interna de las medidas, proporcionando insights sobre la confiabilidad de las puntuaciones obtenidas. Además, se exploraron diversas metodologías para estimar la fiabilidad, destacando la importancia de considerar la estabilidad temporal y la consistencia interna.

En cuanto al análisis factorial exploratorio, se revisaron los pasos esenciales en Jamovi, desde la apertura de datos hasta la interpretación de resultados, resaltando la necesi-



dad de un enfoque teórico sólido. Por último, se abordó el análisis factorial confirmatorio, subrayando la importancia de contrastar modelos teóricos específicos con los datos observados. Estos análisis proporcionan herramientas cuantitativas valiosas y enriquecen la comprensión global de las mediciones utilizadas en el ámbito psicométrico.

REFERENCIAS

- Alavi, M., Visentin, D. C., Thapa, D. K., Hunt, G. E., Watson, R., & Cleary, M. (2020). Chi-square for model fit in confirmatory factor analysis. *Journal of Advanced Nursing*, 76(9), 2209-2211. <https://doi.org/10.1111/jan.14399>
- Ardila Duarte, G. (2022). *Bioestadística descriptiva. Rigurosamente sencilla*. (1era ed., Vol. 1). Ecoe Ediciones.
- Beierl, E., Bühner, M., & Heene, M. (2018). Is that measure really one-dimensional? Nuisance parameters can mask severe model misspecification when assessing factorial validity. *Methodology*, 14(4). <https://ora.ox.ac.uk/objects/uuid:59631d88-9805-4634-8237-3714870971f2>
- Bonita, R., Beaglehole, R., & Kjellström, T. (2008). *Epidemiología Básica* (2da ed.). Organización Panamericana de la Salud.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. <https://doi.org/10.1007/BF02310555>

- De Boeck, P., & Elosua, P. (2016). Reliability and Validity: History, Notions, Methods, and Discussion. En F. T. L. Leong, D. Bartram, F. Cheung, K. F. Geisinger, & D. Ilescu (Eds.), *The ITC International Handbook of Testing and Assessment* (p. 0). Oxford University Press. <https://doi.org/10.1093/med:psych/9780199356942.003.0028>
- Donis, J. H. (2013). Tipos de diseños de los estudios clínicos y epidemiológicos. *Avances en Biomedicina*, 2(2), 76-99. <https://www.redalyc.org/pdf/3313/331327989005.pdf>
- Effendi, M., Zamri Khairani, A., Adnan, R., & Matore, E. M. (2019). Exploratory Factor Analysis (EFA) for Adversity Quotient (AQ) Instrument among Youth (SCOPUS). *Journal of Critical Reviews*, 6(6), 234-242. <https://doi.org/10.22159/jcr.06.06.33>
- Espinoza Casco, R. J., Sánchez Camargo, M. R., Velasco Taipe, M. A., Gonzales Sánchez, A., Romero-Carazas, R., & Mory Chiparra, W. E. (2023). METODOLOGÍA Y ESTADÍSTICA EN LA INVESTIGACIÓN CIENTÍFICA. *En Puerto Madero Editorial Académica* (2020-). Puerto Madero Editorial Académica. <https://doi.org/10.55204/PMEA.17>
- Ferrando, P. J., & Lorenzo-Seva, U. (2014). El Análisis Factorial Exploratorio de los Ítems: Algunas consideraciones

adicionales. *Anales de Psicología*, 30(3), 1170-1175.
<https://doi.org/10.6018/analesps.30.3.199991>

Forsythe, G. B., McGaghie, W. C., & Friedman, C. P. (1986). Construct Validity of Medical Clinical Competence Measures: A Multitrait-Multimethod Matrix Study Using Confirmatory Factor Analysis. *American Educational Research Journal*, 23(2), 315-336. <https://doi.org/10.3102/00028312023002315>

Gadermann, A., Guhn, M., & Zumbo, B. (2019). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research, and Evaluation*, 17(1). <https://doi.org/10.7275/n560-j767>

Hernández Sampieri, R., Fernández Collado, C., & Baptista Lucio, P. (2010). *Metodología de la Investigación* (5ta ed.). McGrawHill.

Hernández-Avila, M., Garrido-Latorre, F., & López-Moreno, S. (2000). Diseño de estudios epidemiológicos. *Salud Pública de México*, 42(2), 144-154. <https://doi.org/10.1590/S0036-36342000000200010>

Hox, J. j. (2021). Confirmatory Factor Analysis. En C. Barnes & D. R. Forde (Eds.), *The Encyclopedia of Research Methods in Criminology and Criminal Justice* (pp. 830-832). John Wiley & Sons, Ltd.

- Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3), 187-200. <https://doi.org/10.1007/BF02289233>
- López Segovia, L. (2020). *Bioestadística y sus aplicaciones*. Lucas López Segovia.
- Lozano Poveda, D. (2012). *Principios de epidemiología* (2da ed.). Pontificia Universidad Javeriana. <http://repository.javeriana.edu.co/handle/10554/47755>
- MacCallum, R. C., Browne, M. W., & Cai, L. (2007). Factor analysis models as approximations. En R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 153-175). Lawrence Erlbaum Associates Publishers.
- Maldonado Reyes, F. B., Álvarez Ochoa, R. I., Maldonado Córdova, P. A., Cordero Cordero, G. del R., & Capote Llanares, M. Á. (2023). METODOLOGÍA DE LA INVESTIGACIÓN: De la teoría a la práctica. *En Puerto Madero Editorial Académica* (2020-; 1era ed.). Puerto Madero Editorial Académica. <https://doi.org/10.55204/pmea.24>
- Manterola, C., Grande, L., Otzen, T., García, N., Salazar, P., & Quiroz, G. (2018). Confiabilidad, precisión o reproducibilidad de las mediciones. Métodos de valoración, utilidad y aplicaciones en la práctica clínica. *Revista*

chilena de infectología, 35(6), 680-688. <https://doi.org/10.4067/S0716-10182018000600680>

Martín Andrés, A., & Luna del Castillo, J. de D. (2004). *Bioestadística para las Ciencias de la Salud (+)* (1era ed.). Capitel Editores.

Martínez Hernández, J. (2013). *Nociones de Salud Pública* (2da ed.). Ediciones Díaz de Santos.

McDonald, R. (1985). *Factor analysis and related methods*. Lawrence Erlbaum Associates Publishers.

Medina, M. C. (2011). Generalidades de las pruebas diagnósticas, y su utilidad en la toma de decisiones médicas. *Revista Colombiana de Psiquiatría*, 40(4), 787-797. http://www.scielo.org.co/scielo.php?script=sci_abstract&pid=S0034-74502011000400015&lng=en&nrm=iso&tlng=es

Meredith, W. (1964). Rotation to achieve factorial invariance. *Psychometrika*, 29(2), 187-206. <https://doi.org/10.1007/BF02289700>

Moreno-Altamirano, A., López-Moreno, S., & Corcho-Berdugo, A. (2000). Principales medidas en epidemiología. *Salud Pública de México*, 42(4), 337-348. <https://doi.org/10.1590/S0036-36342000000400009>

Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Exploratory structural equation modeling. En G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A*

second course (2nd ed., pp. 395-436). IAP Information Age Publishing.

Morin, A. J. S., Myers, N. D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM), and BifactorESEM. En G. Tenenbaum, R. C. Eklund, & N. Boiangin (Eds.), *Handbook of sport psychology: Exercise, methodologies, & special topics, Vol. 2, 4th ed* (pp. 1044-1073). John Wiley & Sons, Inc. <https://doi.org/10.1002/9781119568124.ch51>

Muñiz, J. (2018). *Introducción a la Psicometría. Teoría clásica y TRI*. Pirámide.

Nunnally, J. C. (1978). *Psychometric Theory* (2da ed.). McGraw-Hill.

Organización Panamericana de la Salud. (2011). *Módulo de Principios de Epidemiología para el Control de Enfermedades (MOPECE)* (2da ed.). Organización Panamericana de la Salud.

Revelle, W. R. (2019). *psych: Procedures for Personality and Psychological Research*. Northwestern Scholars. <https://www.scholars.northwestern.edu/en/publications/psych-procedures-for-personality-and-psychological-research/fingerprints/>

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and catego-

- tical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354-373. <https://doi.org/10.1037/a0029315>
- Ricolfi, L., & Testa, S. (2021). Factorial Invariance and Orthogonal Rotation. *Multivariate Behavioral Research*, 56(5), 787-804. <https://doi.org/10.1080/00273171.2020.1770571>
- Romero Sandoval, N., Valcárcer, I., Sánchez Pérez, H. J., & Martín Mateo, M. (2015). *Principios de Epidemiología: Tasas y estandarización, análisis poblacionales y muestrales*. Bubok.
- Schweizer, K., Schreiner, M., & Gold, A. (2009). The confirmatory investigation of APM items with loadings as a function of the position and easiness of items: A two-dimensional model of APM. *Psychology Science*, 51(1), 47-64.
- Shrestha, N. (2021). Factor Analysis as a Tool for Survey Analysis. *American Journal of Applied Mathematics and Statistics*, 9(1), Article 1. <https://doi.org/10.12691/ajams-9-1-2>
- Sokal, R. R., & Rohlf, F. J. (2002). *Introducción a la bioestadística*. Reverte.
- Sureshchandar, G. S. (2021). Quality 4.0 – a measurement model using the confirmatory factor analysis (CFA) approach. *International Journal of Quality &*

Reliability Management, 40(1), 280-303. <https://doi.org/10.1108/IJQRM-06-2021-0172>

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. *Frontiers in Psychology*, 7. <https://www.frontiersin.org/articles/10.3389/fpsyg.2016.00769>

Tukey, J. W. (1977). *Exploratory Data Analysis* (1era ed.). Pearson.

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). Un viaje alrededor de alfa y omega para estimar la fiabilidad de consistencia interna. *Anales de Psicología / Annals of Psychology*, 33(3), Article 3. <https://doi.org/10.6018/analesps.33.3.268401>

Villa Romero, A. R., Moreno Altamirano, L., & García de la Torre, G. (2012). *Epidemiología y estadística en salud pública* (2da ed.). McGrawHill.

Vizcaíno-Salazar, G. J. (2017). Importancia del cálculo de la sensibilidad, la especificidad y otros parámetros estadísticos en el uso de las pruebas de diagnóstico clínico y de laboratorio. *Medicina y Laboratorio*, 23(7-8), 365-386. <https://doi.org/10.36384/01232576.34>

Wrigley, C., Saunders, D. R., & Neuhaus, J. O. (1958). Application of the quartimax method of rotation to Thurstone's primary mental abilities study. *Psychometrika*, 23(2), 151-170. <https://doi.org/10.1007/BF02289011>

Tennant A, Pallant JF. The root mean square error of approximation (RMSEA) as a supplementary statistics to determine fit to the Rasch model with large sample sizes. *Rasch Meas Trans.* 2012; 25(4): 1348.

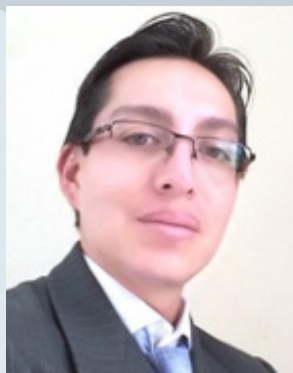
Sobre los autores



FAUSTO BOLÍVAR MALDONADO REYES


Master of Public Health Bélgica. Diplomado en Docencia Universitaria por la Universidad Católica de Cuenca Ecuador. Diplomado en Pensamiento Crítico por la Universidad Católica de Cuenca. Docente Universidad Católica de Cuenca. Hospital San Martín, Azogues, Ecuador. Ha participado como ponente en eventos a nivel local, nacional e internacional. Es autor y coautor de varios artículos en revistas indexadas y arbitradas de alto impacto, así como de libros y capítulos de libro. Sus líneas de investigación abordan temáticas relacionadas con la investigación en educación superior, salud pública y epidemiología.

ORCID: <https://orcid.org/0000-0001-5272-4041>



ROBERT IVÁN ÁLVAREZ OCHOA

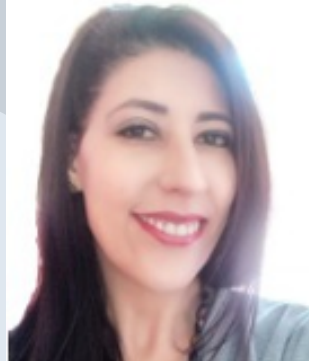
Bioquímico-Farmacéutico por la Universidad de Cuenca. Magister en Nutrición Infantil por la Universidad de Especialidades Espíritu Santo. Magister en Educación, Tecnología e Innovación por la Universidad Católica de Cuenca. Doctorando en Ciencias de la Educación por la Universidad Santander. Diplomado en Nutrición Infantil por el Politécnico de Colombia. Diplomado en Formación de Investigadores por la Universidad Autónoma de Nuevo León. Diplomado en Gestión y Edición de Revistas Científicas por el Centro de Investigación y Desarrollo Ecuador. Docente Investigador de la Universidad Católica de Cuenca (UCACUE, Ecuador). Miembro del Sistema Nacional de Investigadores (SENESCYT, Ecuador). Miembro activo de la Red de Estudios sobre Educación REED.



Miembro de la Red Ecuatoriana de Investigación Científica Inclusiva Multidisciplinar REICIM y de la Red de Docentes de América Latina y el Caribe REDDOLAC. Ha participado como ponente y conferencista en eventos a nivel local, nacional e internacional. Es evaluador de proyectos de investigación y revisor en varias revistas arbitradas. Es autor y coautor de varios artículos en revistas indexadas y arbitradas de alto impacto, así como de libros y capítulos de libro. Es director y colaborador de varios proyectos de investigación. Sus líneas de investigación abordan temáticas relacionadas con la investigación en educación superior, tecnología e innovación, formación y desarrollo del profesorado, alimentación y nutrición, ciencias médicas y de la salud.

Correo electrónico: rialvarezo@ucacue.edu.ec

ORCID: <https://orcid.org/0000-0002-2431-179X>



GABRIELA CORDERO CORDERO

Bioquímica-Farmacéutica por la Universidad de Cuenca. Magister en Nutrición Infantil por la Universidad de Especialidades Espíritu Santo. Magister en Educación, Tecnología e Innovación por la Universidad Católica de Cuenca. Docente de la Universidad Católica de Cuenca (UCACUE, Ecuador). Miembro del Sistema Nacional de Investigadores (SENESCYT, Ecuador). Miembro activo de la Red de Docentes de América Latina y el Caribe REDDOLAC. Ha participado como ponente en eventos a nivel local, nacional e internacional. Es autora y coautora de varios artículos en revistas indexadas y arbitradas de alto impacto, así como de libros y capítulos de libro. Es colaboradora de varios proyectos de investigación. Sus líneas de investigación abordan temáticas relacionadas con la investigación en educación superior, alimentación y nutrición, ciencias médicas y de la salud.

Correo electrónico: grcorderoc@ucacue.edu.ec

ORCID: <https://orcid.org/0000-0001-7278-2177>



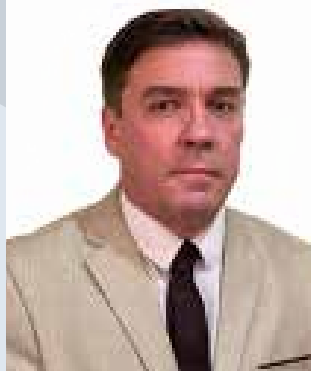
ANDRÉS RAMÍREZ

Licenciado en Ciencias de la Educación mención Psicología Educativa y Orientación Vocacional por la Universidad Católica de Cuenca. Máster en Neuropsicología y Educación con énfasis en Investigación Científica por la Universidad Internacional de la Rioja.

Cursando Doctorado en Psicología en la Universidad de Palermo- Argentina. Cursando Doctorado de Epidemiología y Bioestadística en la Universidad CES de Colombia. Docente e Investigador principal de la facultad de Enfermería sede Azogues, profesor e Investigador principal en la Maestría en Gestión del Cuidado - Posgrados de la Universidad Católica de Cuenca. Coordinador del Laboratorio de Psicometría, Psicología Comparada y Etología de la Universidad Católica de Cuenca. Coordinador investigador en la Unidad Distrital de Apoyo a la Inclusión (UDAI). Revisor de Artículos Científicos en la Universidad Espíritu Santo.

Correo electrónico: andres.ramirez@ucacue.edu.ec

ORCID: <https://orcid.org/0009-0007-3493-6519>



MIGUEL ÁNGEL CAPOTE LLANARES

Doctor en Medicina por la Universidad Médica de la Habana, Cuba. Especialista de primer grado en ortopedia y traumatología por la Universidad Médica de la Habana, Cuba. Miembro titular de la Sociedad Cubana de Ortopedia y Traumatología. Miembro adjunto de las sociedades ecuatoriana, Mozambicana y Francesa de ortopedia y traumatología. Magister en medicina bioenergética y fisioterapia. Magister en medicina deportiva. Diplomado en Gerencia de salud. Diplomado en Docencia y educación médica superior. Diplomado en técnicas de administración y jurídicas. Docente de la Universidad Católica de Cuenca (UCACUE), Sede Azogues. Responsable de Formación Investigativa de la Carrera de Medicina de la UCACUE, Sede Azogues.

Correo electrónico: macapote2010@gmail.com
ORCID: <https://orcid.org/0000-0001-6380-9251>

REDLIC | Red Editorial
Latinoamericana de
Investigación Contemporánea

L o e s e n c i a l e n
BIOESTADÍSTICA MÉDICA

2024

"Lo esencial en Bioestadística Médica" es una obra imprescindible que aborda de manera clara y concisa los fundamentos de la bioestadística aplicada a la Medicina. Este libro se divide en seis capítulos que cubren aspectos fundamentales para comprender y aplicar la estadística en el ámbito de la salud. Desde las medidas epidemiológicas en salud hasta la introducción a la psicometría, cada capítulo proporciona una base sólida para el análisis estadístico en el campo médico.

En los primeros capítulos, se exploran las medidas epidemiológicas clave utilizadas en la investigación en salud, así como la importancia de la validez y la reproducibilidad de las pruebas diagnósticas. Se presentan los diferentes diseños de estudio y se discuten las generalidades de la estadística descriptiva, incluyendo el análisis de una y dos variables. Estos temas sientan las bases para comprender cómo se recopilan y analizan los datos en el contexto de la medicina.

El libro continúa adentrándose en la estadística analítica e inferencial, proporcionando al lector las herramientas necesarias para interpretar y aplicar los resultados de estudios médicos. Se abordan técnicas avanzadas de análisis estadístico y se introduce al lector en el mundo de la psicometría, explorando el concepto de fiabilidad como medida de precisión en el proceso de medición. Con un enfoque práctico y ejemplos relevantes, "Lo esencial en Bioestadística Médica" se convierte en un recurso valioso tanto para estudiantes de medicina como para profesionales de la salud que buscan mejorar su comprensión y aplicación de la estadística en el campo médico. ¡Sumérgete en este libro y descubre el poder de la bioestadística en la práctica clínica y la investigación médica!

REDLIC Red Editorial
Latinoamericana de
Investigación Contemporánea

