

**TECANA AMERICAN UNIVERSITY**  
**ACCELERATED DEGREE PROGRAM**  
**DOCTORATE OF PHILOSOPHY IN**  
**EPISTEMOLOGY AND RESEARCH SCIENTIFIC**



**INFORME N° 02**

**LA MEDICIÓN EN EL PROCESAMIENTO Y ANÁLISIS DE  
DATOS**

**Cursante: Billy Williams Moisés Ríos, Mg.**

**Curso avanzado: Processing and Analysis of Data**

**“Por la presente juro y doy fe que soy el único autor del presente informe y que su contenido es fruto de mi trabajo, experiencia e investigación académica”**

**Ayacucho, Perú, Agosto 2018**

## ÍNDICE GENERAL

	<b>Pág.</b>
<b>RESUMEN</b>	V
<b>INTRODUCCIÓN</b>	1
<b>OBJETIVOS</b>	2
<b>General</b>	2
<b>Específicos</b>	2
<b>JUSTIFICACIÓN</b>	3
<b>CAPÍTULO I</b>	
<b>MEDICIÓN</b>	4
1.1.  Nociones sobre el medición	4
1.2.  Definiciones sobre medición	5
1.3.  Funciones de la medición	5
<b>CAPÍTULO II</b>	
<b>VARIABLES Y ESCALAS DE MEDICIÓN</b>	6
2.1.  Variable	6
2.2.  Clasificación	6
2.3.  Escala de medición	10
<b>CAPÍTULO III</b>	
<b>PRINCIPIOS DE LA MEDICIÓN</b>	13
3.1.  Validez	13
3.2.  Confiabilidad	14
<b>CAPÍTULO IV</b>	
<b>PROCESAMIENTO Y ANÁLISIS DE DATOS</b>	19
4.1.  Datos	19
4.2.  Procesamiento de datos	21
4.3.  Análisis de datos	22

<b>CAPÍTULO V</b>	
<b>ANÁLISIS DESCRIPTIVO</b>	25
5.1. Distribución de frecuencias	25
5.2. Medidas de tendencia central	26
5.3. Medidas de dispersión	29
5.4. Medidas de posición	31
5.5. Medidas de forma	35
<b>CAPÍTULO VI</b>	
<b>ANÁLISIS INFERENCIAL</b>	39
6.1. Análisis paramétrico	39
6.2. Análisis no paramétrico	41
<b>CONCLUSIONES</b>	46
<b>BIBLIOGRAFÍA</b>	49

## ÍNDICE DE FIGURAS

<b>FIGURA</b>	<b>Pág.</b>
1. Valores que adopta la variable continua	9
2. Valores que adopta la variable discreta	10
3. Escalas de medición	10
4. Relación entre variable y dato	19
5. Dato e información significativa	20
6. Tipos de análisis de datos	23
7. Tipo de asimetría	36
8. Tipo de apuntamiento	37
9. Curva normal	39

Tecana American University  
Doctorate of Philosophy (Ph.D) in Epistemology and Research Scientific  
Measurement in Processing and Data Analysis

**Informe N° 1**

**LA MEDICIÓN EN EL PROCESAMIENTO Y ANÁLISIS DE DATOS**

AUTOR: Billy Williams Moisés Ríos

Agosto 2018

**RESUMEN**

El presente informe tiene como objetivo general “Discutir los alcances de la medición en el procesamiento y análisis de datos”. Su fundamentación bibliográfica se basó, entre otros, en los siguientes autores: (Blair & Taylor, 2008), (Moisés Ríos, Valenzuela Oré, Pariona Cahuana, & Bendezú Prado, 2018), (Moisés Ríos, Anjo Bedriñana, Palomino Vargas, & Feria Macizo, 2018), (Valderrama M., 2011), (Posada Hernández, 2016) y (Triola, 2009). Este informe de revisión bibliográfica concluye que en el campo de la investigación cuantitativa, la medición es una condición necesaria para el tratamiento de las variables: procesamiento y análisis de datos.

**Palabras claves:** Medición, variable, escala de medición, confiabilidad, validez, dato, procesamiento, análisis.

## INTRODUCCIÓN

Una de las características de la ciencia fáctica es la precisión para efectivizar la comunicación mediante el lenguaje científico claro e unívoco, evitando ambigüedades. Con la medición de las diferentes propiedades del objeto de estudio se logra una mayor precisión en la comunicación de los conocimientos adquiridos, aunque la pretensión de alcanzar una certeza absoluta acerca del valor real de lo que se pretende medir no sea accesible.

El procesamiento y análisis de datos no exceptúa la medición, porque los datos necesitan ser caracterizados, clasificados, agrupados y sometidos a una valoración cuantitativa para una mejor interpretación. De esta manera, es posible hallar pautas de comportamiento dentro de los datos y entre las variables para la toma de decisiones informadas.

El objetivo general del presente informe de revisión bibliográfica fue discutir los alcances de la medición en el procesamiento y análisis de datos. Para tal efecto, se ha estructurado en seis capítulos. El primer capítulo, medición, presenta las nociones, definiciones más comunes y funciones de la medición. El capítulo II, variables y escalas de medición, diferencia conceptualmente la variable de una constante, clasifica las variables según su naturaleza y contextualiza las escalas de medición. El capítulo III, principios de la medición, trata sobre validez y confiabilidad de las medidas. El capítulo IV, procesamiento y análisis de datos, delimita el significado de dato, define el procesamiento y análisis de datos, describiendo sus etapas. El capítulo V, análisis descriptivo, está referida a la distribución de frecuencias y medidas de resumen (tendencia central, dispersión, posición y forma). El capítulo VI, análisis inferencial, presenta el análisis paramétrico y no paramétrico. Finalmente, se consignan las conclusiones y la bibliografía.

## **OBJETIVOS**

### **GENERAL**

Discutir los alcances de la medición en el procesamiento y análisis de datos.

### **ESPECÍFICOS**

1. Exponer las definiciones sobre medición.
2. Clasificar las variables y sus escalas de medición.
3. Explicar los principios de la medición.
4. Exponer los alcances del procesamiento y análisis de datos.
5. Explicar los criterios para la aplicación de pruebas paramétricas.
6. Explicar los criterios para la aplicación de pruebas no paramétricas.

## **JUSTIFICACIÓN**

En el contexto actual, la estadística se ha convertido en una herramienta imprescindible para describir y analizar datos, también como apoyo en la toma de decisiones informadas. Por tanto, el estudio del procesamiento y análisis de datos permite conocer los métodos y procedimientos para recoger, clasificar, resumir y analizar datos para hallar patrones de comportamiento dentro de los datos y entre las variables.

## CAPÍTULO I

### MEDICIÓN: MEDIO DE VERIFICACIÓN DE LA EXISTENCIA DE LAS VARIABLES

#### 1.1. Nociones sobre medición

“(…) en la antigüedad, cuando se quería medir algún atributo físico de los objetos, como podía ser su peso, su longitud, etc., se trataban de desarrollar instrumentos pertinentes y nadie dudaba acerca de la posibilidad de llevar a cabo tales mediciones” (Barbero García, Vila Abad, & Holgado Tello, 2015, pág. 9).

En el campo de la investigación cuantitativa, la medición es una condición necesaria para el tratamiento de las variables.

Semánticamente una variable se define como la propiedad del objeto de estudio, **factible de medición** y capaz de asumir diferentes valores o categorías.

En el razonamiento de Pitágoras (572-497 a.C.) y de la escuela pitagórica, el número constituía la esencia de las cosas: “todo es número” (Moisés Ríos, Anjo Bedriñana, Palomino Vargas, & Feria Macizo, 2018, pág. 104).

Para Galileo Galilei (1564-1642), astrónomo, filósofo, ingeniero, matemático y físico italiano, la medición era inherente al trabajo investigativo: “Mide lo que se pueda medir; y lo que no, hazlo medible”.

Desde esta lógica, en potencia, todos los datos son cuantificables (Kerlinger & Lee, 2002) o más acertadamente: “casi todas las propiedades son medibles”.

La medición de las propiedades del objeto de estudio de naturaleza concreta y abstracta involucra la utilización de instrumentos con cierto grado de precisión: una especie de exteriorización de los sentidos en condiciones controladas, para evidenciar la existencia de lo que se pretende medir.

Cuanto más precisa sea la medición, más información nos suministra sobre la variable y, por tanto, sobre la unidad de análisis. No es lo mismo decir que una persona es alta, a decir que mide 1,83 metros. (Cazau, 2016)

De lo expresado anteriormente, se concluye que la medición es el proceso de aproximación al valor real de la propiedad del objeto de estudio con algún grado de precisión.

## **1.2. Definiciones sobre medición**

Medir es el proceso según el cual "el hecho se vuelve número" (Badiou, 1972, pág. 25).

“Es la asignación de números a entidades, acontecimientos o sucesos, con el fin de representar sus propiedades y sus relaciones” (Barbero García, Vila Abad, & Holgado Tello, 2015, pág. 9).

Es un procedimiento que consiste en un conjunto de reglas para asignar números a objetos para representar cantidades de atributos (Nunnally & Bernstein, 1995).

La medición “es el proceso de asignar una magnitud a las propiedades del objeto, para establecer un continuo de valores” (Moisés Ríos, Ango Bedriñana, Palomino Vargas, & Feria Macizo, 2018, pág. 80).

## **1.3. Funciones de la medición**

Una de las características de la ciencia fáctica es efectivizar la comunicación mediante el lenguaje científico, preciso, claro e unívoco, evitando ambigüedades, para uniformizar su comprensión y entendimiento (Moisés Ríos, Ango Bedriñana, Palomino Vargas, & Feria Macizo, 2018).

Con la medición de las diferentes propiedades del objeto de estudio se logra una mayor precisión en la comunicación de los conocimientos adquiridos, aunque la pretensión de alcanzar una certeza absoluta acerca del valor real de lo que se pretende medir no sea accesible: tiene asociada una incertidumbre sea por la naturaleza de la magnitud que se mide, el uso de instrumentos de medición, el observador o las condiciones externas.

Por esta razón, “el conocimiento científico procura la precisión; aunque nunca está enteramente libre de vaguedades” (Bunge, 1997, pág. 13).

## CAPÍTULO II

### VARIABLES Y ESCALAS DE MEDICIÓN

#### 2.1. Variable

Las cinco primeras letras del término variable otorgan su significado exacto: **V-A-R-I-A**. Por tanto, la característica o propiedad del objeto de estudio que no varía es una **constante**.

“Se considera como variable a cualquier característica o propiedad general de una población que sea posible medir con distintos valores o describir con diferentes modalidades” (Posada Hernández, 2016, pág. 16).

“La propiedad variante del objeto de estudio, medible directa o indirectamente, (...) da como resultado valores o categorías diferentes (más de uno) que la diferencian de la constante (asume un único valor o categoría)” (Moisés Ríos, Ango Bedriñana, Palomino Vargas, & Feria Macizo, 2018, pág. 82).

#### 2.2. Clasificación

Por su naturaleza las variables se clasifican en cualitativas, cuasi-cuantitativas y cuantitativas.

##### 2.2.1. Variable cualitativa

“Expresa distintas cualidades, características o modalidades. Cada modalidad refleja una categoría o atributo diferente y su clasificación convencional prescinde la medición numérica o el ordenamiento jerárquico entre categorías” (Moisés Ríos, Ango Bedriñana, Palomino Vargas, & Feria Macizo, 2018, pág. 89).

“Las variables cualitativas son aquellas que representan atributos de los elementos y no permiten una representación numérica definida” (Posada Hernández, 2016, pág. 16).

Es toda propiedad o característica del objeto de estudio que varía en modalidades o categorías diferentes: no expresan cantidades. En consecuencia, los “números convencionales” que se asignan a las categorías (en el procesamiento de datos) solo sirven de etiqueta para su mejor identificación. Significa que no puede realizarse ninguna operación aritmética.

**Ejemplo:**

Variable	Categoría
Filiación religiosa	1. Católica 2. Evangélica 3. Adventista
Género	1. Varón 2. Mujer
Lugar de procedencia	1. Arequipa 2. Ayacucho 3. Junín 4. Lima
Rendimiento académico	1. Aprobado 2. Desaprobado
Sexo	1. Masculino 2. Femenino

**2.2.2. Variable cuasi-cuantitativa**

Es toda propiedad o característica del objeto de estudio con categorías jerarquizadas por orden natural. En consecuencia, los “números convencionales” asignados son ordinales: hay un primero, un segundo y así sucesivamente.

“Los valores o categorías poseen un orden, secuencia o progresión natural establecida. Generalmente refleja la intensidad de una propiedad o atributo (...). A pesar de este ordenamiento jerárquico, no es posible obtener una valoración numérica precisa entre dos valores o categorías” (Moisés Ríos, Anjo Bedriñana, Palomino Vargas, & Fera Macizo, 2018, pág. 89).

**Ejemplo:**

<b>Variable</b>	<b>Categoría</b>
Nivel de presión arterial	1. Hipotensión 2. Normotensión 3. Hipertensión
Estado nutricional (IMC)	1. Delgadez 2. Normal 3. Sobrepeso 4. Obesidad
Nivel de instrucción	1. Inicial 2. Primaria 3. Secundaria 4. Superior
Rendimiento académico	1. Inicio 2. Proceso 3. Logro previsto 4. Logro destacado
Condición socioeconómica	1. Baja 2. Media 3. Alta

**2.2.3. Variable cuantitativa**

Es toda propiedad o característica del objeto de estudio que puede ser contada o identificada según su magnitud (cantidad). En consecuencia, sus valores son números enteros o fraccionarios que posibilitan el cálculo de operaciones aritméticas.

“(…) las características de los elementos son observados cuantitativamente a través de una medida y una escala definidas” (Posada Hernández, 2016, pág. 16).

**Ejemplo:**

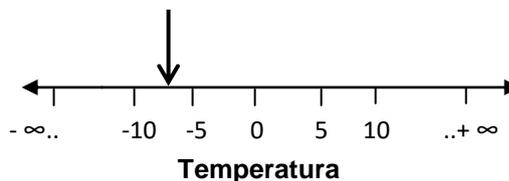
Variable	Valor
Edad	18-28 años
	29-39 años
	40-49 años
Número de hijos	1
	2
	3
Número de cursos aprobados	1
	2
	3
Rendimiento académico	0-5
	6-10
	11-15
	16-20

Las variables cuantitativas se clasifican en dos:

**(a) Variable cuantitativa continua**

“Una variable es cuantitativa continua si a lo largo de un intervalo puede tomar cualquier valor; es decir, entre uno y otro valor de la variable siempre puede existir otro valor intermedio” (Posada Hernández, 2016, pág. 17).

En este tipo de variable la unidad mínima de medición puede ser fraccionada infinitamente. Por tanto, adopta números fraccionarios o decimales.



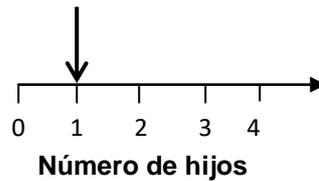
**Figura N° 1. Valores que adopta la variable cuantitativa continua.**

Fuente: (Moisés Ríos, Ango Bedriñana, Palomino Vargas, & Fera Macizo, 2018, pág. 90)

### (b) Variable cuantitativa discreta

Se caracteriza porque la unidad mínima de medición no puede ser fraccionada. Por tanto, adopta únicamente números enteros.

“Una variable es cuantitativa discreta si solo puede tomar un valor de un conjunto de números; existen separaciones entre dos valores sucesivos que no pueden llenarse con valores intermedios; en este caso la variable toma valores aislados” (Posada Hernández, 2016, pág. 17).

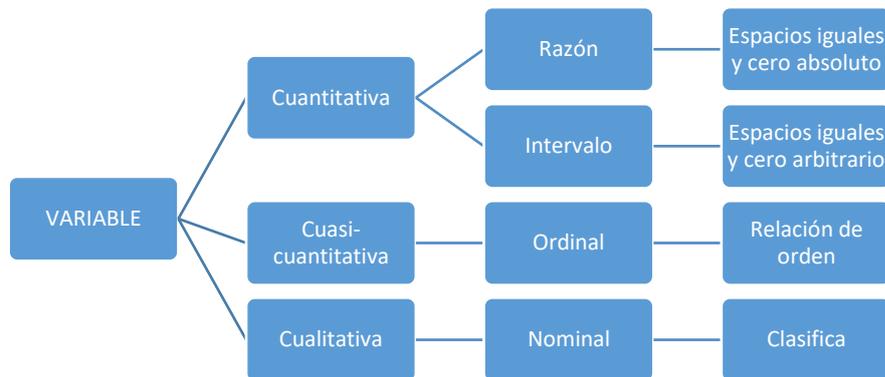


**Figura N° 2. Valores que adopta la variable cuantitativa discreta.**

Fuente: (Moisés Ríos, Ango Bedriñana, Palomino Vargas, & Feria Macizo, 2018, pág. 91)

### 2.3. Escalas de medición

“Es considerada como un sistema que asigna valores numéricos a características susceptibles de medir. Normalmente, las escalas pueden ser de cuatro tipos: nominal, ordinal, de intervalos y de razón” (Posada Hernández, 2016, pág. 18).



**Figura N° 3. Escalas de medición.**

Fuente: (Elaboración propia)

### 2.3.1. Nominal

“(…) posee dos características principales. Primero, produce clasificaciones de personas o cosas con base en una evaluación cualitativa de la característica que se somete a consideración; segundo, su uso no brinda información con respecto a la cantidad” (Blair & Taylor, 2008, pág. 10).

Las categorías de la variable únicamente nominan o nombran características específicas y mutuamente excluyentes del objeto de estudio, a manera de clasificación, sin indicar orden o jerarquía.

#### Ejemplo:

Variable	Categoría		
<b>Filiación religiosa</b>	1. Católica	1. Evangélica	1. Adventista
	2. Evangélica	2. Católica	2. Evangélica
	3. Adventista	3. Adventista	3. Católica

Los números asignados a las categorías, en los diferentes casos, no conservan la propiedad de orden ni jerarquía. En consecuencia, cualquiera de las formas de presentación de las modalidades de la variable es correcta.

### 2.2.2. Ordinal

“(…) clasifica personas o cosas sobre la base de la característica evaluada. Sin embargo, a diferencia de la escala nominal, las clasificaciones producidas por esta escala incorporan los atributos muy importantes de “mayor que” o “menor que” (Blair & Taylor, 2008, pág. 10).

Las categorías de la variable están ordenadas ascendente o descendentemente, reflejando jerarquía inherente entre categorías de una propiedad o característica del objeto de estudio.

Variable	Categoría		
Estado nutricional (IMC)	1. Delgadez	4. Obesidad	<del>1. Delgadez</del>
	2. Normal	3. Sobrepeso	<del>2. Sobrepeso</del>
	3. Sobrepeso	2. Normal	<del>3. Normal</del>
	4. Obesidad	1. Delgadez	<del>4. Obesidad</del>
	<b>Orden natural</b>		<b>Sin orden</b>
	<b>Ascendente</b>	<b>Descendente</b>	

### 2.2.3. Intervalo

“(…) agrega los atributos de cuánto más y cuánto menos a aquellos de la escala ordinal” (Blair & Taylor, 2008, pág. 11).

“Un defecto de la escala de intervalo es la falta de un punto cero verdadero. En otras palabras, el punto cero en esta escala es una designación arbitraria, lo cual significa que no representa una ausencia de la característica medida” (Blair & Taylor, 2008, pág. 11).

Los valores de la variable expresan orden, jerarquía y espacios iguales entre uno y otro. Además conservan un cero arbitrario que indica el punto de partida de la característica: el valor cero existe y representa cierta magnitud de la característica medida. Por ejemplo, un coeficiente intelectual de “0” no indica ausencia de inteligencia, sino que en las condiciones establecidas el valor cero corresponde a una magnitud del coeficiente de inteligencia a partir del cual se estructuran otros niveles de inteligencia.

### 2.2.4. Razón

“(…) posee un verdadero punto cero. Mediciones físicas como la estatura y el peso son ejemplos comunes. Cuando algo tiene peso cero, el cero indica que no hay peso presente” (Blair & Taylor, 2008, pág. 11).

Los valores de la variable expresan orden, jerarquía y espacios iguales entre uno y otro; pero a diferencia de la escala de intervalo, conservan un cero absoluto o real que indica la ausencia de la característica. Por ejemplo, “0 hijos” indica ausencia de hijos.

## CAPÍTULO III

### PRINCIPIOS DE LA MEDICIÓN

#### 3.1. Validez

Validez es la exactitud del instrumento para medir la variable que pretende medir. Un instrumento es válido en la medida que todos los ítems previstos para la variable miden lo mismo: “Cuando el instrumento mide de alguna manera demostrable aquello que trata de medir, libre de distorsiones sistemáticas” (Moisés Ríos, Anjo Bedriñana, Palomino Vargas, & Fera Macizo, 2018, pág. 82).

Es el grado de la evidencia empírica y teórica para proporcionar soporte a las interpretaciones de los puntajes de un instrumento de medición (Furr, 2011).

La literatura bibliográfica reporta diferentes tipos de validez (Moisés Ríos, Valenzuela Oré, Pariona Cahuana, & Bendezú Prado, 2018, pág. 50):

**(a) Validez de contenido.** Grado en que la medición de la variable representa al contenido previsto. Responde a las siguientes interrogantes:

- ¿Cuán representativos son los ítems de un comportamiento elegido respecto al contenido previsto?
- ¿En qué medida los ítems de una conducta elegida corresponden al contenido previsto?

**(b) Validez de constructo.** Grado en que la medición de la variable representa al concepto o dominio teórico. Responde a las siguientes interrogantes:

- ¿Cuán representativos son los ítems de una conducta elegida como muestra del universo que se pretende representar?
- ¿Qué significado tiene la conducta respecto a los atributos del individuo que son de interés para la medición?

**(c) Validez de criterio.** Grado en que una prueba se correlaciona con una variable criterio. El criterio es una variable diferente que sirve de referencia o patrón de correspondencia. Responde a la siguiente interrogante:

- ¿Hasta dónde puede predecirse la actuación del sujeto a partir de la ejecución de la prueba?

**(d) Validez predictiva.** Grado en que la medición predice conductas ejecutorias futuras:

- ¿Hasta dónde la medición puede predecir un resultado futuro?

**(e) Validez concurrente.** Grado en que la medición se relaciona con los índices del criterio, obtenidos aproximadamente al mismo tiempo. Responde a la siguiente interrogante:

- ¿Hasta dónde la medición se relaciona con una conducta criterio?

### **3.2. Confiabilidad**

“En el lenguaje cotidiano el término fiabilidad se asocia a algo que funciona de manera correcta. Nos fiamos de nuestro despertador si suena a la hora que se ha programado, de la báscula si nos proporciona sin error nuestro peso (...)” (Barrios & Cosculluela, 2013, pág. 75).

“Confiabilidad es la precisión del instrumento para medir la variable de interés. A mayor fiabilidad será menor la cantidad de errores aleatorios e impredecibles que aparecerán al utilizarlo” (Moisés Ríos, Valenzuela Oré, Pariona Cahuana, & Bendezú Prado, 2018, pág. 52).

En consecuencia, si la medida toma valores consistentes y precisos, creemos que podemos confiar en los resultados obtenidos (Barrios & Cosculluela, 2013).

Existen diferentes tipos de confiabilidad: estabilidad temporal, equivalencia y consistencia interna (Reidl Martínez, 2013).

### (a) Estabilidad temporal

Es el grado de variación en las puntuaciones de una prueba al medirlas en distintas ocasiones (dos veces) al mismo grupo de personas.

Tiene como propósito asegurar que si se presentan cambios en la variable de interés, se debieron al paso del tiempo (por ejemplo, la hora del día o debido al desarrollo) y no al instrumento de medición (Reidl Martínez, 2013).

- **Test-Retest.** Consiste en aplicar un mismo test a una misma muestra de sujetos en dos momentos temporales distintos para establecer el grado de correlación entre ambas medidas. La función de prueba es:

$$r_{xy} = \frac{n \sum x_i y_j - \sum x_i \sum y_j}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_j^2 - (\sum y_j)^2]}}$$

“Si el test es preciso, se espera que ambas medidas sean muy parecidas y el coeficiente de fiabilidad tenderá a 1. De ser así, se asume que el test es fiable porque independientemente del momento de aplicación se tiende a obtener la misma medida” (Moisés Ríos, Valenzuela Oré, Pariona Cahuana, & Bendezú Prado, 2018, pág. 52).

### (b) Consistencia interna

Grado en que los ítems que componen una prueba miden el mismo atributo psicológico. Refleja la constancia de los ítems para operar sobre un dominio psicológico de modo análogo (grado de homogeneidad entre los reactivos) (Moisés Ríos, Valenzuela Oré, Pariona Cahuana, & Bendezú Prado, 2018, pág. 52).

- **Alfa de Cronbach.** Se aplica a instrumentos de medición con ítems valorados de manera **politómica** (> 2 opciones). La función de prueba es:

$$\alpha = \frac{K}{K-1} \left[ 1 - \frac{\sum \sigma_i^2}{\sigma_T^2} \right]$$

**Donde:**

$\alpha$  : Coeficiente alfa.

$\sigma_i^2$  : Varianza de cada ítem.

$\sigma_T^2$  : Varianza del total.

K : Número de preguntas o ítems.

- **Kuder-Richardson KR<sub>20</sub>**. Se aplica a instrumentos de medición con ítems valorados de manera **dicotómica** (solo 2 opciones). La función de prueba es:

$$KR_{20} = \frac{K}{K-1} \left[ 1 - \frac{\sum pq}{\sigma_T^2} \right]$$

**Donde:**

KR<sub>20</sub> : Coeficiente de correlación de Kuder-Richardson.

p : Proporción de respuestas correctas al ítem.

q : Proporción de respuestas incorrectas al ítem.

$\sigma_T^2$  : Varianza del total.

K : El número de preguntas o ítems.

- **División por mitades**. Consiste en dividir prueba en dos mitades (ítems pares e ítems impares). Es necesario que los reactivos o preguntas estén ordenadas de acuerdo a su nivel de dificultad (de los más fáciles a los más difíciles) (Moisés Ríos, Valenzuela Oré, Pariona Cahuana, & Bendezú Prado, 2018, pág. 53). La función de prueba depende del procedimiento elegido:

**Spearman-Brown:**

$$r_{xy} = \frac{2r}{1+r}$$

**Donde:**

r: Coeficiente de correlación de Pearson entre las sumatorias de las puntuaciones de las dos mitades del test.

**Rulon:**

$$r_{xy} = 1 - \frac{\sigma_d^2}{\sigma_T^2}$$

**Donde:**

$\sigma_d^2$  : Varianza de la diferencia entre las sumatorias de las puntuaciones de ambas mitades del test.

$\sigma_T^2$  : Varianza del total.

**Guttman:**

$$r_{xy} = 2 \left( 1 - \frac{\sigma_x^2 + \sigma_y^2}{\sigma_T^2} \right)$$

**Donde:**

$\sigma_x^2$  : Varianza de la sumatoria de las puntuaciones de la primera mitad del test.

$\sigma_y^2$  : Varianza de la sumatoria de las puntuaciones de la segunda mitad del test.

$\sigma_T^2$  : Varianza del total.

Con los reactivos pares en uno de los conjuntos y los impares en el otro, se asegura de alguna manera la homogeneidad de dificultad en ambos conjuntos, propiciando que ambas distribuciones tengan medias y varianzas semejantes (Reidl Martínez, 2013).

### (c) Equivalencia

Grado en que las puntuaciones de dos versiones de un instrumento de medición son equiparables o equivalentes. Estas versiones, con diferentes reactivos o

estímulos pero con iguales dimensiones, deben medir lo mismo (Moisés Ríos, Valenzuela Oré, Pariona Cahuana, & Bendezú Prado, 2018, pág. 54).

- **Formas paralelas.** Consiste en aplicar dos versiones diferentes de un instrumento a una misma muestra de sujetos y se correlacionan ambas medidas.

La función de prueba es:

$$r_{xy} = \frac{n \sum x_i y_j - \sum x_i \sum y_j}{\sqrt{[n \sum x_i^2 - (\sum x_i)^2][n \sum y_j^2 - (\sum y_j)^2]}}$$

Si ambas versiones son paralelas, deberían medir lo mismo y con igual precisión. De ser así, se asume la fiabilidad de las medidas con cualquiera de las versiones (Moisés Ríos, Valenzuela Oré, Pariona Cahuana, & Bendezú Prado, 2018, pág. 54).

Los coeficientes de fiabilidad de 0,90 y más son los mejores (incluyendo la posibilidad de la existencia de redundancia en los reactivos); los coeficientes de 0,80 y más son aceptables. Sin embargo, coeficiente de fiabilidad de 0,65 a menos son inaceptables (Reidl Martínez, 2013).

## CAPÍTULO IV

### PROCESAMIENTO Y ANÁLISIS DE DATOS

#### 4.1. Datos

Datos son las observaciones recolectadas como mediciones, respuestas de encuestas, etc. (Triola, 2009, pág. 4).

“(…) los datos son una representación simbólica que sirve de base para procesar y transmitir los atributos de esas entidades” (Gutiérrez Vargas, 2002, pág. 204).

Es el valor numérico, alfabético, lógico o de otra índole que representa a las modalidades de la variable. Es la respuesta a un estímulo o a las medidas realizadas.



**Figura N° 4. Relación entre variable y dato.**

Fuente: (Elaboración propia)

El dato “en bruto”, por sí mismo, no constituye información: es la materia prima. El procesamiento y análisis de los datos le otorgan la condición de información significativa.

VARIABLE	DATOS BRUTOS
Sexo	F-F-F-M-M-M-M-F-F-M-M-F-M-F-F-F-F-M-F-M-F

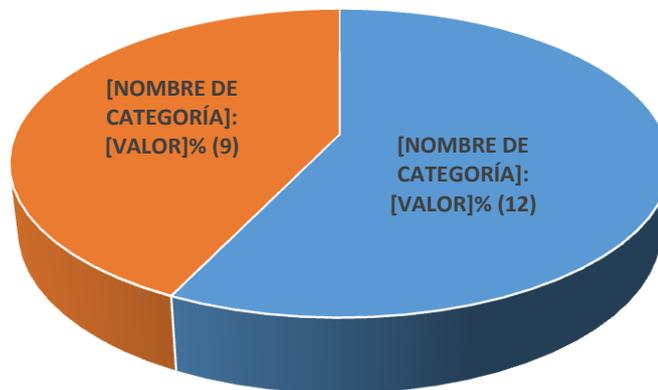


### PROCESAMIENTO Y ANÁLISIS DE DATOS

SEXO	PALOTEO	f <sub>i</sub>	h <sub>i</sub> %
Femenino	//// //	12	57,1
Masculino	////	9	42,9
<b>TOTAL</b>		<b>21</b>	<b>100,0</b>



### INFORMACIÓN SIGNIFICATIVA



**Figura N° 5. Dato e información significativa.**

Fuente: (Elaboración propia)

“Los datos tienen valor solo cuando son comprensibles; de otra forma, no son más que un montón de observaciones aleatorias” (BSA: The Software Alliance, 2015, pág. 10).

## **4.2. Procesamiento de datos**

Es el conjunto de operaciones que se realiza con una serie de datos para producir información significativa.

“Si bien el mundo es cada vez más autónomo, siguen siendo necesarias la curiosidad personal, las habilidades humanas y el trabajo intenso para liberar las respuestas del interior de los datos” (BSA: The Software Alliance, 2015, pág. 10).  
Consiste en el tratamiento de los datos o el procedimiento para traducir los datos en información significativa para la toma de decisiones informadas.

### **4.2.1. Etapas del procesamiento de datos**

#### **(a) Origen:**

Consiste en el acopio y evaluación de los datos iniciales. Un registro original de datos recibe el nombre de "documento fuente".

#### **(b) Entrada:**

Está referida a la construcción de la base de datos para el posterior procesamiento.

#### **(c) Procesamiento:**

Es el conjunto de operaciones necesarias para convertir los datos en información significativa.

#### **(d) Salida:**

Es el reporte o resumen de los datos en forma impresa o electrónica.

**(e) Distribución:**

Consiste en el envío de los registros de los datos de salida (reportes) a los tomadores de decisiones. Tiene como propósito la utilización de la información significativa.

**(f) Almacenamiento**

Está referida a la custodia de la base de datos para utilizarlos posteriormente como datos de entrada.

**4.3. Análisis de datos**

Es el uso simultáneo de estadísticas y algoritmos (...) para descubrir ideas, patrones y conexiones coherentes dentro de los datos (BSA: The Software Alliance, 2015, pág. 28).

Consiste en producir información relevante para la toma de decisiones informadas.

“Al examinar cuidadosamente los datos, las herramientas analíticas pueden atravesar el conjunto de datos desordenados para ayudar a los usuarios a descubrir pautas y tendencias nuevas, encontrar ideas inesperadas a partir de datos aparentemente inconexos y descubrir automáticamente relaciones interesantes desde el punto de vista estadístico” (BSA: The Software Alliance, 2015, pág. 10).

**4.3.1. Etapas del análisis de datos**

**(a) Inspección de los datos**

Consiste en evaluar la procedencia o fuente de los datos así como la naturaleza de los mismos.

“Una fuente de datos es el lugar principal de donde provienen los datos, por ejemplo, de una base de datos, una hoja de cálculo o un flujo de datos” (BSA: The Software Alliance, 2015, pág. 31).

### **(b) Limpieza de datos**

“La limpieza de los datos es el proceso de detección y corrección o eliminación de datos erróneos, incompletos o duplicados de una base de datos” (BSA: The Software Alliance, 2015, pág. 32).

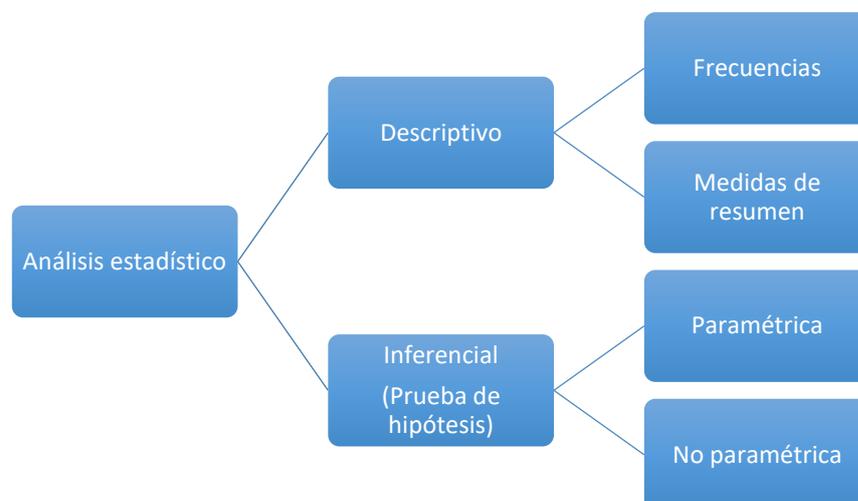
### **c) Transformación de datos**

Se utiliza con frecuencia para descubrir patrones escondidos o correlaciones desconocidas, y como una ayuda para la toma de decisiones” (BSA: The Software Alliance, 2015, pág. 28).

## **4.3.2. Importancia**

Predecir el futuro solía parecer inverosímil, pero ahora parece inevitable. Hoy, gracias a los análisis de datos modernos, realizamos predicciones confiables todo el tiempo (BSA: The Software Alliance, 2015, pág. 10).

## **4.3.3. Tipos de análisis de datos**



## **Figura N° 6. Tipos de análisis de datos.**

Fuente: (Elaboración propia)

### **(a) Análisis descriptivo**

Está referido al conteo, ordenación y clasificación de los datos mediante la representación tabular y gráfica de frecuencias o de medidas de resumen.

“Permiten simplificar la complejidad de los datos que intervienen en la distribución (...) No se hace uso del cálculo de probabilidades y únicamente se limita a realizar deducciones directamente a partir de los datos y parámetros obtenidos” (Borrego del Pino, 2008, pág. 2)

### **(b) Análisis inferencial**

Su propósito es generalizar resultados para una población a partir de una muestra seleccionada aleatoriamente, utilizando las leyes de la probabilidad.

“Plantea y resuelve el problema de establecer previsiones y conclusiones generales sobre una población a partir de los resultados obtenidos de una muestra. Los modelos estadísticos actúan de puente entre lo observado (muestra) y lo desconocido (población)” (Borrego del Pino, 2008, pág. 4).

## CAPÍTULO IV

### ANÁLISIS DESCRIPTIVO

#### 5.1. Frecuencias

Es la agrupación de datos en categorías y modalidades de las variables mediante números y porcentajes. A continuación se presenta algunos conceptos y procedimientos comunes para la presentación de frecuencias:

##### (a) Frecuencia absoluta simple

Es el número de veces que se repite un valor en un conjunto de valores. Se representa con  $f_i$  donde el subíndice representa cada uno de los valores. La suma de las frecuencias absolutas es igual al número total de datos, representado por N.

##### (b) Frecuencia absoluta acumulada

Es igual a la suma de las frecuencias absolutas simples inferiores o iguales a las frecuencias acumuladas que se desea encontrar. Se representa por  $F_i$ .

$$F_i = \sum f_i$$

##### (c) Frecuencia relativa simple

Es el cociente entre la frecuencia absoluta simple de un determinado valor y el número total de datos. Se representa con  $h_i$ . La suma de las frecuencias relativas es igual a 1.

$$h_i = \frac{f_i}{N}$$

#### **(d) Frecuencia relativa acumulada ( $H_i$ )**

Cociente entre la frecuencia absoluta acumulada y el número total de datos. Se calcula también sumando las frecuencias relativas simples inferiores o iguales a las frecuencias relativas acumuladas que se desea encontrar. Se representa por  $H_i$ .

$$H_i = \frac{F_i}{N} = \sum h_i$$

### **5.2. Medidas de tendencia central**

Son medidas estadísticas que resumen en un solo valor un conjunto de datos. Representan un centro en torno al cual están ubicados el conjunto de datos (Moisés Ríos, Chachaima Pumasoncco, Leguía Franco, Ochatoma Palomino, & Bendezú Prado, 2018).

#### **5.2.1. Media aritmética**

“(…) es la suma de todos los datos correspondientes a una muestra, dividida entre el total de los datos” (Valderrama M., 2011, pág. 27).

Es el promedio aritmético en una distribución de datos. Es la medida más usada cuando la serie no presenta valores extremos que distorsionen su valor.

#### **Media aritmética para datos no agrupados:**

##### **A. Para la muestra:**

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

##### **B. Para la población:**

$$\mu = \frac{\sum_{i=1}^n x_i}{N}$$

Donde:

- $\Sigma$  : Suma de un conjunto de valores.
- $x_i$  : Datos individuales de la variable.
- $\bar{X}$  : Media aritmética para la muestra.
- $\mu$  : Media aritmética para la población.
- $n$  : Número de datos en la muestra.
- $N$  : Número de datos en la población.

**Media aritmética para datos agrupados:**

$$\bar{X} = \frac{\sum_{i=1}^n f_i x_i}{n}$$

Donde:

- $f_i$  : Frecuencia de clase i.
- $x_i$  : Marca de clase i.

**5.2.2. Mediana**

Valor de la variable que ocupa la posición central en el conjunto de datos ordenado ascendentemente. Por tanto, equidista de ambos extremos de la distribución: el 50% de las observaciones tiene valores iguales o inferiores a la mediana y el otro 50% tiene valores iguales o superiores a la mediana (Moisés Ríos, Chachaima Pumasoncco, Leguía Franco, Ochatoma Palomino, & Bendezú Prado, 2018).

**Mediana en datos no agrupados:**

**A. Cuando el tamaño del conjunto de datos es impar.** La mediana es el valor central de la distribución. Se obtiene sumando 1 a la frecuencia total y dividiendo el resultado por dos:

$$Md = \frac{x_n + 1}{2}$$

**B. Cuando el tamaño del conjunto de datos es par.** La mediana está representada por la media aritmética de los dos valores centrales.

$$Md = \frac{\frac{x_n}{2} + \frac{x_{(n+2)}}{2}}{2}$$

### **Mediana en datos agrupados:**

En este caso, la mediana se obtendrá mediante la fórmula siguiente:

$$Md = L_i + \left[ \frac{\left(\frac{n}{2}\right) - F_{(ant)}}{f_i} \right] \times A$$

Donde:

$n/2$  : Posición aproximada de la mediana en la distribución de acuerdo al número de datos que se disponga.

$L_i$  : Límite inferior real de la clase mediana.

$F_{(ant)}$  : Frecuencia acumulada de la clase anterior a la clase mediana  $i$  ( $F_{i-1}$ ).

$f_i$  : Frecuencia absoluta de la clase mediana.

$A$  : Amplitud real del intervalo.

### **5.2.3. Moda**

Valor que se repite con más frecuencia en una serie de datos. Puede que no exista o bien que existan varios valores modales.

**Moda en datos no agrupados.** Es el **valor** que tiene **mayor frecuencia absoluta**.

### **Moda en datos agrupados:**

$$Mo = L_i + \left[ \frac{\Delta_1}{\Delta_1 + \Delta_2} \right] \times A$$

Donde:

$L_i$  : Límite inferior real de la clase modal.

$\Delta_1$  : Diferencia absoluta entre la frecuencia de la clase modal y la de la clase anterior ( $f_i - f_{i-1}$ ).

$\Delta_2$  : Diferencia absoluta entre la frecuencia de la clase modal y la de la clase posterior ( $f_i - f_{i+1}$ ).

A : Amplitud real del intervalo.

### 5.3. Medidas de dispersión

Medidas que indican cuan dispersos o alejados están los datos respecto de la media aritmética. Sirven como indicador de la variabilidad de los datos (Moisés Ríos, Chachaima Pumasoncco, Leguía Franco, Ochatoma Palomino, & Bendezú Prado, 2018).

#### 5.3.1. Varianza

Es el promedio de la suma de los cuadrados de las desviaciones de la variable respecto a su media. Se obtiene de acuerdo a lo siguiente:

#### Varianza para datos no agrupados:

##### A. Para la muestra:

$$S^2 = \frac{\sum_{i=1}^n x_i^2 - n(\bar{X})^2}{n - 1}$$

##### B. Para la población:

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2 - N(\bar{X})^2}{N}$$

Donde:

$s^2$  : Varianza muestral.

$\sigma^2$  : Varianza poblacional.

$x_i$  : Valor de cada dato.

- n : Tamaño de la muestra (número de datos).
- $\bar{X}$  : Media aritmética para datos simples.
- N : Tamaño de la población.

**Varianza para datos agrupados:**

**A. Para la muestra:**

$$S^2 = \frac{\sum_{i=1}^n x_i^2 f_i - n(\bar{X})^2}{n - 1}$$

**B. Para la población:**

$$\sigma^2 = \frac{\sum_{i=1}^n x_i^2 f_i - N(\bar{X})^2}{N}$$

Donde:

- $x_i$  : Marca de clase de cada intervalo (punto medio del intervalo).
- $f_i$  : Frecuencia absoluta de cada intervalo.
- n : Tamaño de la muestra (número de datos).
- $\bar{X}$  : Media aritmética para datos simples.
- N : Tamaño de la población.

**Nota:** La varianza es una medida teórica y **carece de interpretación práctica**. Asimismo, sus valores **siempre** son positivos.

**5.3.2. Desviación estándar (típica)**

Es la raíz cuadrada de la varianza. Esta medida tiene interpretación práctica.

**A. Para la muestra:**

$$S = \sqrt{S^2}$$

**B. Para la población:**

$$\sigma = \sqrt{\sigma^2}$$

Donde:

- $S^2$  : Varianza muestral.
- $\sigma^2$  : Varianza poblacional.

### 5.3.3. Coeficiente de variación (CV%)

Es el cociente que resulta de la división de la desviación estándar con la media aritmética. Se multiplica por 100 para obtener un porcentaje.

#### A. Para la muestra:

$$CV\% = \frac{S}{|\bar{X}|} \times 100$$

#### B. Para la población:

$$CV\% = \frac{\sigma}{|\mu|} \times 100$$

Donde:

S : Desviación estándar muestral.

$|\bar{X}|$  : Valor absoluto de la media aritmética muestral.

$\sigma$  : Desviación estándar poblacional.

$|\mu|$  : Valor absoluto de la media aritmética poblacional.

Es un indicador de la dispersión relativa de los datos y tiene mayor utilidad cuando se comparan dos o más distribuciones pertenecientes a poblaciones distintas (Moisés Ríos, Chachaima Pumasoncco, Leguía Franco, Ochatoma Palomino, & Bendezú Prado, 2018).

#### Interpretación:

- $CV\% < 30\%$ : La media aritmética es una medida representativa del conjunto de datos (la distribución de los datos tienden a la homogeneidad).
- $CV\% \geq 30\%$ : La media aritmética no es una medida representativa del conjunto de datos (la distribución de los datos tienden a la heterogeneidad).

### 5.4. Medidas de posición

Son valores que dividen la distribución de los datos en partes iguales (intervalos). Los más utilizados son los cuartiles, quintiles, deciles y percentiles.

### 5.4.1. Cuartiles

Dividen a la distribución en 4 partes. Es decir, existen 3 cuartiles que dividen a una serie en 4 partes iguales.

#### Para datos no agrupados:

##### Primer cuartil ( $Q_1$ ):

Si n es par:

$$Q_1 = \frac{(1)(n)}{4}$$

Si n es impar:

$$Q_1 = \frac{(1)(n+1)}{4}$$

##### Tercer cuartil ( $Q_3$ ):

Si n es par:

$$Q_3 = \frac{(3)(n)}{4}$$

Si n es impar:

$$Q_3 = \frac{(3)(n+1)}{4}$$

#### Para datos agrupados:

$$Q_i = L_i + \left[ \frac{\frac{(i)\sum f}{4} - F_{ant}}{f_i} \right] \times A$$

Donde:

$L_i$  : Límite exacto inferior de la clase que contiene el número de cuartil objeto de cálculo.

n : Número total de datos.

$F_i$  : Frecuencia acumulada de la clase que contiene el número de cuartil objeto de cálculo: se calcula como la frecuencia acumulada de clase inmediatamente superior al valor de  $\frac{(i)\sum f}{4}$ .

$F_{ant}$  : Frecuencia acumulada de la clase anterior a la clase que contiene el número de cuartil objeto de cálculo ( $F_{i-1}$ ).

$f_i$  : Frecuencia absoluta de la clase que contiene el número de cuartil objeto de cálculo.

$A$  : Amplitud real del intervalo.

De acuerdo a los aspectos detallados, las fórmulas estadísticas del primer cuartil ( $Q_1$ ) y tercer cuartil ( $Q_3$ ) serían las siguientes:

$$Q_1 = L_i + \left[ \frac{\left[ \frac{(1)\sum f}{4} - F_{ant} \right]}{f_i} \right] \times A$$

$$Q_3 = L_i + \left[ \frac{\left[ \frac{(3)\sum f}{4} - F_{ant} \right]}{f_i} \right] \times A$$

#### 5.4.2. Deciles

Dividen a la distribución en 10 partes. Es decir, existen 9 deciles que dividen a una serie en 10 partes iguales.

##### Para datos no agrupados:

Cuando el decil es par:

$$D_i = \frac{(D)(n)}{10}$$

Cuando el decil es impar:

$$D_i = \frac{(D)(n+1)}{10}$$

##### Para datos agrupados:

$$D_i = L_i + \left[ \frac{\left[ \frac{(i)\sum f}{10} - F_{ant} \right]}{f_i} \right] \times A$$

Donde:

$L_i$  : Límite exacto inferior de la clase que contiene el número de decil objeto de cálculo.

$n$  : Número total de datos.

- $F_i$  : Frecuencia acumulada de la clase que contiene el número de decil objeto de cálculo: se calcula como la frecuencia acumulada de clase inmediatamente superior a  $\frac{(i)\Sigma f}{10}$ .
- $F_{ant}$  : Frecuencia acumulada de la clase anterior a la clase que contiene el número de decil objeto de cálculo ( $F_{i-1}$ ).
- $f_i$  : Frecuencia absoluta de la clase que contiene el número de decil objeto de cálculo.
- $A$  : Amplitud real del intervalo.

### 5.4.3. Percentiles

“Se conoce como percentil, al puntaje de una distribución que deja por debajo de él un cierto porcentaje de los casos de dicha distribución” (Valderrama M., 2011, pág. 38).

Dividen a la distribución en 100 partes. Es decir, existen 99 percentiles que dividen a una serie en 100 partes iguales.

#### Para datos no agrupados:

Cuando el percentil es par:

$$P_i = \frac{(P)(n)}{100}$$

Cuando el percentil es impar:

$$P_i = \frac{(P)(n+1)}{100}$$

#### Para datos agrupados:

$$P_i = L_i + \left[ \frac{\frac{(i)\Sigma f}{100} - F_{ant}}{f_i} \right] \times A$$

Donde:

$L_i$  : Límite exacto inferior de la clase que contiene el número de percentil objeto de cálculo.

$n$  : Número total de datos.

$F_i$  : Frecuencia acumulada de la clase que contiene el número de percentil objeto de cálculo: se calcula como la frecuencia acumulada de clase inmediatamente superior a  $\frac{(i)\Sigma f}{100}$ .

$F_{ant}$  : Frecuencia acumulada de la clase anterior a la clase que contiene el número de percentil objeto de cálculo ( $F_{i-1}$ ).

$f_i$  : Frecuencia absoluta de la clase que contiene el número de percentil objeto de cálculo.

$A$  : Amplitud real del intervalo.

## 5.5. Medidas de forma

Permiten identificar el tipo de distribución del conjunto de datos.

### 5.5.1. Asimetría

Permite identificar hacia qué sector se acumulan los datos de una distribución. Se obtiene mediante las siguientes fórmulas:

$$AS = \frac{\bar{X} - Mo}{S}$$

Donde:

$\bar{X}$  : Media aritmética.

$Mo$  : Moda.

$S$  : Desviación estándar.

$$AS = \frac{Q_3 + Q_1 - 2Q_2}{Q_3 - Q_1}$$

Donde:

$Q_1$  : Primer cuartil.

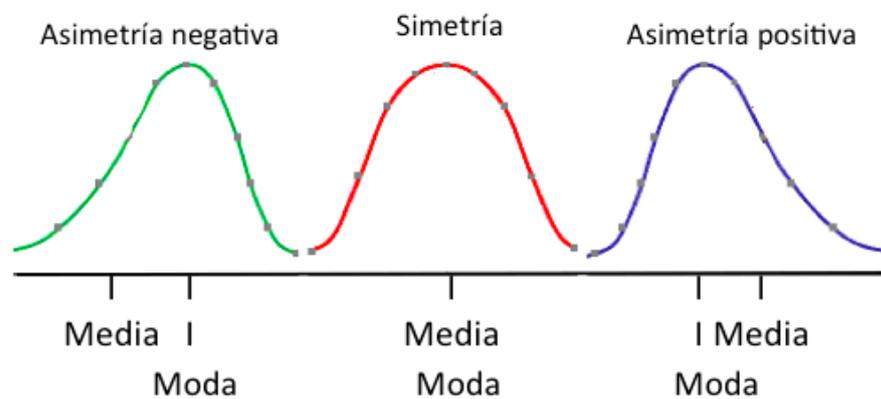
$Q_2$  : Segundo cuartil.

$Q_3$  : Tercer cuartil.

**Interpretación:**

- $As = 0$ : La distribución es simétrica (existe el mismo número de valores a la izquierda y derecha de la media. En este caso, la media, mediana y moda coinciden en un mismo punto).
- $As > 0$ : La distribución es asimétrica positiva (la cola de la distribución se alarga para valores mayores que la media: existen más valores superiores a la media).
- $As < 0$ : La distribución es asimétrica negativa (la cola de la distribución se alarga para valores menores que la media: existen más valores inferiores a la media).

**Figura N° 7.**  
*Tipología de asimetría.*



Fuente:  
(Elaboración propia)

**En la práctica, existe una interpretación más tolerante** (Moisés Ríos, Chachaima Pumasoncco, Leguía Franco, Ochatoma Palomino, & Bendezú Prado, 2018):

$AS > 0,5$  : Distribución asimétrica positiva.

$AS \pm 0,5$  : Distribución simétrica.

AS < -0,5 : Distribución asimétrica negativa.

### 5.5.2. Coeficiente de curtosis o apuntamiento (K)

Indica el grado de aplastamiento (deformación vertical) que tiene la distribución.

$$k = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

Donde:

P<sub>90</sub> : Percentil 90.

P<sub>10</sub> : Percentil 10.

Q<sub>1</sub> : Primer cuartil.

Q<sub>3</sub> : Tercer cuartil.

#### Interpretación:

- k = 0,263: La distribución de frecuencias es mesocúrtica: tiene igual apuntamiento que la normal o campana de Gauss (concentración media de datos alrededor de los valores centrales de la variable).
- k < 0,263: La distribución de frecuencias es platicúrtica (reducida concentración de datos alrededor de los valores centrales de la variable).
- k > 0,263: La distribución de frecuencias es leptocúrtica (alta concentración de datos alrededor de los valores centrales de la variable).

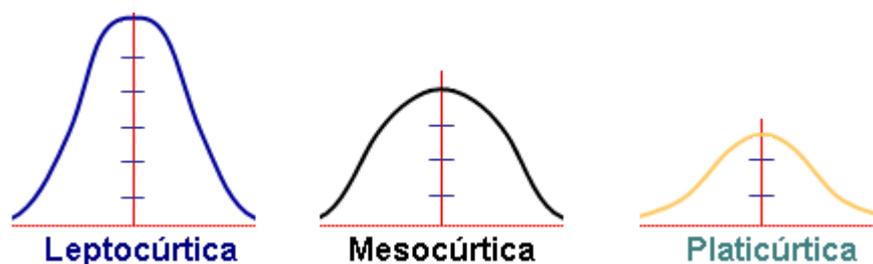


Figura N° 8. Tipo de apuntamiento.

Fuente: (Elaboración propia)

**En la práctica, existe una interpretación más tolerante** (Moisés Ríos, Chachaima Pumasoncco, Leguía Franco, Ochatoma Palomino, & Bendezú Prado, 2018):

$K > 0,5$  : Distribución leptocúrtica.

$K \pm 0,5$  : Distribución mesocúrtica.

$K < -0,5$  : Distribución platicúrtica.

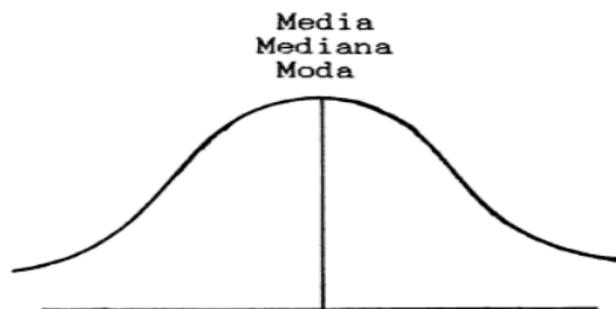
## CAPÍTULO V

### ANÁLISIS INFERENCIAL

#### 5.1. Análisis paramétrico

Se basan en el supuesto de una distribución normal en la población estudiada. La distribución de probabilidad normal y la curva normal que la representa, tienen las siguientes características (Moisés Ríos, Valenzuela Oré, Pariona Cahuana, & Bendezú Prado, 2018):

- a) **Simétrica.** Con un solo pico en el centro de la distribución donde la media, mediana y moda son iguales (la curva normal desciende suavemente en ambas direcciones a partir del valor central). Así, la mitad del área bajo la curva se encuentra a la derecha de este punto central y la otra mitad a la izquierda del punto en referencia.
- b) **Mesocúrtica.** Es una distribución con una curtosis media.
- c) **Asintótica.** Las colas de la curva se acercan cada vez más al eje X sin llegar a tocarlo. Es decir, las colas de la curva se extienden de manera indefinida en ambas direcciones.



**Figura N° 9. Curva normal.**

Fuente: (Moisés Ríos, Valenzuela Oré, Pariona Cahuana, & Bendezú Prado, 2018)

Existe una multiplicidad de variables aleatorias continuas que siguen una distribución normal o aproximadamente normal.

<b>Pruebas</b>	<b>Una muestra</b>	<b>Variables</b>
Prueba de Ji Cuadrada de Bartlett	Prueba que la varianza de una población normal es igual a un valor específico.	V: Intervalo
T de Student	Es una prueba referida a la proporción poblacional. La hipótesis se refiere al parámetro P, la proporción de individuos de la población con la característica de interés ( $n < 30$ ).	V: Intervalo
Z estándar	Es una prueba referida a la proporción poblacional. La hipótesis se refiere al parámetro P, la proporción de individuos de la población con la característica de interés ( $n \geq 30$ ).	V: Intervalo
<b>Pruebas</b>	<b>Dos muestras independientes</b>	<b>Variables</b>
F	Es una prueba referida a la comparación de dos varianzas poblacionales (homocedasticidad u heterocedasticidad).	VI: Dicotómica VD: Intervalo
T de Student	Es una prueba referida a la comparación de dos medias poblacionales ( $n_x + n_y - 2 < 30$ ) cuando las varianzas poblacionales son desconocidas: a) Aproximadamente iguales o b) diferentes.	VI: Dicotómica VD: Intervalo
Z estándar	Es una prueba referida a la comparación de dos medias poblacionales ( $n_x + n_y - 2 \geq 30$ ) cuando las varianzas poblacionales son conocidas.	VI: Dicotómica VD: Intervalo
<b>Pruebas</b>	<b>2 muestras relacionadas</b>	<b>Variables</b>
T de Student pares	Está diseñada para la comparación de grupos relacionados (dependientes). Por tanto, los requisitos para su aplicación (distribución normal y homocedasticidad de varianzas) exceptúa la independencia de los grupos.	VI: Dicotómica VD: Intervalo
Coefficiente de correlación r de Pearson	Es una medida de asociación lineal especialmente apropiada para estudiar la relación entre variables de intervalo o razón (no indica necesariamente causación). El término correlación denota covariación o interdependencia entre las dos variables (variación conjunta)	V: Intervalo
Coefficiente de	Establece la relación entre una variable	VI: Dicotómica

Correlación biserial	dicotomizada a partir de una variable normalmente distribuida y una variable con escala de intervalo.	VD: Intervalo
Regresión lineal simple	Tiene como propósito estudiar la naturaleza y forma de la asociación entre dos variables, siempre que dicha relación pueda ser expresada matemáticamente por la ecuación de la línea recta	V: Intervalo
<b>Pruebas</b>	<b>K muestras independientes</b>	<b>Variables</b>
Análisis de varianza (ANOVA)	Es la generalización de la prueba t de Student para grupos independientes. Se utiliza generalmente para probar diferencias en las medias procedentes de más de dos grupos.	VI: Politémica VD: Intervalo

### 5.1.1. Ventajas

- Mayor poder de eficiencia.
- Mayor sensibilidad a la variabilidad de los datos recolectados.
- Menor posibilidad de errores.
- Robustas (proporcionan estimaciones probabilísticas más exactas).

### 5.1.2. Desventajas

- Son más difíciles de calcular.
- Limitaciones en ciertos tipos de datos (cualitativos y cuasi-cuantitativos).

## 5.2. ANÁLISIS NO PARAMÉTRICO

Se denominan pruebas no paramétricas a aquellas que no presuponen una distribución de probabilidad conocida para los datos, por ello se denominan de distribución libre (distribución free). A pesar de basarse en determinadas suposiciones, no parten de la base de que los datos analizados adoptan una distribución normal (Moisés Ríos, Valenzuela Oré, Pariona Cahuana, & Bendezú Prado, 2018).

Por tanto, se utilizan siempre con variables cualitativas (nominales) y cuasi-cuantitativas (ordinales); y a veces con variables cuantitativas: solo si no cumplen con los supuestos de normalidad.

La literatura bibliográfica reporta la siguiente clasificación de pruebas estadísticas no paramétricas (Berlanga Silvente & Rubio Hurtado, 2012, págs. 103-105):

<b>Pruebas</b>	<b>Una muestra</b>	<b>Variables</b>
Prueba de Chi-cuadrado de Pearson	Es una prueba de bondad de ajuste, que permite averiguar si la distribución empírica de una variable categórica se ajusta o no (se parece o no) a una determinada distribución teórica (uniforme, binomial, multinomial, etc.).	V: Nominal
Prueba Binomial	Es una prueba de bondad de ajuste, que permite averiguar si una variable dicotómica sigue o no un determinado modelo de probabilidad. Permite contrastar la hipótesis de que la proporción observada de aciertos se ajusta a la proporción teórica de una distribución binomial (lo cual se traduce en la posibilidad de contrastar hipótesis sobre proporciones y sobre cuartiles).	V: Nominal
Prueba de Rachas	Es una prueba de independencia o de aleatoriedad que permite determinar si el número de rachas (R) observado en una determinada muestra de tamaño n es lo suficientemente grande o lo suficientemente pequeño para poder rechazar la hipótesis de independencia (o aleatoriedad) entre las observaciones. Una racha es una secuencia de observaciones de un mismo atributo o cualidad. Una serie de datos en los que hay muchas o pocas rachas permite concluir que estas no han ocurrido por azar.	V: Nominal
Prueba de Kolmogorov-Smirnov (K- S)	Es una prueba de bondad de ajuste, que sirve para contrastar la hipótesis nula de que la distribución de una variable se ajusta a una determinada distribución teórica de probabilidad que puede ser con tendencia a la normal, a la de Poisson o exponencial	V: Ordinal/Intervalo
<b>Pruebas</b>	<b>Dos muestras relacionadas</b>	<b>Variables</b>

Prueba de McNemar	Sirve para contrastar hipótesis sobre igualdad de proporciones. Se usa cuando hay una situación en la que las medidas de cada sujeto se repiten, por lo que la respuesta de cada uno de ellos se obtiene dos veces: una vez antes y otra después de que ocurra un evento específico	VI: Dicotómica VD: Nominal
Prueba de los Signos	Permite contrastar la hipótesis de igualdad entre dos medianas poblacionales. Puede ser usada para saber si una variable tiende a ser mayor que otra. También es útil para probar la tendencia que sigue una serie de variables ordinales positivas, o para una valoración rápida de un estudio exploratorio.	VI: Dicotómica VD: Ordinal/intervalo
Prueba de Wilcoxon	Permite contrastar la hipótesis de igualdad entre dos medianas poblacionales. Paralela a la prueba paramétrica de contraste t para muestras relacionadas.	VI: Dicotómica VD: Ordinal/intervalo
<b>Pruebas</b>	<b>K-muestras relacionadas</b>	<b>Variables</b>
Prueba de Friedman	Es una extensión de la prueba de Wilcoxon para incluir datos registrados en más de dos periodos de tiempo o grupos de tres o más sujetos pareados, con un sujeto de cada grupo que ha sido asignado aleatoriamente a una de las tres o más condiciones. La prueba examina los rangos de los datos generados en cada periodo de tiempo para determinar si las variables comparten la misma distribución continua de su origen.	VI: Politómica VD: Ordinal/Intervalo
Prueba de Cochran	Esta prueba es idéntica a la prueba de Friedman, pero se aplica cuando todas las respuestas son binarias. La Q de Cochran prueba la hipótesis de que varias variables dicotómicas que están relacionadas entre sí, tienen el mismo promedio. En observaciones múltiples las variables son medidas en el mismo individuo o en individuos pareados. Tiene la ventaja de examinar cambios en las variables categóricas.	VI: Dicotómica VD: Nominal
Coefficiente de concordancia de	Tiene las mismas indicaciones que la prueba de Friedman, aunque su uso en	VI: Politómica VD: Ordinal

W de Kendall	investigación ha sido, principalmente, para conocer la concordancia entre rangos, más que para probar que existe una diferencia entre las medianas.	/Intervalo
<b>Pruebas</b>	<b>Dos muestras independientes</b>	<b>Variables</b>
Prueba U de Mann-Whitney	Es equivalente a la prueba de suma de rangos de Wilcoxon y a la prueba de dos grupos de Kruskal-Wallis. Es la alternativa no paramétrica a la comparación de dos promedios independientes a través de la t de Student	VI: Dicotómica VD: Ordinal/Intervalo
Prueba de Kolmogorov-Smirnov	Sirve para contrastar la hipótesis de que dos muestras proceden de la misma población. Para ello, compara las funciones de distribución (funciones de probabilidad acumuladas) de ambas muestras.	VI: Dicotómica VD: Ordinal/Intervalo
Prueba de Rachas de Wald-Wolfowitz	Contrasta si dos muestras con datos independientes proceden de poblaciones con la misma distribución.	VI: Dicotómica VD: Ordinal/Intervalo
Prueba de reacciones extremas de Moses	Sirve para estudiar si existe diferencia en el grado de dispersión o variabilidad de dos distribuciones. Esta prueba se centra en la distribución del grupo de control y es una medida para saber cuántos valores extremos del grupo experimental influyen en la distribución cuando se combinan con el grupo de control.	VI: Dicotómica VD: Ordinal/Intervalo
<b>Pruebas</b>	<b>K-muestras independientes</b>	<b>Variables</b>
Prueba de la Mediana	Contrasta diferencias entre dos o más grupos en relación con su mediana, bien porque no cumplen las condiciones de normalidad para usar el promedio como medida de tendencia central, bien porque la variable es cuantitativa discreta. Es similar a la prueba Chi-cuadrado.	VI: Plicotómica VD: Ordinal/Intervalo
Prueba de Jonckheere-Terpstra	Es más potente que sus homónimas, la prueba de Kruskal-Wallis y la de la mediana, cuando existe una ordenación a priori (ascendente o descendente) de las K poblaciones de las que se extraen las muestras.	VI: Politómica VD: Ordinal/Intervalo
Prueba H de Kruskal-Wallis	Es una extensión de la de U de Mann-Whitney y representa una excelente alternativa al ANOVA de un factor completamente aleatorizado.	VI: Politómica VD: Ordinal/Intervalo

### 5.2.1. Ventajas

- Pueden ser aplicados a una amplia variedad de situaciones porque sus requerimientos son más flexibles.
- Pueden ser aplicados en datos no numéricos.
- Más fáciles de calcular y, por tanto, más fáciles de entender y aplicar.

### **5.2.2. Desventajas**

- Mayor tendencia a perder información porque los datos numéricos son frecuentemente reducidos a datos cualitativos o cuasi-cuantitativos.
- Menor poder de eficiencia: se necesita evidencia más fuerte.

## CONCLUSIONES

En este capítulo, se presentan las conclusiones generadas del análisis de los resultados de la investigación bibliográfica. Con el objeto de organizar el cuerpo de conclusiones, se agrupan atendiendo a los objetivos generales y específicos, a saber:

En cuanto al objetivo general de discutir los alcances de la medición en el procesamiento y análisis de datos, este objetivo fue alcanzado ya que:

En el campo de la investigación cuantitativa, la medición es una condición necesaria para el tratamiento de las variables. Con la medición de las diferentes propiedades del objeto de estudio se logra una mayor precisión en la comunicación de los conocimientos adquiridos, aunque la pretensión de alcanzar una certeza absoluta acerca del valor real de lo que se pretende medir no sea accesible.

En cuanto a los objetivos específicos:

- Exponer las definiciones sobre medición, el mismo fue logrado porque se presentan diferentes puntos de vista de los autores bibliográficos. El análisis de estas definiciones permitió identificar aspectos comunes como la asignación de números a las propiedades de los objetos de estudio para ser mensuradas. Como aporte personal, se concluye que la medición es el proceso de aproximación al valor real de la propiedad del objeto de estudio con algún grado de precisión.
- En cuanto a la clasificación de las variables y sus escalas de medición, se llegó a la conclusión que existen diferentes formas de clasificar las variables atendiendo criterios convencionales. En lo que respecta a la

medición, la clasificación más adecuada es de acuerdo a la naturaleza de las propiedades del objeto de estudio: cualitativas, cuasi-cuantitativas y cuantitativas. Estas, a su vez, guardan correlato con las escalas de medición: nominal para las cualitativas, ordinal para las cuasi-cuantitativas e intervalo y razón para las cuantitativas. El estudio de las escalas de medición es relevante en el procesamiento y análisis de datos, porque define la presentación de la información y el tipo de prueba de hipótesis.

- Referente al objetivo explicar los principios de la medición, fue efectivizado porque la utilización de “instrumentos de medición” demanda el cumplimiento de dos indicadores de calidad: (a) validez para referirse a la exactitud de la medida y (b) confiabilidad como criterio de precisión de la medida. Un instrumento de medición puede ser válido pero no confiable o confiable pero no válido: es necesario que sea válido y confiable a la vez.
- Acerca de los alcances del procesamiento y análisis de datos, se concluye que el dato “en bruto”, por sí mismo, no constituye información: es la materia prima. El procesamiento y análisis de los datos le otorgan la condición de información significativa. El análisis descriptivo está referido al conteo, ordenación y clasificación de los datos mediante la representación tabular y gráfica de frecuencias o de medidas de resumen. El análisis inferencial tienen como propósito generalizar resultados para una población a partir de una muestra seleccionada aleatoriamente, utilizando las leyes de la probabilidad.

- Sobre los criterios para la aplicación de pruebas paramétricas se concluye que corresponde exclusivamente a variables cuantitativas con una distribución normal: simétrica, mesocúrtica y asintótica.
- En cuanto a los criterios para la aplicación de pruebas no paramétricas, se llegó a la conclusión que se utilizan siempre con variables cualitativas (nominales) y cuasi-cuantitativas (ordinales); y a veces con variables cuantitativas: solo si no cumplen con los supuestos de normalidad.

## BIBLIOGRAFÍA

- Badiou, A. (1972). *El concepto de modelo: introducción a una metodología materialista de las matemáticas*. Buenos Aires, Argentina: Siglo XXI.
- Barbero García, M. I., Vila Abad, E., & Holgado Tello, F. (2015). *Psicometría*. Madrid, España: Sanz y Tores, S.L.
- Barrios, M., & Cosculluela, Á. (2013). Fiabilidad. En J. Meneses, M. Barrios, A. Bonillo, Á. Cosculluela, L. Lozano, J. Turbany, & S. Valero, *Psicometría* (págs. 75-137). Barcelona, España: Eureca Media S.L.
- Berlanga Silvente, V., & Rubio Hurtado, M. J. (2012). Clasificación de pruebas no paramétricas. Cómo aplicarlas en SPSS. *REIRE*, 5(2), 101-113.
- Blair, R. C., & Taylor, R. A. (2008). *Bioestadística*. Juárez, México: Pearson Educación.
- Borrego del Pino, S. (2008). Estadística descriptiva e inferencial. *Innovación y experiencias educativas*(13), 1-12. Recuperado el 20 de Julio de 2018, de [https://archivos.csif.es/archivos/andalucia/ensenanza/revistas/csicsif/revista/pdf/Numero\\_13/SILVIA\\_BORREGO\\_2.pdf](https://archivos.csif.es/archivos/andalucia/ensenanza/revistas/csicsif/revista/pdf/Numero_13/SILVIA_BORREGO_2.pdf)
- BSA: The Software Alliance. (2015). *¿Por qué son tan importantes los datos?* (Primera ed.). España, España: BSA.
- Bunge, M. (1997). *La investigación científica: Su estrategia y su filosofía* (Cuarta ed.). España: Ariel.
- Cazau, P. (2016). *Fundamentos de estadística*. Buenos Aires: UBA.
- Furr, R. (2011). *Scale construction and psychometrics for social and personality psychology*. California: SAGE.
- Gutiérrez Vargas, M. E. (2002). El aprendizaje de la ciencia y de la información científica en la educación superior. *Anales de documentación*(5), 197-212.
- Kerlinger, F. N., & Lee, H. B. (2002). *Investigación del comportamiento: métodos de investigación en ciencias sociales* (Cuarta ed.). México, D.F., México: McGraw Hill/Interamericana.
- Moisés Ríos, B. W., Ango Bedriñana, J., Palomino Vargas, V. A., & Feria Macizo, E. E. (2018). *Diseño del proyecto de investigación científica* (Primera ed.). Lima: San Marcos.

- Moisés Ríos, B. W., Valenzuela Oré, F., Pariona Cahuana, B., & Bendezú Prado, J. (2018). *Ejecución del proyecto de investigación científica*. Lima, Perú: San Marcos.
- Moisés Ríos, B., Chachaima Pumasoncco, N., Leguía Franco, W., Ochatoma Palomino, M., & Bendezú Prado, J. (2018). *Epidemiología y estadística más elemental*. Lima, Perú: San Marco.
- Nunnally, J., & Bernstein, I. (1995). *Teoría psicométrica* (Tercera ed.). (J. Velázquez, Trad.) México D.F.: McGraw Hill.
- Posada Hernández, G. J. (2016). *Elementos básicos de estadística descriptiva para el análisis de datos*. Medellín, Colombia: Funlam.
- Reidl Martínez, L. M. (12 de Mayo de 2013). Confiabilidad en la medición. *Investigación en Educación Médica*. Recuperado el 12 de Mayo de 2017, de <http://riem.facmed.unam.mx/node/76>
- Triola, M. F. (2009). *Estadística* (Décima ed.). México: Pearson Educación.
- Valderrama M., G. A. (2011). *Estadísticas aplicadas en psicología, ciencias sociales y educación*. Panamá, Panamá: Instituto Panameño de Terapia Cognitivo-Conductual.