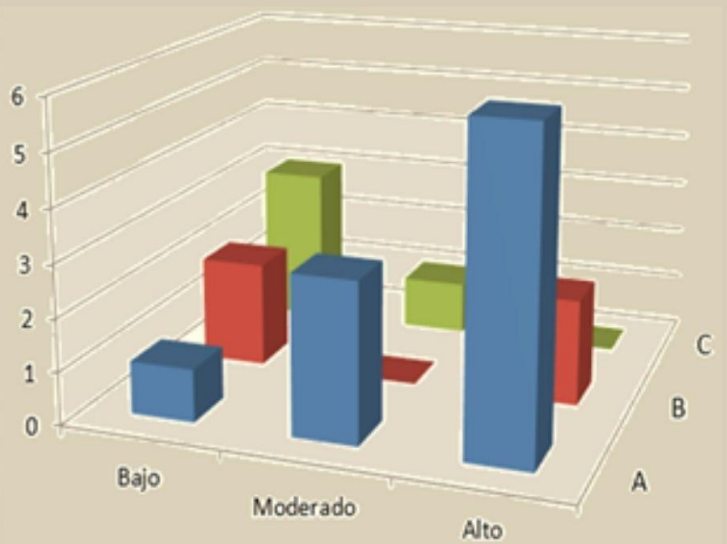
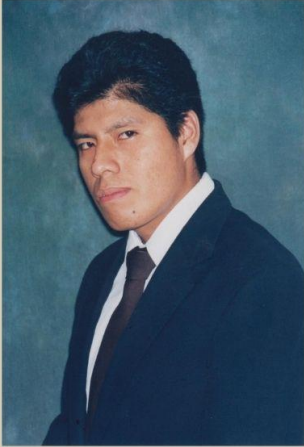


ESTADÍSTICA DESCRIPTIVA PARA TRABAJOS DE INVESTIGACIÓN

Presentación e interpretación
de los resultados



Jorge Sucasaire Pilco



Jorge Sucasaire Pilco

Nacido en la ciudad de Lima. Es docente egresado de la Facultad de Educación de la Universidad Nacional Mayor de San Marcos con especialidad en Matemática y Física. Maestro y Doctor en Administración de la Educación por la Universidad César Vallejo. Ha preparado y dictado cursos de Matemática y Estadística en el Instituto Superior Tecnológico Paul Müller, Instituto de Educación Superior Avansys (Cibertec), Instituto de Formación Bancaria IFB CERTUS, Instituto Superior IDAT, Centro Preuniversitario de la Universidad Autónoma del Perú, Universidad Privada César Vallejo, Universidad Tecnológica del Perú, Universidad Privada del Norte y la Universidad Le Cordon Bleu.

Jorge Sucasaire Pilco

**ESTADÍSTICA DESCRIPTIVA PARA
TRABAJOS DE INVESTIGACIÓN**

**Presentación e interpretación de los
resultados**

©Jorge Sucasaire Pilco
Primera edición digital – marzo 2021

Editado por:
Jorge Sucasaire Pilco
Jr. Los Lirios 380, Urb. La Primavera – El Agustino
E-mail: jorgesuca1982@gmail.com

Libro electrónico disponible en:
<https://repositorio.concytec.gob.pe/>

Hecho el Depósito Legal en la Biblioteca Nacional
del Perú N° 2021-02593

ISBN 978-612-00-6118-3

ESTADÍSTICA DESCRIPTIVA PARA TRABAJOS DE INVESTIGACIÓN

Presentación e interpretación de los resultados

Diseño de portada: Jorge Sucasaire Pilco

No se permite reproducir, almacenar en sistemas de recuperación de la información ni transmitir parte o el total de esta publicación, cualquiera que sea el medio empleado, sin el permiso previo del autor o editor

Lima-Perú

Presentación

El presente texto tiene el propósito de acercar a las personas que se inician en el desarrollo de una investigación a los principales elementos del análisis estadístico descriptivo. La estadística es una herramienta indispensable para el análisis de datos de una investigación, por ello, el investigador debe comprender los conceptos estadísticos básicos y seleccionar de manera adecuada aquellos que le servirán para presentar y analizar sus resultados.

Los términos estadísticos, en este libro, se explican de forma sencilla y se complementan con ejemplos básicos, pero importantes para reforzar los conceptos y su aplicación pertinente dentro del tratamiento estadístico de acuerdo con el objetivo de un trabajo de investigación. Muchos textos de estadística descriptiva hacen énfasis en el cálculo, mas no en el significado de los resultados, por ello, aquí se aborda de manera práctica la forma de elaborar una interpretación estadística.

Para comprender el contenido de este libro no se requiere tener como base algún entrenamiento previo en estadística o tener conocimientos profundos de matemática, puesto que se hace el uso necesario de las fórmulas sin distraer la atención del lector en simbologías y notaciones estrictamente formales.

En muchas ocasiones los resultados de este texto se calculan por medio de software estadístico como el Minitab, SPSS o Excel. Generalmente, los investigadores requieren los servicios de asesores estadísticos quienes hacen uso de software especializado y brindan la orientación oportuna al investigador en el uso de los métodos estadísticos para el trabajo de investigación. Sin embargo, cuando el investigador comprende lo elementos básicos de estadística puede entablar una comunicación más efectiva con el asesor estadístico, ya que este brinda el análisis de los resultados que obtiene del software, pero no siempre hace referencia a los fundamentos teóricos o al cálculo.

Por tanto, el contenido de este libro es importante porque brinda al investigador las herramientas estadísticas necesarias para que él mismo esté en capacidad de comprender el soporte estadístico de su análisis y sustentar su trabajo de investigación ante la comunidad científica.

Aquí se abordan elementos básicos como la población, la muestra, la variable, las medidas de posición central y no central, las medidas de dispersión y las medidas de forma. En el último capítulo se brindan ejemplos aplicativos para el análisis de resultados de diferentes tipos de variable.

Jorge Sucasaire Pilco

Contenido

Capítulo 1: Estadística e investigación	13
1.1. Estadística	13
1.2. Estadística descriptiva y estadística inferencial	15
1.3. Investigación científica	17
1.4. Tipos de investigación	18
1.5. La investigación cuantitativa y cualitativa.....	20
1.6. Estructura de la investigación científica	22
Capítulo 2: Elementos básicos del análisis estadístico	24
2.1. Población	24
2.2. Muestra y unidad de análisis:.....	24
2.3. Muestras representativas.....	29
2.4. Variables.....	31
2.5. Dato:.....	32
2.6. Medición.....	33
2.7. Niveles o escalas de medición	35
2.7.2. Escala ordinal	37
2.7.3. Escala de intervalo.....	39
2.7.4. Proporción o Razón	41
2.8. Tipos de variables	45
2.8.1. Variables cualitativas:.....	46
2.8.2. Variables cuantitativas.....	49
2.8.3. Variable independiente	52
2.8.4. Variable dependiente	53

2.8.5. Variable de control	53
2.8.6. Variable interviniente	54
2.8.7. Variable simple	56
2.8.8. Variable compuesta	57
2.8.9. Variable dicotómica.....	58
2.8.10. Variable politómica	58
Capítulo 3: Tablas de distribución de frecuencias	61
3.1. Elementos de una tabla.....	63
3.2. Tabla de distribución de frecuencias para variable cualitativa	66
3.3. Tabla de distribución de frecuencias sin intervalos para variable cuantitativa	69
3.4. Tabla de distribución de frecuencias con intervalos.....	72
3.5. Tabla de contingencia.....	79
Capítulo 4: Gráficos estadísticos	82
4.1. Elementos de un gráfico estadístico.....	84
4.2. Tipos de gráficos estadísticos	88
4.3. Diagrama de barras	88
4.3.1. Diagrama de barras verticales	88
4.3.2. Diagrama de barras horizontales	91
4.3.3. Gráfico de barras apiladas.....	92
4.4. Diagrama circular.....	94
4.5. Histograma.....	98
4.6. Polígono de frecuencias.....	100
4.7. Diagrama escalonado	102
4.8. Ojiva.....	103
4.9. Grafica de una distribución bidimensional	105

4.10. Diagrama de dispersión	106
4.11. Gráfico de líneas	111
4.12. Criterios para la interpretación de tablas y gráficos estadísticos.....	113
Capítulo 5: Medidas de posición central.....	118
5.1. Estadígrafos	118
5.2. Media.....	118
5.2.1. Cálculo de la media para datos sin agrupar	120
5.2.2. Cálculo de la media para datos agrupados	120
5.2.3. Cálculo de la media para datos agrupados en intervalos.....	121
5.2.4. Otros promedios	123
5.3. Mediana.....	125
5.3.1. Cálculo de la mediana para datos sin agrupar	126
5.3.2. Cálculo de la mediana para datos agrupados.....	127
5.3.3. Cálculo de la mediana para datos agrupados en intervalos	129
5.4. Moda	131
5.4.1. Cálculo de la moda para datos no agrupados	131
5.4.2. Cálculo de la moda para datos agrupados	132
5.4.3. Cálculo de la moda para datos agrupados en intervalos	133
Capítulo 6: Medidas de posición no central	135
6.1. Cuartiles.....	135
6.1.1. Cálculo de cuartiles para datos sin agrupar	136
6.1.2. Cálculo de cuartiles para datos agrupados	137
6.1.3. Cálculo de cuartiles para datos agrupados en intervalos.....	139
6.2. Deciles.....	143
6.2.1. Cálculo de los deciles	143
6.3. Percentiles.....	146

6.3.1. Cálculo de los percentiles	146
Capítulo 7: Medidas de dispersión.....	150
7.1. Rango	151
7.1.1. Cálculo del rango para datos no agrupados.....	151
7.1.2. Cálculo del rango para datos agrupados	151
7.1.3. Cálculo del rango para datos agrupados en intervalos.....	152
7.2. Rango intercuartil.....	153
7.3. Desviación estándar.....	154
7.3.1. Cálculo de la desviación estándar para datos no agrupados.....	155
7.3.2. Cálculo de la desviación estándar para datos agrupados	157
7.3.3. Cálculo de la desviación estándar para datos agrupados en intervalos.....	160
7.4. Varianza.....	161
7.5. Coeficiente de variación	161
Capítulo 8: Medidas de forma de la distribución.....	165
8.1. Asimetría o sesgo	166
8.1.1. Coeficiente de asimetría.....	169
8.2. Curtosis.....	173
8.2.1. Coeficiente de curtosis.....	176
8.3. Diagrama de cajas	181
Capítulo 9: Análisis descriptivo de los resultados de la investigación	190
9.1. ¿Qué resultados se deben presentar?.....	190
9.2. Criterios básicos para el análisis descriptivo	192
9.3. Análisis descriptivo para variables cualitativas	194
9.3.1. Análisis descriptivo para variable nominal	194
9.3.2. Análisis descriptivo para variable ordinal.....	200
9.4. Análisis descriptivo para variables cuantitativas	210

9.4.1. Análisis descriptivo para variable discreta.....	210
9.4.2. Análisis descriptivo para variable continua	222
Referencias bibliográficas.....	239

Capítulo 1: Estadística e investigación

Contenido

- ❖ Estadística.
- ❖ Estadística descriptiva y estadística inferencial.
- ❖ Investigación científica.
- ❖ Tipos de investigación.
- ❖ Investigación cuantitativa e investigación cualitativa.
- ❖ Estructura de la investigación científica.

1.1. Estadística

La dinámica del mundo actual se rige por la transmisión de información en todos sus formatos. Es difícil concebir la idea de un ciudadano moderno que no acceda a la información mediante la tecnología que tenga a su disposición. El manejo de la información permite la creación de conocimiento y uno de los instrumentos que utilizan los investigadores para procesar la información es la estadística.

La estadística es un término que no es ajeno a nuestro quehacer cotidiano, muchas de las noticias que observamos en los diarios, en los programas informativos de TV o los contenidos de las redes sociales se basan en la estadística. Solo observando la información de nuestra cuenta en Facebook podemos verificar la cantidad de contactos que tenemos, las solicitudes de amistad, el registro de actividad, la cantidad de comentarios por publicación, la cantidad de reacciones a una publicación y recibimos encuestas periódicamente. Todo lo anterior constituye solo una parte de lo que se obtiene mediante el uso de la estadística, por ello, es necesario definirla para reconocer la importancia de su manejo y alcance.

Wackerly et al. (2010) realizan una recopilación de diferentes definiciones para la estadística y encuentran elementos comunes concluyendo que la estadística es una teoría de información que tiene como objetivo la inferencia. Con el análisis estadístico se trata de conocer las características de la población, pero se trabaja con un subconjunto de esta. Este

subconjunto llamado muestra es la base para poder establecer inferencias sobre la población.

Lind et al. (2016) definen a la estadística como una ciencia que recoge, organiza, presenta, analiza e interpreta datos con la finalidad de tomar las decisiones más adecuadas. Esta definición resalta la importancia de la estadística como herramienta en sí misma, es decir, la estadística es un instrumento que nos permite analizar determinados fenómenos o situaciones problemáticas y los resultados obtenidos deben generar alguna respuesta.

Spiegel y Stephens (2009) indican que la estadística se ocupa de los métodos que se usan para recolectar, organizar, resumir, presentar y analizar datos, así como para obtener conclusiones válidas y tomar decisiones razonables.

También se considera que la estadística proporciona métodos que tratan sobre la presentación gráfica de la información, el resumen de datos a través de indicadores, la estimación de parámetros y la prueba de hipótesis (Vargas, 2007).

Podemos concluir, que la estadística es una herramienta para el adecuado manejo de la información. Los procesos estadísticos que señalan los diferentes autores son específicos para cada etapa del análisis estadístico que se inicia con la recolección de datos y culmina con la toma de decisiones.

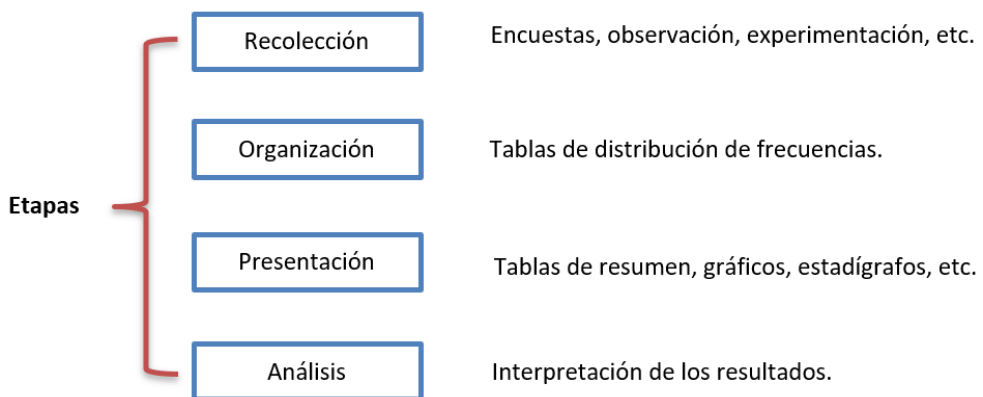


Figura 1.1 Etapas del proceso estadístico

Se resalta, nuevamente, el carácter utilitario de la estadística como soporte para la investigación. Es decir, la estadística no representa un fin en sí mismo, sino que su utilidad depende de los objetivos y etapas de la investigación. Sin embargo, Martínez (2012) nos muestra algunos fines u objetivos que se pueden alcanzar con la estadística:

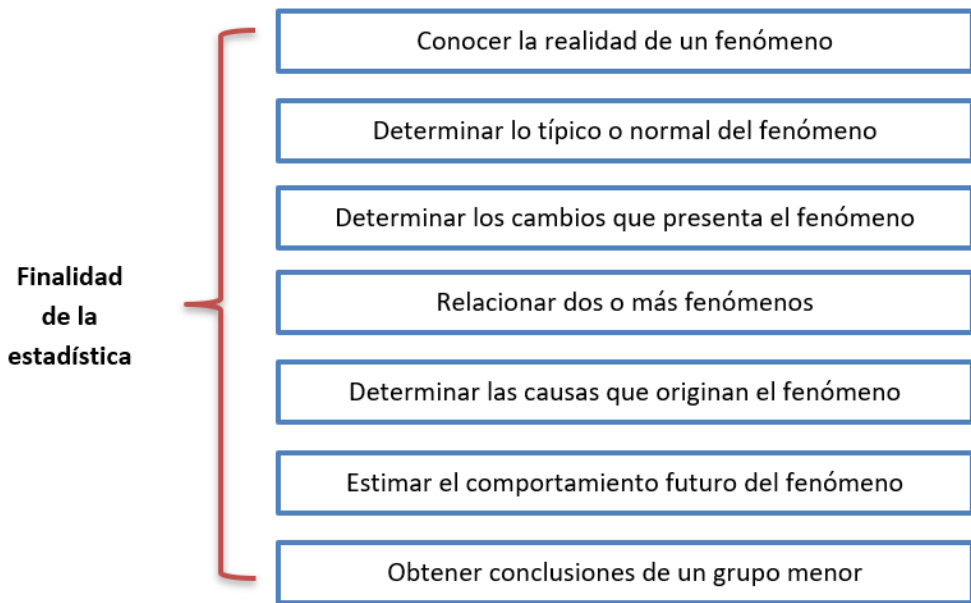


Figura 1.2 Finalidad del análisis estadístico

1.2. Estadística descriptiva y estadística inferencial

Gorgas et al. (2011) consideran que la estadística se divide en dos partes: La primera parte comprende el trabajo relacionado con la organización y análisis inicial de los datos recogidos, la segunda parte comprende la elaboración de conclusiones válidas y la toma de decisiones adecuadas a partir de estas. Esta división realizada por los autores mencionados es la división general que se encuentra con mayor frecuencia en los textos de estadística. La denominación que reciben estas dos partes también es de uso generalizado: la estadística descriptiva y la estadística inferencial.

Medenhall et al. (2010) definen a la estadística descriptiva como aquella parte de la estadística que proporciona procedimientos para resumir y describir las características más importantes de un conjunto de observaciones. Asimismo, definen a la estadística inferencial como un conjunto de procedimientos que permiten establecer conclusiones sobre la población a partir de una muestra tomada de ella.

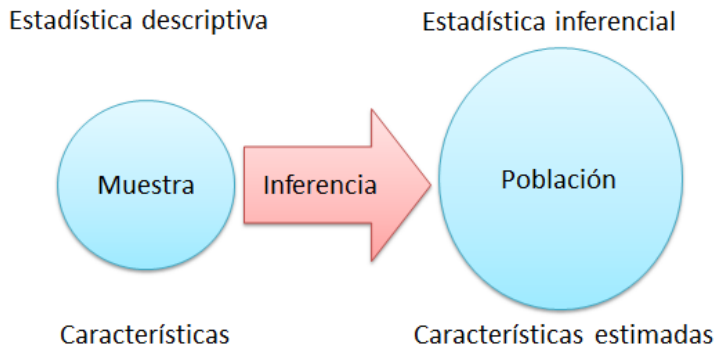


Figura 1.3 Note que el análisis estadístico se inicia con la estadística descriptiva y se complementa con la estadística inferencial

De la figura 1.3 se concluye que las características de la población se pueden estimar a partir de las características de la muestra. En el siguiente capítulo se explica con más detalle lo que es una población y una muestra, por ahora entiéndase que la población es el conjunto de todos los elementos que se quieren analizar y la muestra una parte de la población.

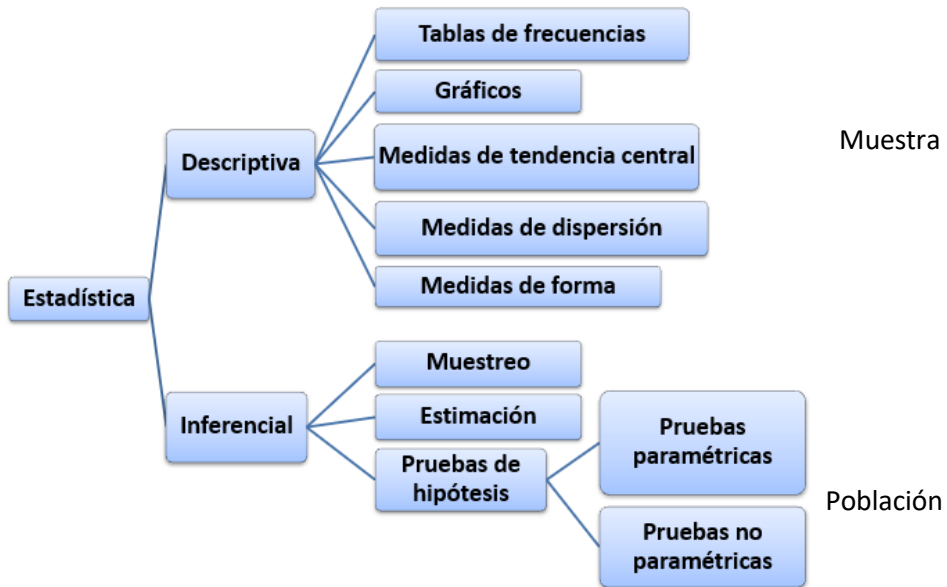


Figura 1.4 Temas que abarcan la estadística descriptiva e inferencial

En la figura 1.4 se muestran los contenidos que abarcan el desarrollo de la estadística descriptiva y la estadística inferencial. Se observa que en la estadística descriptiva se resume las características de la muestra y en la estadística inferencial se estiman las características de la población.

1.3. Investigación científica

La investigación es una actividad que implica la obtención de nuevo conocimiento. Esta actividad no está limitada a un grupo reducido de personas, puesto que, dentro del quehacer de un profesional se incluye esta práctica. En nuestra realidad encontramos diferentes fenómenos que presentan un cambio continuo, lo que genera oportunidad para investigar. Por ello, es necesario fortalecer las competencias propias de la labor investigativa. Actualmente existen diferentes medios y recursos accesibles para poder iniciarse en la investigación.

La investigación en las diferentes disciplinas se basa en la aplicación del método científico. La investigación científica es un proceso compuesto de

diferentes etapas secuenciales con la finalidad de obtener conocimientos nuevos y válidos sobre algún aspecto de la realidad.

Martínez (2012) señala que la investigación científica es una actividad que consiste en estudiar algún fenómeno de nuestra realidad, de forma sistemática con la finalidad de comprenderlo y explicarlo. Hernández, Fernández et al. (2010) señalan que la investigación es un conjunto de procesos sistemáticos, críticos y empíricos que se utilizan para estudiar un fenómeno o problema.



Figura 1.5 La investigación es un instrumento para conocer la realidad.

1.4. Tipos de investigación

En la literatura referente a metodología de la investigación científica existen diferentes formas de clasificar los tipos de investigación. Tamayo (2003) resalta que casi nunca se presentan estudios con un tipo de investigación único; generalmente se combinan entre sí y se adecúan a los requerimientos de la investigación.

Hernández, Fernández et al. (2010) utilizan el término alcance para denominar los diferentes tipos de investigación cuantitativa resaltando la importancia del orden en que se debe presentar los elementos de esta clasificación. Es decir, se debe tener en cuenta la profundidad del análisis y los objetivos de la investigación. Así, la investigación puede tener un alcance exploratorio, descriptivo, correlacional o explicativo (figura 1.6). Además, consideran que el alcance depende principalmente del estado de conocimiento que se tiene sobre el problema de estudio y de la perspectiva que se pretende dar al estudio.

Hernández, Ramos et al. (2018) clasifican la investigación según su diseño, así, distinguen entre diseños experimentales, no experimentales, transversales, longitudinales, etc. La elección del diseño depende de las

características de la unidad de análisis, de los objetivos y del alcance de la investigación.

Alcance	Objetivo
Exploratorio	Abordar problemas nuevos o poco estudiados.
Descriptivo	Especificar las características más importantes del fenómeno .
Correlacional	Descubrir la relación entre las variables en estudio.
Explicativo	Determinar las causas del fenómeno en estudio.

Figura 1.6 Alcances de la investigación cuantitativa según la profundidad del análisis y los objetivos según Hernández, Fernández et al. (2010).

Tamayo (2003) considera formas y tipos de investigación.

Forma	Objetivo
Pura	Plantear la teoría.
Aplicada	Confrontar la teoría con la realidad.
↓	
Tipos	
Histórica	Describir lo que era.
Descriptiva	Interpretar lo que es.
Experimental	Describir lo que será.

Figura 1.7 Formas y tipos de investigación.

Ñaupas, et al. (2014) resaltan que, en investigación científica, generalmente se distinguen la investigación básica y la investigación aplicada.

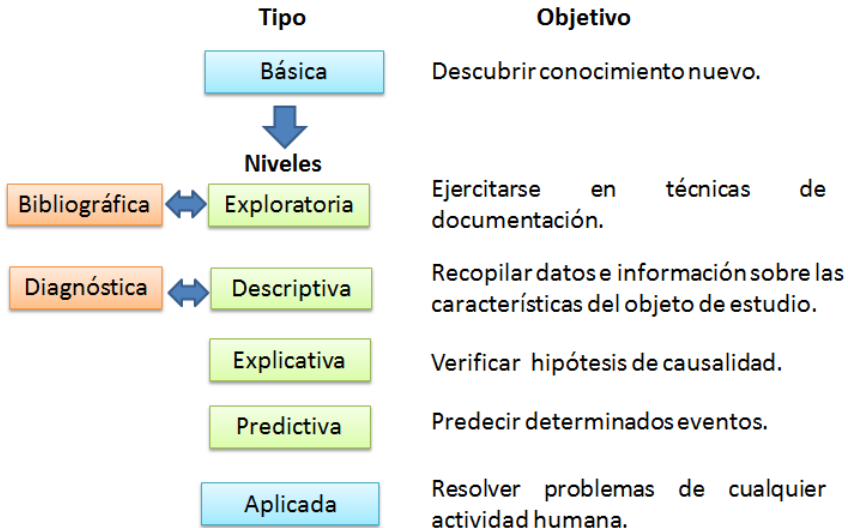


Figura 1.8 Tipos y niveles de la investigación.

De esta breve clasificación se puede entender que en la investigación exploratoria no se realiza un tratamiento estadístico con la información recolectada (figura 1.8), puesto que se trata de un problema nuevo sobre el que hay que revisar lo poco que se ha escrito al respecto y no se tiene el conocimiento necesario para establecer el método estadístico correspondiente. Sin embargo, la investigación exploratoria sirve como base para investigaciones futuras que aborden la misma problemática. En los demás tipos de investigación si es necesario el uso de herramientas estadísticas para analizar la información.

1.5. La investigación cuantitativa y cualitativa

Hernández, Fernández et al. (2010) consideran que en investigación existen 3 enfoques, la investigación cualitativa, la investigación cuantitativa y la investigación mixta. De esta manera, señalan que en la investigación cuantitativa se recolectan datos para probar hipótesis utilizando el análisis estadístico, en cambio, en la investigación cualitativa la recolección de datos no implica una medición numérica. Con base en el párrafo anterior

y en la naturaleza de este libro se justifica que en adelante solo centremos nuestro interés en la investigación cuantitativa.

Leyton (2018) diferencia el campo de acción de la investigación, así, sostiene que la investigación cuantitativa se utiliza en las Ciencias Naturales que son disciplinas que tratan con objetos. De otro lado señala que la investigación cualitativa abarca las Ciencias Sociales que son las disciplinas que tratan con sujetos. Sin embargo, hay que resaltar que existen métodos estadísticos que permiten hacer investigación cuantitativa en diferentes disciplinas como la psicología, la pedagogía o las ciencias de la comunicación.

Bologna (2011) afirma que el uso de procedimientos estadísticos en la investigación aporta objetividad al análisis y esto es muy cierto, dado que la estadística brinda el soporte matemático para entender las características y relaciones entre las variables de la investigación. Sin embargo, esto no resta importancia a la investigación cualitativa que cuenta con procedimientos propios para el análisis de la información.

A continuación, se presentan algunas características de la investigación cuantitativa.

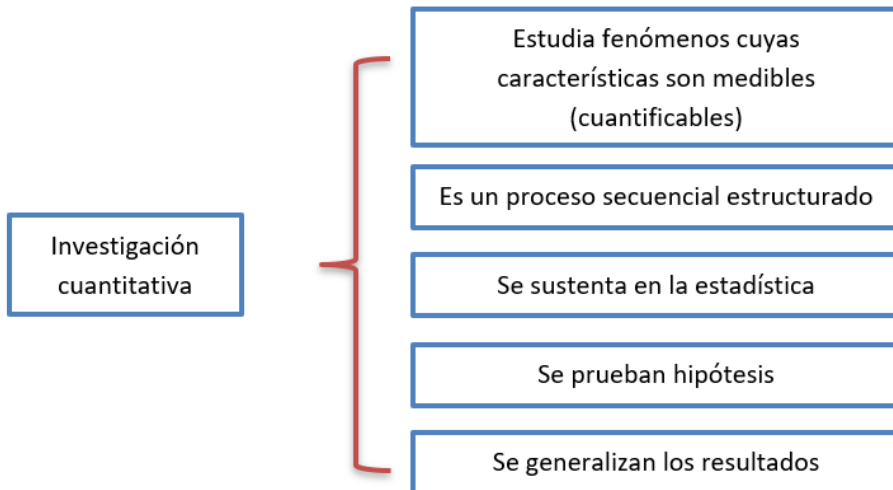


Figura 1.9 Las características presentadas son las más resaltantes en investigación cuantitativa.

1.6. Estructura de la investigación científica

Tamayo (2003) muestra un esquema general sobre los pasos a seguir para elaborar una investigación. La investigación es un proceso secuencial y en el siguiente esquema adaptado del libro “el proceso de la investigación científica” se aprecia claramente que cada paso se fundamenta en el paso anterior.

Pasos	Descripción
Elección del tema	Planteamiento.
Delimitación del tema	Revisión del conocimiento, Alcances, límites y recursos.
Problema	Identificación, descripción , elementos y formulación.
Objetivos	Generales y específicos.
Marco teórico	Antecedentes, definición conceptual, hipótesis y variables.
Metodología	Población, muestra, recolección y procesamiento de datos.
Informe	Resultados, discusión, conclusiones y recomendaciones.

} Uso de herramientas estadísticas

Figura 1.10 Estructura de la investigación científica según Tamayo (2003).

La estructura mostrada en la figura 1.10 sirve para que el investigador pueda orientarse dentro del proceso investigativo y entender que el tratamiento estadístico se desarrolla durante la elaboración de la metodología y el informe. Sin embargo, la elección del tratamiento estadístico no es independiente de los pasos anteriores, sino que se fundamenta en ellos (figura 1.11).

La elección del procedimiento estadístico responde al nivel de medición de las variables, la formulación de las hipótesis, los objetivos y problemas de investigación. Por ello, el investigador debe nutrirse con la teoría existente con respecto a su problema de investigación con la finalidad de establecer hipótesis coherentes e identificar y definir adecuadamente las variables.

La correcta definición de las variables permite la elaboración de un instrumento de medición que aborde todas las dimensiones e indicadores posibles de la variable. De esta manera se busca construir instrumentos válidos y confiables.

En la actualidad las instituciones manejan su propio protocolo de presentación para trabajos de investigación. Sin embargo, su estructura no dista mucho de la presentada en la figura 1.10.

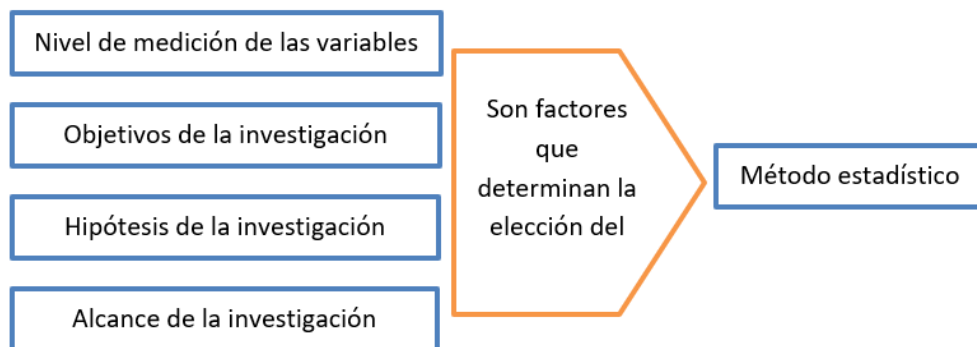


Figura 1.11 Algunos factores que determinan la elección del método estadístico.

Capítulo 2: Elementos básicos del análisis estadístico

Contenido

- ❖ Población
- ❖ Muestra y unidad de análisis
- ❖ Muestras representativas
- ❖ Variables
- ❖ Datos
- ❖ Medición
- ❖ Escalas de medición
- ❖ Tipos de variables

Para iniciarse en la estadística es necesario comprender ciertos conceptos que serán parte del lenguaje utilizado en los capítulos posteriores. Como toda disciplina la estadística maneja conceptos y simbología propia.

2.1. Población

Es el conjunto compuesto por todos los individuos u objetos donde se observa la característica que se quiere analizar.

Vargas (2007) señala que la población se define de acuerdo con los objetivos de la investigación. Esta afirmación es importante para que el investigador delimite adecuadamente la población sobre la que planea realizar su investigación.

Por ejemplo:

- Institutos superiores del Perú.
- Hoteles de la ciudad de Lima.
- Países de Latinoamérica.

2.2. Muestra y unidad de análisis:

La muestra es un subconjunto de una población, que intenta reflejar las características de la población de la forma más cercana posible.

Por ejemplo:

- 20 institutos superiores del Perú.
- 25 hoteles de la ciudad de Lima.
- 8 países latinoamericanos.

La unidad de análisis hace referencia a cada uno de los elementos que constituyen la población y por lo tanto la muestra.

Por ejemplo:

- El Instituto Superior Tecnológico “José Pardo”.
- El hotel Sheraton.
- El Perú.

En investigación experimental se utiliza el término unidad experimental que hace referencia al individuo u objeto que es parte del experimento y en el que se mide la variable de interés (Medenhall et al., 2010).

A continuación, se presentan tres situaciones donde se debe reconocer la población, la muestra y la unidad de análisis.

Situación 1: Una empresa de créditos quiere colocar una oficina en el supermercado “A” y necesita conocer el tipo de tarjeta que utilizan los clientes que hacen compras a crédito en dicho supermercado. Por ello, se realiza una encuesta a 280 clientes que hacen compras a crédito.

Población: Clientes que hacen compras a crédito en el supermercado A.

Muestra: 280 clientes encuestados.

Unidad de análisis: Un cliente que realiza compras a crédito en el supermercado A.

Situación 2: Un grupo de estudiantes realiza una investigación para conocer el gasto semanal en transporte público que realizan los estudiantes

de su universidad. Por ello encuesta a 50 estudiantes de cada ciclo.

Población: Estudiantes de dicha universidad que utilizan transporte público.

Muestra: Los estudiantes encuestados.

Unidad de análisis: Un estudiante de la universidad que utiliza el transporte público.

Situación 3: En una urbanización viven 2 500 personas y se quiere conocer los hábitos alimentarios de las personas mayores a 30 años que viven en ese lugar. Por ello, se selecciona a 320 personas con esa característica y se les toma una encuesta.

Población: Personas mayores a 30 años que viven en la urbanización.

Muestra: Las 320 personas encuestadas.

Unidad de análisis: Una persona mayor de 30 años que vive en dicha urbanización.

¿Por qué utilizar muestras?

A la cantidad de elementos que conforman la población se le puede llamar tamaño de la población y se le simboliza con la letra N y a la cantidad de elementos que conforman la muestra se le puede llamar tamaño de la muestra y se le simboliza con la letra n .

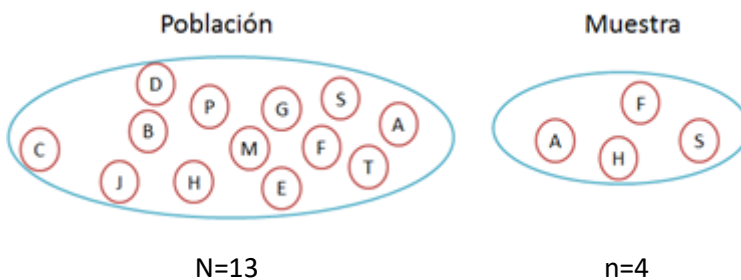


Figura 2.1 Tamaño de la población y tamaño de la muestra.

El uso de muestras se justifica porque frecuentemente no es posible analizar a toda la población. Es decir, existen limitaciones para poder acceder a toda la población. Un claro ejemplo son las encuestas donde se

intenta conocer las características de toda la población, sin embargo, esta tarea no es posible por diversos factores. Por ejemplo, el medio geográfico puede ser un factor adverso ya que existen lugares de difícil acceso, el tiempo es otro factor que afecta una investigación, ya que no se puede extender de forma indeterminada, inclusive el idioma es otro factor limitante que no permite la comunicación efectiva (en países que tienen comunidades con etnias de diferente lengua) entre el encuestador y el encuestado.

La información estadística que se recibe a diario, generalmente se basa en características de una muestra. Las empresas encuestadoras seleccionan un grupo de individuos y en base a ellos nos presentan la información estimada sobre las características de la población.

Ejemplo: A continuación, se presenta parte de la ficha técnica de una encuesta realizada por la encuestadora Ipsos Opinión y Mercado en el año 2020.

Ficha técnica	
Encuesta nacional urbana	
Nombre de la encuestadora	: IPSOS OPINION Y MERCADO S.A.
Objetivo del estudio	: Evaluar la percepción, opinión y actitudes hacia temas políticos, económicos y sociales del país.
Tamaño de la población objetivo	: 19 064 735 personas
Tamaño de la muestra	: 1019 personas entrevistadas
Procedimiento de selección del entrevistado	: Fueron elegidos de manera aleatoria dentro de cada vivienda respetando las cuotas de sexo y edad.
Fecha de trabajo de campo	: Del 09 al 10 de julio del 2020

Figura 2.2 Extracto de una ficha de muestreo. Fuente: IPSOS.

En la figura 2.2 se puede observar la gran diferencia entre el tamaño de la población y el tamaño de la muestra.

Ejemplo: A continuación, se muestra parte de la ficha técnica de la encuesta realizada por la empresa SEDAPAL para su estudio de impacto ambiental detallado del proyecto: “Optimización de sistemas de agua potable y alcantarillado, sectorización, rehabilitación de redes y actualización de catastro – Área de influencia planta Huachipa – Área de drenaje Comas – Chillón - Lima”.

Ficha técnica	
Objetivo del estudio	: Obtener información sobre la opinión sobre el proyecto y percepción de impactos.
Universo	: Hombres y mujeres mayores de edad residentes permanentes en el distrito de Comas, de edades entre los 18 a 70 años.
Tamaño de la muestra	: 202 personas elegidas en forma aleatoria en base a una distribución geográfica por conglomerado de superficies geográficas (viviendas) en las zonas donde se ejecutarán las obras generales.
Segmentación	: Por habilitación urbana.
Fecha de trabajo de campo	: Domingo 07 al martes 09 de febrero del 2010.

Figura 2.3 Extracto de una ficha de muestreo. Fuente: SEDAPAL

En esta ficha de la figura 2.3 también se observa que la muestra es pequeña en comparación con la población que se compone de todos los residentes en el distrito de Comas.

Ejemplo: En la siguiente tabla se muestra parte de la ficha técnica de la encuesta de opinión pública realizada por la empresa DATUM.

Ficha técnica Encuesta de opinión pública a nivel nacional	
Objetivo del estudio	: Obtener información de coyuntura política, económica y social.
Universo	: Hombres y mujeres, de 18 a 70 años, pertenecientes a todos los niveles socioeconómicos, considerando zonas urbana y rural.
Tamaño de la muestra	: 1201 encuestas efectivas.
Selección	: Selección aleatoria de personas.
Fecha de campo	: 25 al 29 de noviembre del 2016.

Figura 2.4 Extracto de una ficha de muestreo. Fuente: DATUM.

En la figura 2.4 no se precisa el tamaño de la población, pero sabemos que es una cantidad muy grande, dado que la encuesta se realiza a nivel nacional y los individuos son varones y mujeres de 18 a 70 años.

2.3. Muestras representativas

Es necesario hacer énfasis en el carácter representativo de la muestra. En la actualidad se cuenta con el apoyo de la tecnología, pero muchas veces se utiliza de manera incorrecta. Un claro ejemplo, es el uso indiscriminado de las encuestas online. Este tipo de encuestas permiten al investigador tener un mayor alcance y acceder a un número muy grande de personas. El problema es que los individuos que acceden a responder a estas encuestas son los que tienen la voluntad de hacerlo y se deja de lado a individuos que pueden ser elementos representativos de nuestra población de estudio. Triola (2004) sostiene que con muestras de este tipo solo se pueden elaborar conclusiones sobre esa muestra. Es decir, no es correcto establecer conclusiones para la

población a partir de estas muestras voluntarias, puesto que estas muestras carecen de representatividad.

De lo anterior no se infiere que se deba prescindir de las encuestas online. Lo que se trata de resaltar es que antes de enviar la encuesta se debe realizar un proceso estadístico riguroso denominado muestreo. Es decir, la encuesta online no se debe enviar a cualquier persona, sino a aquellos individuos que presentan las características principales de la población que se quiere analizar.

El muestreo es el estudio de la relación que existe entre una población y las muestras que se pueden obtener de esta población. El muestreo es importante porque permite realizar inferencias de una muestra hacia la población por medio de prueba de hipótesis (Spiegel y Stephens, 2009, p. 203).

Cuando se obtiene datos de toda la población, sólo se requiere establecer conclusiones a partir de la estadística descriptiva (Medenhall et al., 2010). En estos casos no es necesario hacer uso de estadística inferencial. Por ejemplo, no se necesita estimar la media o la proporción, dado que el conjunto de observaciones corresponde a toda la población.

Leyton (2018) considera que toda muestra debe cumplir con dos requisitos de representatividad: En primer lugar, la representatividad cuantitativa referente a la cantidad de elementos de la muestra, con respecto al tamaño de la población y estratos, que son suficientes para establecer el análisis inferencial. En segundo lugar, está la representatividad cualitativa que considera que los elementos de la muestra deben contar con las características generales y específicas de la población en estudio.

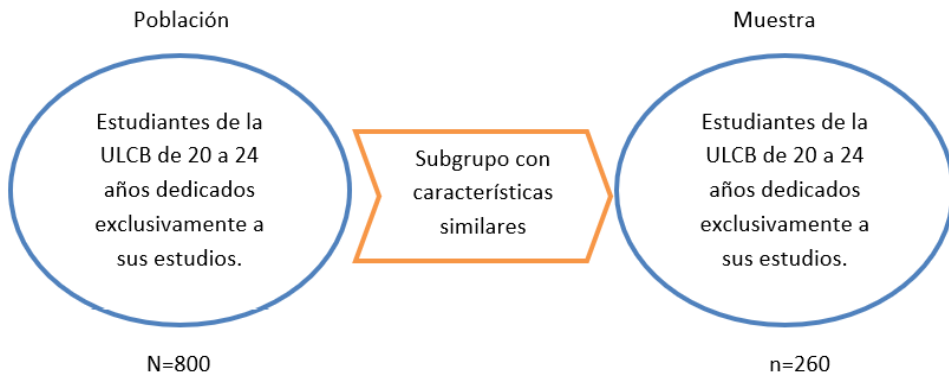


Figura 2.5 La muestra debe ser representativa tanto en calidad como en cantidad.

En el gráfico 2.5 se resalta que las características de la muestra deben ser las mismas características de la población. El tamaño de la muestra “n” se obtiene mediante un cálculo estadístico.

2.4. Variables

Una variable se puede entender como la característica o propiedad que se desea estudiar en las unidades de análisis (personas, cosas, entidades, etc.). Una variable es susceptible de ser medida y dentro del proceso de medición se establecen los criterios de comparación.

Entendida como característica, la variable se presenta en mayor o menor magnitud en los diferentes objetos de estudio. Se debe resaltar que en la investigación no se comparan los objetos sino las características de los objetos (Bologna, 2011). Es decir, no se analiza la unidad de análisis en sí misma sino sus características (Martínez, 2012). Por ello, las variables deben estar bien definidas para establecer un correcto proceso de recolección de datos.

Hay que tener en cuenta que se pueden analizar muchas variables en una sola unidad de análisis. Por ejemplo, si observamos una pelota o una persona:

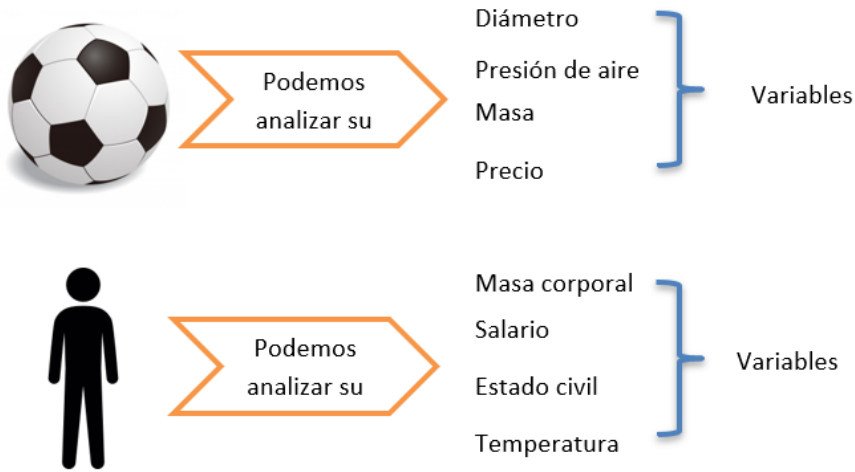


Figura 2.6 Las variables son características observables de los objetos.

2.5. Dato: Es el valor que se obtiene al realizar la medición de la variable en estudio.

Los datos son la base para la posterior organización y análisis de la variable en estudio. Cabe resaltar que luego de recolectar datos, estos, en conjunto no generan información directa (Vargas, 2007), sino que hay que establecer criterios para su tratamiento y posterior interpretación.

Ejemplos:

Existen variables cuyos datos recolectados no son numéricos como el caso de la variable nivel de estrés.

No	Nivel de estrés
1	Alto
2	Bajo
3	Bajo
4	Medio
5	Alto
6	Alto
7	Alto
8	Medio

Datos

Los datos recolectados son atributos

Figura 2.7 Bajo, medio y alto son datos que indican el nivel de estrés.

Otras variables presentan datos recolectados que están representados por números como la variable consumo diario de calorías (Kcal) y la variable masa corporal (kg).

No	Consumo diario de calorías
1	1400
2	1800
3	1760
4	1640
5	1600
6	1800
7	1540
8	1700

No	Masa corporal
1	55.70
2	54.80
3	70.34
4	73.80
5	68.00
6	65.20
7	70.50
8	72.86

Los datos recolectados son cantidades

Figura 2.8 Los datos numéricos pueden ser enteros o decimales.

2.6. Medición

Medir una variable implica hacer una comparación entre el valor observado y un valor de referencia. Ruiz (2005) concluye que la medición consiste en asignar números a los objetos bajo ciertas reglas, de modo que estos números representen relaciones y no solamente cantidades. Este concepto de medición es más amplio e incorpora el conjunto de los datos no numéricos dentro del proceso de medición.

Existen procedimientos muy diferentes para efectuar la medición de las variables, pues algunas variables se pueden medir directamente con un instrumento mecánico, como es el caso de una balanza que permite medir la masa corporal. Sin embargo, existen variables cuya medición es más compleja.

Existen variables cuya medición se realiza mediante dispositivos mecánicos como la variable masa, la temperatura o la longitud.

Tabla 2.1 Instrumentos mecánicos para la medición de las variables.

Variable	Instrumento de medición
Masa	Balanza
Temperatura	Termómetro
Longitud	Cinta métrica

Existen variables que se pueden medir directamente por medio de otro tipo de procedimientos.

Tabla 2.2 Instrumentos no mecánicos para la medición de las variables.

Variable	Instrumento de medición
Edad	Encuesta o documentos administrativos
Estado civil	
Profesión	

Existen otras variables que no pueden ser medidas directamente por medios mecánicos ni con una simple pregunta de encuesta o revisión de documentos. Para efectuar la medición de estas variables se debe establecer las características observables y medibles que componen dicha variable. Este último proceso se denomina operacionalización de la variable. Mendoza y Garza (2009) consideran que los conceptos por sí mismos no son directamente observables, por ello es necesario utilizar algún elemento observable que refleje al concepto (figura 2.9).

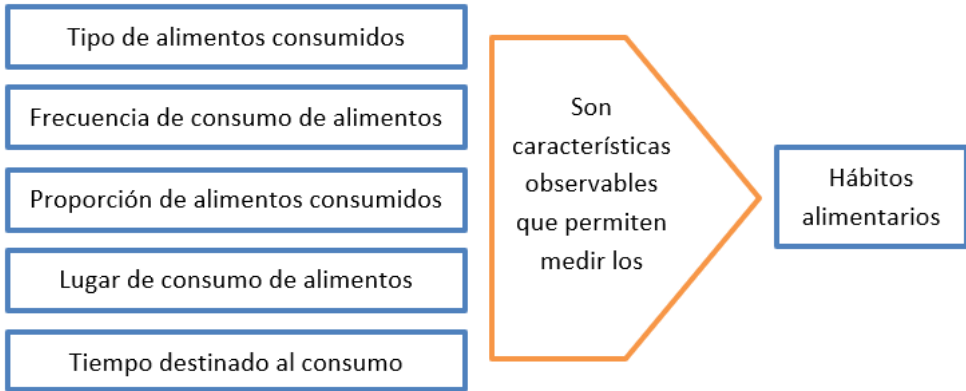


Figura 2.9 Las variables complejas requieren más de una observación para ser medidas.

Siegel y Castellan (1995) resaltan que la medición que realiza un investigador de las ciencias físicas y un investigador social son muy diferentes. En las ciencias físicas se les asignan números a las observaciones y con dichos números es coherente realizar operaciones aritméticas, pero esto no se puede generalizar a las ciencias de la conducta. Las operaciones aritméticas aplicadas a los números que se asignan a las observaciones en las ciencias de la conducta no siempre tienen un sentido concreto.

2.7. Niveles o escalas de medición

Anderson et al. (2012) resaltan que las escalas de medición nos indican la forma más apropiada de analizar los datos, puesto que, cada escala representa el tipo de información que presentan los datos.

El nivel de medición de una variable está determinado por el significado que tengan los símbolos numéricos que se asignan a las categorías (Bologna, 2011). Una categoría se entiende como el conjunto de valores que puede tomar una variable.

Es importante conocer y comprender las escalas de medición, puesto que el uso de una determinada escala orienta al investigador en la elección del tratamiento estadístico adecuado.

2.7.1. Escala nominal

Implica la agrupación de objetos en categorías basándose en atributos o propiedades observadas.

En esta escala se pueden utilizar números, pero estos se asignan solo como etiqueta, cada número representa una categoría diferente. Por lo tanto, no se pueden realizar operaciones aritméticas con dichos números.

Ejemplo:

- Estado civil
- Grupo sanguíneo
- Sexo
- Nacionalidad
- Partido político
- Profesión

Para entender la escala de medición nominal consideremos la variable medio de transporte utilizado por los estudiantes de una comunidad de Cusco. Para obtener la información preguntamos ¿Qué medio de transporte utiliza con mayor frecuencia? Establecemos criterios para clasificar los medios de transporte en: público y privado. Podemos utilizar los números 1 y 2 para designar a cada uno de estos grupos.

Tabla 2.3 Categorías de una variable en la escala nominal.

Variable	Categorías	Valor asignado
Medio de transporte	Público	1
	Privado	2

En la tabla 2.3 se observa que las respuestas de los encuestados pertenecen a una u otra categoría. Note que el número 2 no significa que el medio de transporte privado sea mejor que el medio de transporte público al que se le

ha asignado el número 1. Los números asignados solo se utilizan con fines de codificación.

Si observamos la variable nacionalidad con las categorías presentadas (peruano, boliviano, panameño y uruguayo) en diferentes posiciones:

Caso 1		Caso 2	
Variable	Categorías	Variable	Categorías
Nacionalidad	Uruguayo (1)	Nacionalidad	Peruano (1)
	Boliviano (2)		Boliviano (2)
	Peruano (3)		Panameño (3)
	Panameño (4)		Uruguayo (4)

Caso 3	
Variable	Categorías
Nacionalidad	Boliviano (1)
	Uruguayo (2)
	Panameño (3)
	Peruano (4)

Figura 2.10 La forma de presentar las categorías queda a criterio del investigador.

Se puede concluir que no existe un orden específico para ubicar las categorías, cualquiera de los tres casos se puede utilizar para presentar la información estadística en una tabla. Es decir, no existe un criterio establecido para subordinar una nacionalidad en comparación con otra.

2.7.2. Escala ordinal

Se utiliza cuando las observaciones pueden ubicarse según cierto orden o jerarquía con respecto a la característica que se evalúa. Esta escala de medición tiene las características de la escala nominal, pero se agrega la característica de que entre las categorías se puede establecer una relación de orden.

En esta escala de medición los números asignados a las categorías indican el orden.

Ejemplo:

- Nivel de estrés
- Hábitos alimentarios
- Nivel socioeconómico
- Grado de instrucción
- Dominio de un idioma

Si observamos el nivel de estrés en un grupo de personas, el posible resultado puede ser que cada individuo presente un bajo, moderado o alto nivel de estrés. Esta información se puede organizar de la siguiente forma:

Tabla 2.4 Categorías de una variable en la escala ordinal.

Variable	Categorías
Nivel de estrés	Bajo (1)
	Moderado (2)
	Alto (3)

Observe que las categorías se pueden ordenar de acuerdo a la intensidad del estrés registrado.

Si observamos el mayor grado académico de los docentes de una universidad, los posibles resultados pueden ser los siguientes:

Tabla 2.5 Categorías de una variable en la escala ordinal.

Variable	Categorías
Máximo grado académico	Bachiller (1)
	Maestría (2)
	Doctorado (3)
	Post doctorado (4)

En la tabla 2.5 la lógica para ordenar las categorías es evidente, puesto que, algunos grados académicos tienen mayor jerarquía que otros. Por ejemplo, para obtener la maestría primero se tiene que obtener el grado de bachiller.

En la escala de medición ordinal se adiciona la relación de orden “mayor que”. En la escala nominal no se puede establecer esta relación. Si adecuamos la relación de orden utilizando “mejor que” ¿el transporte público es mejor que el privado? O ¿el transporte privado es mejor que el público? Esta relación no se puede determinar en la tabla 2.3 porque la información que se obtiene con la pregunta establecida solo hace referencia al medio de transporte que utilizan con más frecuencia. En ningún momento se planteó preguntar la opinión de los estudiantes.

Pero si analizamos la variable dominio del idioma inglés en un grupo de estudiantes por medio de una encuesta, podemos clasificar las respuestas como: básico, intermedio y avanzado.

Tabla 2.6 Categorías de una variable en la escala ordinal.

Variable	Categoría	Valor asignado
Dominio del idioma ingles	Básico	1
	Intermedio	2
	Avanzado	3

En la tabla 2.6 además de clasificar los resultados se puede establecer la relación de orden “mayor que”. Los estudiantes del nivel intermedio tienen un mayor dominio del idioma inglés que los estudiantes del nivel básico. Asimismo, los estudiantes del nivel avanzado tienen un mayor dominio del idioma inglés que los estudiantes del nivel intermedio.

2.7.3. Escala de intervalo

Esta escala de medición tiene las características de la escala nominal y ordinal. Sin embargo, los datos de esta escala son numéricos y se puede cuantificar la diferencia entre dos valores. Aquí el cero es relativo y no implica la ausencia del atributo.

Ejemplo:

- Temperatura
- Contribución al PBI

- Puntaje de aptitud matemática
- Coeficiente intelectual

En la escala de medición de intervalo se observa que las distancias entre los valores de los objetos tienen un significado determinado. Esto no se cumple en la escala anterior. Si sabemos que un estudiante del nivel avanzado tiene un mayor dominio del idioma inglés que un estudiante del nivel intermedio, pero ¿Cuánto más domina? Sabemos que un estudiante del nivel intermedio tiene un mayor dominio del idioma inglés que un estudiante del nivel básico, pero ¿Cuánto más domina?

Estas preguntas no se pueden responder con la información recabada en una escala ordinal, para ello sería necesario hacer una evaluación más minuciosa.

Por ejemplo, si registramos los puntajes de una prueba de aptitud:

Tabla 2.7 Resultados de la prueba de aptitud.

Estudiante	Puntaje
Luis	7
Aarón	8
Pedro	9
Lucas	10
Andrés	11
Liam	12
José	13
Noel	14
Nilton	15

Con base en la tabla 2.7 se puede afirmar que José superó a Pedro con 4 puntos y Nilton superó a Lucas con 5 puntos. En esta escala, además del orden, las distancias entre valores tienen significado y permiten realizar comparaciones.

En el nivel de medición de intervalo, si se tiene dos formas de medir la misma variable con puntajes x e y . El valor de y puede obtenerse a partir de x según la siguiente relación (Bologna, 2011):

$$y = a + bx$$

Donde a y b son constantes arbitrarias.

Por ejemplo, sabemos que existen diferentes escalas para medir la temperatura. Si consideramos la relación entre los grados Celsius (x) y los grados Kelvin (y) tenemos aproximadamente:

$$y = 273 + x$$

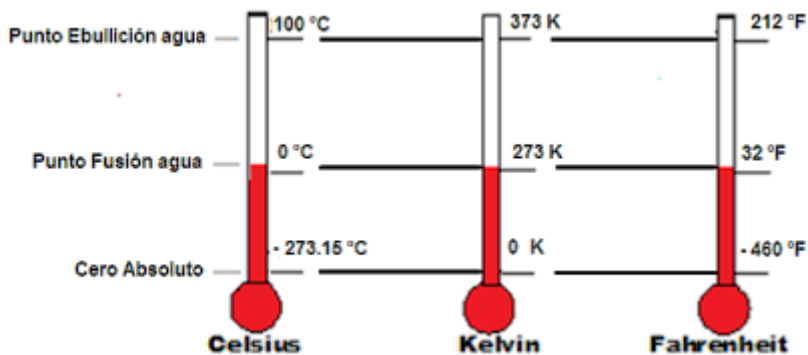


Figura 2.11 Escalas de medición para la temperatura. Recuperado de: <https://sites.google.com/site/matematicasjuanmanuelista/fisica-11/termodinamica/escalas-de-temperatura>

En la figura 2.11 se puede observar que el cero de la escala Celsius es arbitrario, es decir, no representa la temperatura más baja.

La escala de medición de intervalo es la primera escala cuantitativa. Aquí los datos pueden ser tratados con operaciones aritméticas y los resultados tienen un significado concreto.

2.7.4. Proporción o Razón

Tiene las mismas propiedades que la escala de intervalo, pero, el cero representa la ausencia del atributo. Es decir, existe el cero absoluto.

Ejemplo:

- Salario
- Estatura
- Gasto
- Consumo de energía eléctrica
- Masa corporal
- Costo de producción

En el nivel de medición de razón, si se tiene dos formas de medir la misma variable con puntajes x e y . El valor de y puede obtenerse a partir de x según la siguiente relación (Bologna, 2011):

$$y = bx$$

Donde b es una constante arbitraria.

Por ejemplo, sabemos que existen diferentes escalas para medir la longitud. Si consideramos la relación entre los centímetros (y) y las pulgadas (x) tenemos aproximadamente:

$$y = 2.5x$$

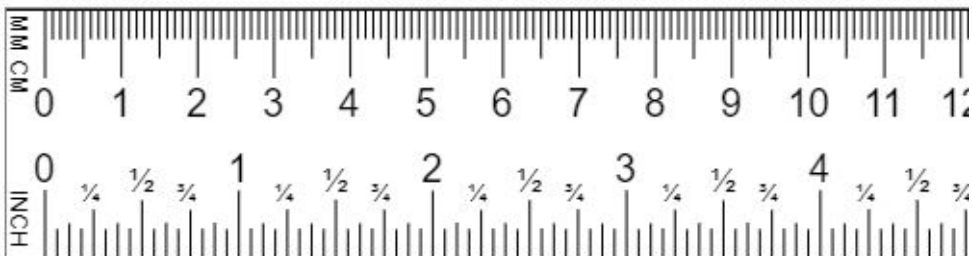


Figura 2.12 Escalas de medición para la longitud. Observe que tanto en el sistema de medición en centímetros como en pulgadas el origen es cero.

En la figura 2.12 se puede observar que el cero es absoluto. Es decir, no importa el valor de la constante b , el cero siempre será el menor valor para cualquier forma de medición de la longitud.

Otra característica de la escala de intervalo es que las distancias entre los valores asignados a la medición tienen significado preciso. Por ejemplo, si un auto recorre 10 Km y camión recorre 20 Km, se puede afirmar que el camión recorrió el doble de distancia que el auto. Esto no sucede en la escala de intervalo, si en una prueba de aptitud el estudiante A obtiene 8 puntos y el estudiante B obtiene 24, no tiene sentido afirmar que el estudiante B es triplemente más apto que el estudiante A.

Veamos otro ejemplo para diferenciar entre una escala de intervalo y una escala de razón, para esto usaremos las variables gasto y temperatura:

Tabla 2.8 Gasto en soles de dos personas.

	Gasto (S/)
Persona 1	300
Persona 2	1200

Con base en la tabla 2.8 tiene sentido afirmar que la persona 1 gasta la cuarta parte de lo que gasta la persona 2.

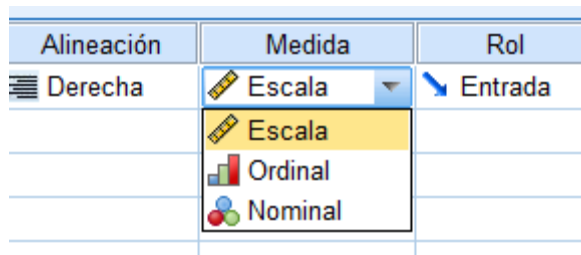
Tabla 2.9 Temperatura de dos objetos.

	Temperatura (°C)
Objeto 1	20
Objeto 2	40

En este caso de la tabla 2.9 no tiene sentido afirmar que el objeto 1 está la mitad de caliente en comparación con el objeto 2. La variable temperatura pertenece al nivel de medición de intervalo.

Del mismo modo que en la escala de medición de intervalo, en la escala de razón, los datos también pueden ser operados de forma aritmética y se pueden calcular diferentes estadígrafos, además de aplicar las diferentes pruebas estadísticas.

El programa SPSS agrupa las escalas de medición de intervalo y de razón en una sola, debido a las características similares que presentan, y la denomina como nivel escalar. Por ello, en el menú para indicar la escala de una variable solo aparecen 3 niveles.



Alineación	Medida	Rol
Derecha	Escala	Entrada
	Escala	
	Ordinal	
	Nominal	

Figura 2.13 Niveles de medición según programa estadístico SPSS.

Siegel y Castellan (1995) presentan una clasificación más completa de acuerdo a la complejidad de cada escala y a las propiedades formales que las caracterizan. En base a ello se muestra la siguiente tabla:

Tabla 2.10 Comparación de las escalas de medición.

Escala de medición	Nominal	Ordinal	Intervalo	Razón
Características	Uso de símbolos para clasificar objetos.	La clasificación de objetos responde a un criterio de jerarquía.	Posee las características anteriores. Además, el cero y la unidad de medida son arbitrarios.	Posee todas las características anteriores. Pero, el cero es el origen de los valores de la escala y la distancia entre valores asignados tiene significado.
Propiedades formales	Relación de equivalencias ($=$ o \neq)	Relación de equivalencias ($=$ o \neq) Relación de orden mayor que ($>$)	Relación de equivalencia y relación de orden mayor que ($>$). Además, $y = a + bx$	Relación de equivalencia y relación de orden mayor que ($>$). Además. $y = bx$
Operaciones aritméticas	No	No	Si	Si

2.8. Tipos de variables

Cuando se realiza una investigación es necesario tener bien definidas nuestras variables y conocer sus características. Cuando establecemos adecuadamente el tipo de variable que utilizaremos en nuestro estudio la selección del método estadístico representa un proceso más objetivo. No todos los procedimientos estadísticos son aplicables a todos los tipos de variables. Por ello es importante conocer los tipos de variables de tal manera

que el análisis de nuestros resultados sea coherente con nuestros objetivos de investigación.

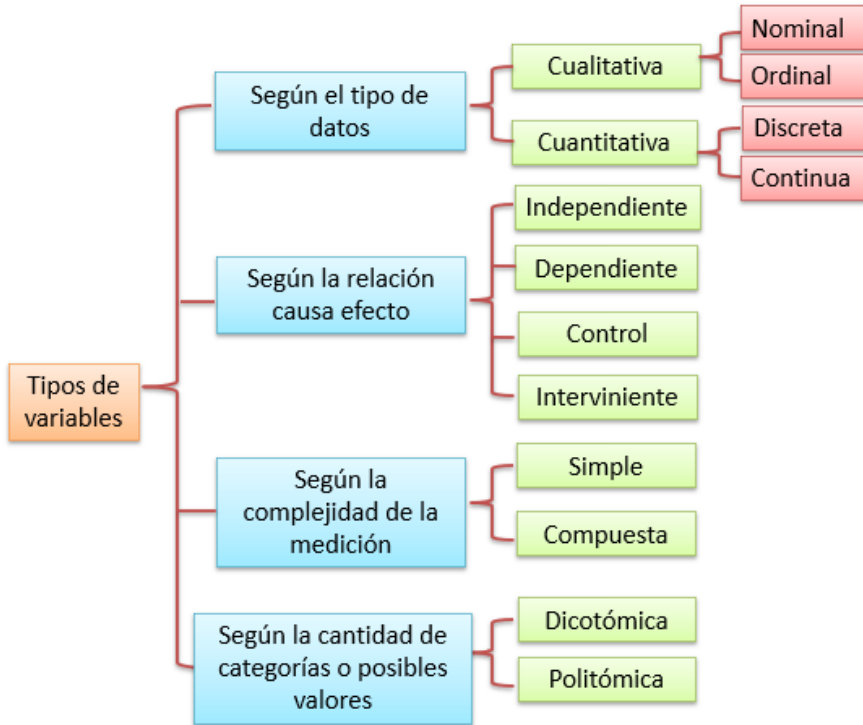


Figura 2.14 Clasificación de las variables.

Según el carácter numérico o no numérico de los datos recabados.

Para medir una variable se necesitan datos. Estos datos, recolectados por medio de un instrumento, pueden estar representados por números o categorías. Por ello, de acuerdo al carácter numérico de los datos se puede diferenciar entre variables cualitativas y cuantitativas.

2.8.1. Variables cualitativas:

Son aquellas variables cuyos datos no son numéricos, pero pueden ser clasificados en categorías. Estas categorías son subconjuntos unitarios del grupo total de datos que se obtienen al medir este tipo de variables.

En el siguiente cuadro se muestra algunas variables cualitativas y sus respectivas categorías.

Tabla 2.11 Variables cualitativas y algunas de sus categorías.

Variable	Categorías
Tipo de papa de mayor consumo	Blanca, negra, huayro, amarilla y otros.
Tipo de producción	Artesanal, por lotes y en masa.
Calidad de un producto	Baja, media y alta
Método de cocción en seco	A la plancha, parrilla, brasa y al horno

Como se puede apreciar, las categorías de una variable representan los datos recolectados por medio de un instrumento. Si usamos una encuesta y preguntamos: ¿Qué tipo de papa consume con mayor frecuencia? La respuesta a esta interrogante es alguna de las categorías que se han considerado en el cuadro.

Las categorías pueden ser codificadas con números, pero no tiene sentido realizar operaciones matemáticas con ellos.

Si consideramos la variable tipo de carne preferida para el consumo familiar y codificamos los posibles datos:

- 1: Carne de pollo.
- 2: Carne de pescado.
- 3: Carne de cerdo.
- 4: Carne de res.

Luego intentamos calcular el promedio:

$$\frac{1 + 2 + 3 + 4}{4} = 2.5$$

En la codificación no existe el 2.5, es decir no hay un tipo de carne para el código 2.5. Entonces, es evidente que no tiene sentido realizar operaciones numéricas con los datos de la variable cualitativa.

Las variables cualitativas pueden ser nominales u ordinales.

2.8.1.1. Variable cualitativa nominal

Es aquella cuyas categorías no son factibles de ser clasificados a través de un criterio de jerarquía. Es decir, las categorías de la variable no se pueden ordenar por niveles, no existe una categoría con mayor jerarquía que otra.

Ejemplos:

- Sexo de los estudiantes (Masculino, femenino).
- Estado civil de una persona (Soltera, casada, viuda, etc.).
- Marca preferida de automóvil (Toyota, Nissan, Mazda, etc.).
- Tipo de comida (Criolla, andina, china, etc.).

2.8.1.2. Variable ordinal

Es aquella variable en la que sí se puede establecer un criterio de jerarquía entre sus categorías.

Ejemplos:

- Nivel de instrucción (básico, técnico, superior).
- Nivel socioeconómico (bajo, medio, alto).
- Rango militar (soldado, sargento, teniente, etc.).
- Dominio de un idioma extranjero (básico, intermedio, superior).

2.8.2. Variables cuantitativas

Son aquellas cuyos datos se expresan en forma numérica y tiene sentido realizar operaciones matemáticas con dichos números.

Si consideramos la variable masa corporal, los datos obtenidos son numéricos. Si asumimos la masa corporal de tres personas 68 kg, 72 kg y 74 kg, podemos calcular el promedio de la masa de las tres personas.

$$\frac{70 + 73 + 75}{3} = 72.67$$

Como se observa con los datos de las variables cuantitativas se pueden efectuar cálculos de valores como el promedio y otros valores más que se verán en los capítulos posteriores.

Las variables cuantitativas pueden ser discretas o continuas.

2.8.2.1. Variable cuantitativa discreta

Es aquella que cumple con la condición de que los datos solo pueden tomar un número finito de valores dentro de un intervalo.

Por ejemplo, si se considera la variable cantidad de docentes por universidad y tomamos dos datos registrados en el intervalo de 180 a 184, tiene sentido pensar que 181, 182 y 183 son otros posibles resultados.

Cantidad de docentes por universidad: 180 **181** **182** **183** 184

De acuerdo con la variable ¿Se podrán presentar más valores entre 180 y 184?

La respuesta es negativa, puesto que, no tiene sentido utilizar números decimales para contabilizar personas ¿182.5 personas? Entonces, se verifica la característica principal de la variable discreta que indica la existencia de un número finito de posibles observaciones dentro de un intervalo.

Ejemplos:

- Cantidad de tesis registradas por universidad
- Cantidad de clientes atendidos diariamente en un banco.
- Cantidad de alumnos por aula.
- Número de hijos.
- Número de piezas defectuosas en la producción diaria.

2.8.2.2. Variable cuantitativa continua

Es aquella cuyos datos pueden tomar cualquier valor numérico dentro de un intervalo continuo.

Si consideramos como variable el interés mensual generado en la cuenta de ahorros de un grupo de empleados y observamos el intervalo de S/ 400 a S/ 404 de interés mensual. Entre estos números pueden ubicarse S/ 401, S/ 402 y S/ 403.

Interés mensual: 400 **401** **402** **403** 404

De acuerdo con la variable ¿Se pueden encontrar más valores?

La respuesta es afirmativa ya que entre S/ 400 y S/ 404 también se encuentra el S/ 400.5 o el S/ 403.5.

Interés mensual: 400 **400.5** **401** ... **402** **403** **403.5** 404

Estos montos son válidos ya que representan el interés generado en unidades monetarias y se pueden encontrar aún más montos como por ejemplo el S/ 400.58 que representa un interés mensual generado de 400 soles con 58 céntimos.

Entonces, se verifica la característica principal de la variable discreta que indica la existencia de un número infinito de posibles observaciones dentro de un intervalo.


Ejemplos:

- Gasto mensual familiar (S/).
- Tiempo para llegar a la universidad.
- Precio de una computadora en una ciudad.
- Masa corporal.
- Estatura.

¿solo las variables continuas admiten valores decimales?

Si analizamos el costo por viaje para los usuarios del servicio de transporte “Metropolitano” en la ciudad de Lima, podemos ver que las tarifas expresadas en decimales, no necesariamente indican que la variable costo por viaje sea una variable continua.

	<i>General</i>	<i>Medio Pasaje</i>
<i>Alimentador</i>	<i>S/ 0.50</i>	<i>S/ 0.25</i>
<i>Troncal</i>	<i>S/ 2.50</i>	<i>S/ 1.25</i>
<i>Troncal - Alimentador</i>	<i>S/ 2.50</i>	<i>S/ 1.25</i>
<i>Alimentador - Troncal - Alimentador</i>	<i>S/ 2.50</i>	<i>S/ 1.25</i>



10 de noviembre de 2018

Figura 2.15 Tarifas del servicio de transporte “Metropolitano”. Recuperado de: <https://rpp.pe/lima/actualidad/metropolitano-anuncia-que-la-tarifa-de-s-250-se-retomara-este-lunes-12-noticia-1162344>

Se debe tener en cuenta que en el caso de la variable continua los datos son numéricos y pueden tomar cualquier valor dentro de cualquier intervalo, de lo contrario, la variable es discreta.

En la figura 2.15 se observan las tarifas por el servicio de transporte “Metropolitano”. Si se usa la ruta troncal y no se justifica el medio pasaje, se debe pagar S/ 2.50. Las cantidades son fijas, no existe la posibilidad de pagar S/ 2.10 o S/ 2.30. Es decir, no se puede tomar cualquier valor numérico dentro de un intervalo, por lo tanto, la variable costo por viaje en el servicio de transporte “Metropolitano” es una variable discreta.

Según la relación de causa y efecto entre las variables.

En algunas investigaciones se desea conocer cómo los valores de una variable influyen en los valores de otra variable. Es decir, se quiere determinar la relación causa-efecto entre dos variables. Por ejemplo, si queremos analizar cómo la cantidad de leche influye en la textura de un tipo de bizcochuelo, podemos identificar que la cantidad de leche es la causa y la textura del bizcochuelo es el efecto.

La relación causa-efecto no siempre se analiza solo entre dos variables. En una investigación puede existir más de una variable considerada como causa o más de una variable considerada como efecto. En general los tipos de estudios donde se identifican causas y efectos se ubican en el nivel explicativo de la investigación. Aquí abordaremos las variables independientes, dependientes, de control y las intervinientes.

2.8.3. Variable independiente

Es la variable que el investigador manipula para observar los cambios que produce en otra variable relacionada. También se denomina variable explicativa y se considera la causa de un determinado fenómeno. Dentro del diseño experimental se les conoce como factores y se les puede representar con la letra X. Por ejemplo, cuando en un estudio se desea conocer el efecto de la cantidad de polietileno en la resistencia de un tipo de empaque.

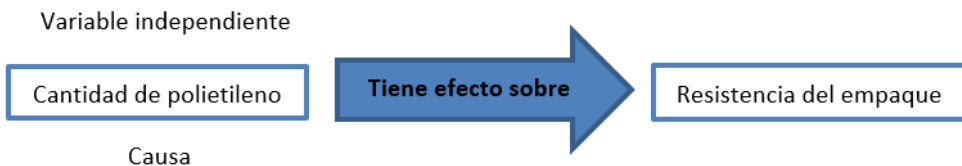


Figura 2.16 La variable independiente también es conocida como causa o factor.

Al realizar las pruebas del experimento la variación en la cantidad de polietileno determina una resistencia diferente para el empaque resultante en cada prueba.

2.8.4. Variable dependiente

Es la variable cuyos valores se espera que cambien a causa de la manipulación de la variable independiente. La manipulación de las variables las efectúa el investigador y responde a los objetivos de la investigación. También se denomina variable explicada, efecto o variable respuesta. Esta variable se representa con la letra Y.

En el ejemplo anterior, donde se quiere conocer el efecto de la cantidad de polietileno en la resistencia de un tipo de empaque, se puede identificar la variable dependiente.

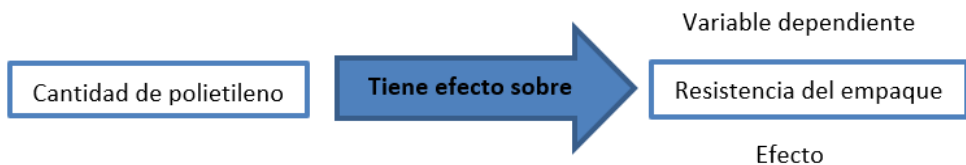


Figura 2.17 La variable dependiente también es conocida como respuesta o efecto.

2.8.5. Variable de control

Es aquella variable que tienen efecto sobre la variable dependiente pero no son parte del conjunto de variables que se desea analizar. Estas variables son manipulables y se reducen o suprimen para eliminar sus posibles efectos sobre la variable dependiente.

Por ejemplo, cuando en un diseño experimental se quiere precisar el efecto que tiene la cantidad de levadura en la altura de cierto tipo de pan, se sabe que existen otros factores que afectan el tamaño del pan resultante.

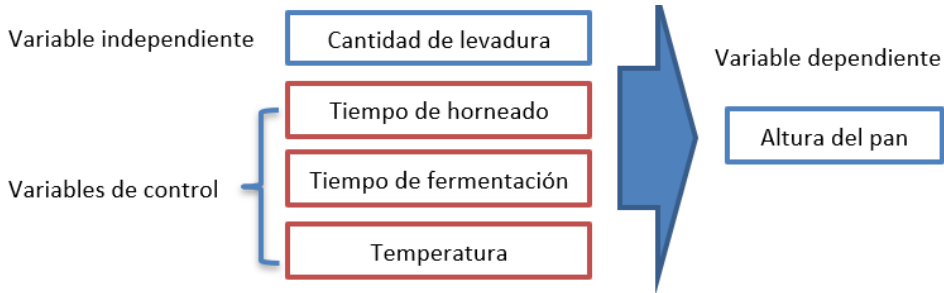


Figura 2.18 La variable de control tiene efecto sobre la variable dependiente, por ello, se debe controlar su efecto.

El tiempo de horneado, la temperatura y el tiempo de fermentación son variables de control, puesto que, tienen efecto sobre la altura del pan elaborado pero su efecto debe eliminarse para conocer únicamente el efecto que tiene la cantidad de levadura. El efecto de estas variables de control se elimina utilizando la misma temperatura y el mismo tiempo de horneado y fermentación para cada prueba del experimento.

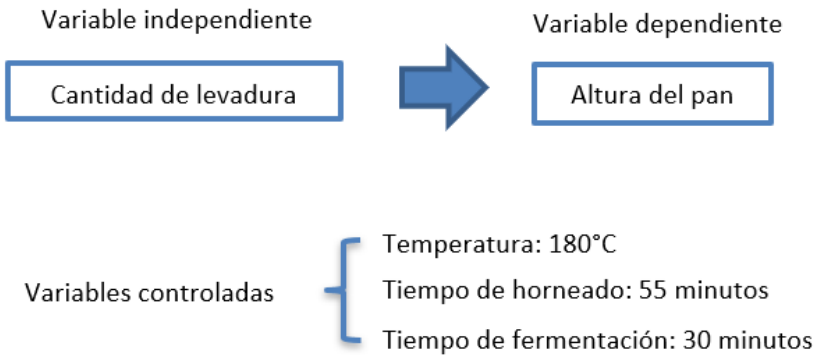


Figura 2.19 Las variables controladas tienen un valor constante durante todo el proceso. Por ello, solo se analiza el efecto de la cantidad de levadura.

2.8.6. Variable interviniente

Es aquella variable que tienen efecto sobre la variable dependiente, pero son difíciles de identificar o manipular. Estas variables no son de interés directo para analizar la relación causa efecto entre variables, pero están presentes en el proceso.

Dentro del diseño experimental se les conoce como factores no controlables. Este tipo de factores pueden ser ambientales, emocionales, factores de calidad, etc.

Por ejemplo, cuando se desea conocer el efecto del tipo de dieta en la masa corporal de un grupo de estudiantes se debe considerar la existencia de otros factores.

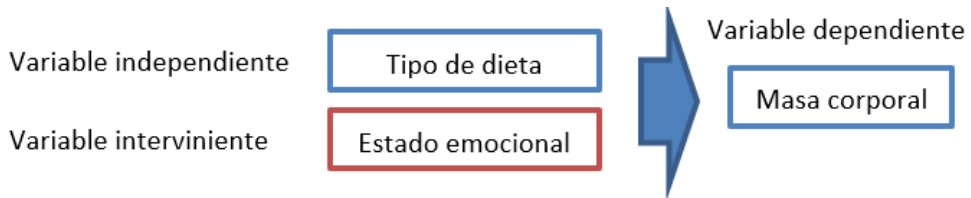


Figura 2.20 La variable interviniente tiene efecto sobre la masa corporal y es difícil controlar su efecto.

La variación de la masa corporal también depende del estado emocional que tiene el estudiante para cumplir de forma estricta con la dieta que le asigna el investigador. Sin embargo, es difícil cuantificar y manipular el estado emocional.

Para los diseños experimentales se puede considerar el siguiente esquema:

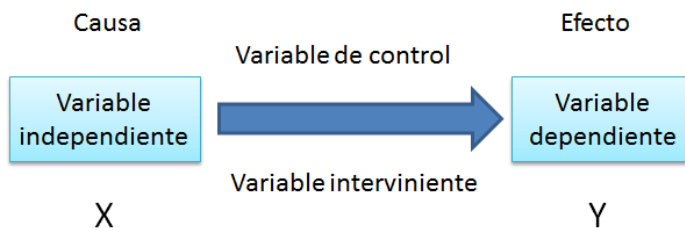


Figura 2.21 Las variables dentro del proceso de experimentación.

Según la complejidad para efectuar la medición

Existen variables cuyo proceso de medición es tan simple que basta con hacer una sola pregunta u observación. Sin embargo, otras variables requieren de mayor información para establecer una medición. La variable

masa corporal es fácil de medir y solo se requiere colocar al individuo sobre una balanza y registrar la información. La variable hábitos de consumo, en cambio, requiere de una serie de observaciones para medirla adecuadamente (tipos de productos que se consume, frecuencia con que los consume, lugar donde los consume, etc.). Por ello, se puede diferenciar dos grupos de variables: variables simples y variables compuestas.

2.8.7. Variable simple

Son aquellas variables cuya medición se realiza mediante un solo indicador (Soto, 2018). Es decir, cuando se elabora la matriz de datos y esta presenta una sola columna de datos recolectados, la variable que se puede medir con esa única columna de datos se denomina variable simple.

Por ejemplo, las siguientes son variables simples:

Individuo	Edad
1	19
2	20
3	19
4	22
5	20
6	19

Columna de datos para la variable edad.

N°	Tiempo de cocción (Minutos)
1	15
2	16
3	15
4	20
5	18
6	15
7	16

Columna de datos para la variable tiempo de cocción.

Figura 2.22 Las variables simples se miden mediante una sola observación por unidad de análisis.

Otros ejemplos de variables simples son:

- La masa corporal.
- La temperatura.
- El tipo de cocción.
- La cantidad de personal por restaurante.
- El ingreso mensual (S/).

2.8.8. Variable compuesta

Son aquellas variables en cuyo proceso de medición se requiere conocer las dimensiones o indicadores que componen la variable. Es decir, cuando en la matriz de datos se requiere más de una columna para poder medir la variable.

Por ejemplo, el IMC es una variable compuesta.

Individuo	Altura (m)	Masa corporal (Kg)	IMC
1	1.70	75	25.95
2	1.68	72	25.51
3	1.66	64	23.23
4	1.72	70	23.66
5	1.78	75	23.67
6	1.72	72	24.34

Para medir la variable índice de masa corporal (IMC) se requiere conocer la masa (M) y la estatura (E).

$$IMC = \frac{E}{M^2}$$

Figura 2.23 Observe que para medir el IMC se requiere de las dos columnas precedentes.

Otros ejemplos de variables compuestas son:

- El grado de estrés laboral.

- El nivel de autoestima.
- La satisfacción con respecto al servicio recibido.
- Los hábitos alimentarios.
- El nivel consumo de alimentos ultra procesados.

Según la cantidad de valores que pueden asumir los datos recolectados.

2.8.9. Variable dicotómica

Es aquella variable cuya medición presenta resultados que solo pueden tener dos valores o categorías, según sea la variable numérica o no numérica respectivamente.

Ejemplos:

- Experiencia laboral (sí, no).
- Preferencia por el producto “A” (sí, no).
- Afiliación al sistema de pensiones (estatal, privado).
- Forma de pago (crédito, al contado).

2.8.10. Variable politómica

Es aquella variable cuya medición arroja resultados que pueden tener más de dos valores.

Ejemplos:

- Edad (16, 19, 24, etc.).
- Profesión (médico, abogado, nutricionista, etc.).

- Nivel de conocimientos (bajo, regular, alto).
- Horas de sueño (5, 6, 7, etc.).

¿Cómo clasifico mis variables de investigación?

Se ha mostrado que existen diferentes formas de clasificar una variable. Por ello, una variable puede recibir diferentes denominaciones. Por ejemplo, la variable estado civil, es una variable nominal, simple, politómica, incluso puede ser una variable independiente.

Para clasificar a nuestras variables nos podemos orientar según el nivel de la investigación. Es decir, si se realiza una investigación descriptiva o relacional, no tiene sentido clasificar a las variables como independientes o dependientes. En cambio, si se realiza una investigación explicativa, entonces si se deben identificar adecuadamente la variable independiente, dependiente, de control e interviniente. Es importante reconocer el tipo de variable, puesto que la correcta identificación de la variable permite la elección del método estadístico más adecuado para la investigación.

Por ejemplo, si analizamos la variable índice de masa corporal:

Tabla 2.12 Clasificación de la variable IMC.

Variable	Tipo	Consideraciones
Índice de masa corporal	Continua	Se pueden calcular e interpretar todos los estadígrafos.
	Compuesta	Se debe recolectar diferentes datos para una misma unidad de análisis.
	Politómica	La información se organiza en tablas con más de dos categorías.
	Dependiente o Independiente	Esta denominación se utiliza solo si la investigación es explicativa.

Si analizamos la variable frecuencia de consumo del producto X:

Tabla 2.13 Clasificación de la variable frecuencia de consumo.

Variable	Tipo	Consideraciones
Frecuencia de consumo	Ordinal	Se puede calcular e interpretar la moda, el tanto por ciento y la mediana.
	Simple	Los datos se obtienen por medio de una sola pregunta.
	Politómica	La información se organiza en tablas con más de dos categorías.
	Dependiente o Independiente	Esta denominación se utiliza solo si la investigación es explicativa.

Dentro del trabajo de investigación, no es indispensable redactar la clasificación completa de nuestras variables de estudio como la clasificación realizada en las tablas 2.12 y 2.13. Sin embargo, es necesario reconocer el tipo de variable para dar un tratamiento adecuado a la información que se ha de recolectar.

Capítulo 3: Tablas de distribución de frecuencias

Contenido

- ❖ Elementos de una tabla estadística.
- ❖ Tabla de distribución de frecuencias para variable cualitativa.
- ❖ Tabla de distribución de frecuencias para variable cuantitativa.
- ❖ Tabla de distribución de frecuencias con intervalos.
- ❖ Tabla de contingencia.

Una forma simple de presentar la información es por medio de una tabla de distribución de frecuencias. Esta tabla indica el número de unidades de análisis que pertenecen a cada una de las clases de la variable cuantitativa o cualitativa. También se le puede definir como una lista valores de datos de forma individual o por intervalos y sus respectivos conteos (Triola, 2004).

Luego de la recolección de los datos se considera que el primer paso para el estudio estadístico de una muestra es su organización y presentación en una tabla de frecuencias (Gorgas, et al., 2011).

La elaboración de tablas de distribución de frecuencias va a depender de la cantidad de variables que se analizan. Si se estudia una sola variable se trata de una distribución de frecuencias univariada, si se estudia 2 variables en base a una sola muestra se trata de una distribución de frecuencias bivariada. Cuando se estudia más de dos variables se trata de una distribución multivariada (Vargas, 2007). En este libro trabajaremos con distribuciones univariadas y bivariadas.

Ejemplo: Se registra la cantidad de celulares que han tenido los estudiantes del quinto ciclo de una universidad hasta la actualidad y se obtienen los siguientes datos:

3	3	3	5	4	5
4	5	3	5	5	5
6	5	4	5	4	5
6	5	4	5	4	5

Los datos, así presentados, no permiten establecer una conclusión rápida y adecuada, esto se complica más cuando la muestra es grande. Sin embargo, se puede observar que muchos datos se repiten. Por ello, estos datos se pueden agrupar con la finalidad de contabilizar la cantidad de veces que aparecen los valores en la muestra.

Tabla 3.1 Tabla de distribución de frecuencias.

Cantidad de celulares	Frecuencia
3	4
4	6
5	12
6	2
Total	24

La tabla 3.1 representa la tabla de distribución de frecuencias más simple. Como se puede ver, la frecuencia determina la cantidad de veces que se contabiliza un dato. Nótese, también que los datos se han organizado en forma ascendente. Esto último es importante para la construcción e interpretación de las demás frecuencias.

De la misma forma se puede construir tablas de distribución de frecuencias para variables cualitativas. A diferencia del caso anterior donde la variable es numérica, aquí no existe un orden determinado para ubicar las categorías.

Ejemplo: Se encuesta a un grupo de gerentes sobre el tipo de ceviche de su preferencia y se obtiene la siguiente información:

Mixto	Pescado	Pescado	Mixto	Pescado
Mariscos	Conchas negras	Camarones	Pulpo	Mixto
Pescado	Conchas negras	Pescado	Mixto	Camarones
Mariscos	Pescado	Mixto	Pescado	Mixto
Pescado	Pulpo	Mariscos	Mixto	Pulpo

Como en el caso anterior muchas respuestas se repiten. Por lo tanto, agrupamos los datos y realizamos el conteo.

Tabla 3.2 Tabla de distribución de frecuencias.


Tipo de ceviche	Frecuencia
Pescado	8
Mixto	7
Pulpo	3
Conchas negras	2
Camarones	2
Mariscos	3
Total	25

Con la información ya organizada en una tabla de distribución de frecuencias es más sencillo establecer las interpretaciones. En adelante a esta tabla básica se le agregarán otras frecuencias según el tipo de variable que se desea analizar.

3.1. Elementos de una tabla

Una tabla estadística debe contener ciertos elementos básicos que hacen comprensible la lectura correcta de la información presentada. Los elementos básicos son los siguientes: El título, los encabezados de las columnas, los encabezados de las filas, notas al pie y la fuente (Naciones Unidas, 2009). A estos elementos se le puede agregar el código de la tabla.

Código: Las tablas estadísticas deben presentar una numeración para establecer una correcta referencia cuando se realiza la presentación e interpretación de resultados.

Código  **Tabla 6**


Nivel de instrucción	Porcentaje
Básico	24%
Técnico	46%
Superior	30%
Total	100%

Figura 3.1 Las tablas siempre deben estar enumeradas.

El título: Describe de forma clara y precisa la información que contiene la tabla. Además, como sugiere el INEI y la Naciones Unidas el título debe responder a las preguntas ¿Qué? ¿Cómo?, ¿Dónde? Y ¿Cuándo? Además, en la redacción del título se debe evitar el uso de verbos.

Ejemplos:

- Ingreso familiar según distrito de residencia en Lima, 2018.
- Nivel de estrés laboral según la ocupación en Cusco, junio 2015.
- Consumo de energía eléctrica según tipo de vivienda en Lima, 2018-2019

Cuadro N°5  Título
LA LIBERTAD: PRODUCCIÓN DE PRINCIPALES PRODUCTOS AGROPECUARIOS,
2018-2019
 (Toneladas)


Producto	2018	2019	Var. (%)
Caña de azúcar	4 795 513	5 514 278	15,0
Papa	496 524	540 771	8,9
Arroz cáscara	387 546	328 797	-15,2
Espárrago	134 200	137 588	2,5
Arándano	75 114	115 635	53,9
Ave	364 683	386 472	6,0
Huevos	76 210	84 438	10,8

Fuente: Ministerio de Agricultura y Riego (MINAGRI).

Figura 3.2 El título de una tabla resume el contenido mostrado.

Encabezados de las columnas: El encabezado de cada columna debe indicar el significado de los datos que están en dicha columna. Estos encabezados pueden indicar la cantidad de observaciones, unidades de medida, tanto por ciento, etc.

Encabezados

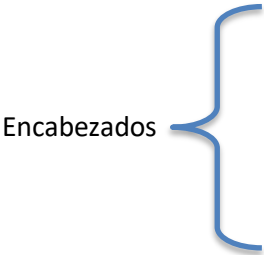


Variable	Frecuencia	Frecuencia porcentual	Frecuencia % acumulada
Total			

Figura 3.3 Los encabezados describen el contenido de cada columna.

Encabezados de las filas: De la misma forma que los encabezados de las columnas, en los encabezados de cada fila se debe describir la característica general de los datos de la fila.

Tabla 7



Nivel de instrucción	Frecuencia	Porcentaje
Básico incompleto		
Básico completo		
Técnico incompleto		
Técnico completo		
Superior incompleto		
Superior completo		
Total		

Figura 3.4 Los encabezados describen el contenido de cada fila.

Notas al pie: Se ubican en la parte inferior de la tabla y contienen información que ayuda a comprender el contenido de la tabla.

Fuente: En la parte inferior de la tabla se debe colocar la fuente de información, es decir, que organización o investigadores elaboraron dicha tabla.

A continuación, se muestra una tabla elaborada por el INEI (2020) con todos los elementos de la tabla descritos anteriormente.

Tabla 12
ÍNDICES DE PRECIOS DE COMERCIO EXTERIOR
ABRIL 2020
 (BASE: AÑO 2002 = 100)

ÍNDICE DE PRECIOS	NÚMERO ÍNDICE	VARIACIÓN PORCENTUAL ABRIL 2020 ^{a/}
DE EXPORTACIÓN (IPEX)	233,03	-14,84
DE IMPORTACIÓN (IPIM)	132,15	-7,24

a/ Con respecto a similar mes del año anterior

Fuente: Instituto Nacional de Estadística e Informática.

 **Fuente**

Figura 3.5 La fuente informa sobre el autor de la tabla. Si no coloca la fuente, se asume que es una elaboración propia.

Existen diferentes tipos de tablas estadísticas, el uso de cada una de ellas depende del tipo de variable, de la información que se quiere mostrar y del tipo de análisis. En este trabajo se organiza la clasificación de las tablas según los criterios expuestos. Por ello, se aborda las características de las tablas para variables cualitativas, cuantitativas y las tablas de contingencia.

3.2. Tabla de distribución de frecuencias para variable cualitativa

Cuando se analiza variables cualitativas lo más conveniente es utilizar la frecuencia absoluta, la frecuencia relativa y la frecuencia relativa porcentual, ya que con estas frecuencias se puede describir las características básicas de la muestra de forma coherente. Existen otras frecuencias, sin embargo, no son útiles cuando se realizan las interpretaciones.

Frecuencia absoluta (f_i): Indica el número de elementos que pertenecen a una determinada categoría o clase.

Frecuencia relativa (h_i): Nos indica que fracción del total se encuentra en cada clase. Se obtiene a través del cociente de la frecuencia absoluta y el total de elementos.

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

Donde “n” es el tamaño de la muestra.

Frecuencia relativa porcentual ($h_i\%$): Es el porcentaje que representa cada frecuencia absoluta con respecto al total. Se obtiene multiplicando la frecuencia relativa por 100%.

$$h_i\% = h_i \times 100\%$$

Ejemplo: Para conocer las características del grupo de estudiantes que se han matriculado de manera regular se verifica el registro de matrículas y se obtiene la siguiente información:

	Carrera	Estudiantes matriculados	
Clases	Ingeniería	40	Frecuencias absolutas
	Administración	50	
	Gastronomía	48	
	Nutrición	32	
	Contabilidad	30	
	Total	200	

Figura 3.6 La frecuencia absoluta es la base de donde se obtienen los demás tipos de frecuencias.

Con esta información se puede elaborar la tabla de distribución de frecuencias.

Tabla 3.3 Cálculo de la frecuencia relativa y la frecuencia porcentual

Carrera	f_i	h_i	$h_i \%$
Ingeniería	40	$40/200 = 0.20$	$0.20 \times 100 = 20$
Administración	50	$50/200 = 0.25$	$0.25 \times 100 = 25$
Gastronomía	48	$48/200 = 0.24$	$0.24 \times 100 = 24$
Nutrición	32	$32/200 = 0.16$	$0.16 \times 100 = 16$
Contabilidad	30	$30/200 = 0.15$	$0.15 \times 100 = 15$
Total	200	1	100

Finalmente, solo consideramos los resultados.

Tabla 3.4 Tabla de distribución de frecuencias.

Carrera	f_i	h_i	$h_i \%$
Ingeniería	40	0.20	20
Administración	50	0.25	25
Gastronomía	48	0.24	24
Nutrición	32	0.16	16
Contabilidad	30	0.15	15
Total	200	1	100

De la tabla 3.4 se pueden extraer interpretaciones de cada frecuencia. Lo ideal es que se interprete la información más resaltante. No es necesario interpretar cada una de las frecuencias.

Algunas interpretaciones:

Para esto se debe tener presente que los subíndices mostrados nos indican el número de la fila. Por ejemplo, h_3 representa a la frecuencia relativa en la fila 3.

$h_2\%$: 25% de los estudiantes matriculados son de la carrera de administración.

h_5 : 15% de los estudiantes matriculados son de la carrera de contabilidad.

h_1 : ($40/200=1/5$) 1 de cada 5 estudiantes matriculados es de la carrera de ingeniería.

h_2 : ($50/200=1/4$) la cuarta parte de los estudiantes matriculados son de la carrera de administración.

f_4 : 32 estudiantes matriculados son de la carrera de nutrición.

f_3 : 48 estudiantes matriculados son de la carrera de gastronomía.

En muchas investigaciones solo se muestra e interpreta la frecuencia absoluta y la relativa porcentual (tabla 3.5). La tabla queda reducida con la finalidad de que los lectores puedan entender mejor los resultados obtenidos.

Tabla 3.5 Distribución de estudiantes matriculados por carrera.

Carrera	Frecuencia	Frecuencia porcentual
Ingeniería	40	20
Administración	50	25
Gastronomía	48	24
Nutrición	32	16
Contabilidad	30	15
Total	200	100

3.3. Tabla de distribución de frecuencias sin intervalos para variable cuantitativa

Para las variables con datos numéricos se pueden utilizar, además de las frecuencias anteriores, las frecuencias acumuladas. Estas frecuencias permiten realizar interpretaciones sobre un grupo de clases. Con las frecuencias anteriores se podía realizar una interpretación para cada clase. Con las frecuencias acumuladas se puede interpretar, en conjunto, clases consecutivas desde la primera clase.

Frecuencia absoluta acumulada (F_i): Es la acumulación ordenada de cada una de las frecuencias absolutas.

$$F_i = f_1 + f_2 + f_3 + \dots + f_i$$

La última frecuencia absoluta acumulada debe ser igual al tamaño de la muestra.

Frecuencia relativa acumulada (H_i): Es la acumulación ordenada de las frecuencias relativas.

$$H_i = h_1 + h_2 + h_3 + \dots + h_i$$

La última frecuencia relativa acumulada debe ser igual a 1.

Frecuencia relativa porcentual acumulada ($H_i\%$): Es la acumulación ordenada de las frecuencias relativas porcentuales.

$$H_i\% = h_1\% + h_2\% + h_3\% + \dots + h_i\%$$

La última frecuencia relativa porcentual acumulada debe ser igual a 100.

Ejemplo: Se realizó una encuesta con la finalidad de conocer la cantidad de hijos de los clientes de una zapatería.

Tabla 3.6 Cantidad de hijos por cliente.

Cantidad de hijos	f_i
0	14
1	60
2	80
3	70
4	26
n=250	

Con esta información se puede completar la tabla con las demás frecuencias

Tabla 3.7 Distribución de frecuencias de la cantidad de hijos por cliente.

Cantidad de hijos	f_i	F_i	h_i	H_i	$h_i \%$	$H_i \%$
0	10	10	0.04	0.04	4	4
1	60	70	0.24	0.28	24	28
2	80	150	0.32	0.60	32	60
3	70	220	0.28	0.88	28	88
4	30	250	0.12	1	12	100
	n=250		1		100	

Observe que los valores de F_i , H_i y $H_i \%$ se obtienen sumando los valores de f_i , h_i y $h_i \%$ respectivamente.

Algunas interpretaciones:

$h_1 \%$: El 4% de los clientes no tiene hijos.

$h_3 \%$: El 32% de los clientes tiene 2 hijos.

$H_3 \%$: El 60% de los clientes tiene como máximo 2 hijos.

F_4 : 220 clientes tienen menos de 4 hijos.

H_3 : 6 de cada 10 clientes tiene menos de 3 hijos.

En los trabajos de investigación generalmente se utiliza la frecuencia absoluta, la frecuencia porcentual y la frecuencia porcentual acumulada. Aunque esto depende de cómo se quiere presentar la información. Es decir, el investigador debe seleccionar el tipo de frecuencias e interpretar sus resultados de tal manera que sea coherente con sus objetivos. Por ello, la tabla se puede resumir y presentarse de la siguiente forma:

Tabla 3.8 Tabla de distribución de frecuencias.

Cantidad de hijos	frecuencia	Porcentaje	Porcentaje acumulado
0	10	4	4
1	60	24	28
2	80	32	60
3	70	28	88
4	30	12	100
Total	250	100	

3.4. Tabla de distribución de frecuencias con intervalos

En estas tablas se utilizan intervalos debido a que las variables son cuantitativas continuas o existe una variedad muy grande entre los valores numéricos de los datos recolectados y es necesario agruparlos, pues de otra forma la tabla tendría una cantidad muy grande de categorías.

Se debe recordar que un intervalo es una parte del conjunto de los números reales. Por ejemplo, el intervalo $[4; 8)$ contiene a los números enteros 4, 5, 6 y 7. Pero este intervalo contiene muchos más números como el 4.2 o el 5.56.

La notación que se ha utilizado en el ejemplo servirá para el desarrollo de las tablas de distribución con intervalos de clase.

$\langle 3; 8 \rangle$ Contiene números reales entre 3 y 8.

$[3; 8]$ Contiene números reales del 3 al 8.

$\langle 3; 8]$ Contiene números reales mayores que 3 y menores o iguales a 8.

$[3; 8)$ Contiene números reales mayores o iguales a 3 pero menores que 8.

Una tabla de distribución de frecuencias con intervalos de clase tiene la siguiente forma:

	Consumo de galletas (S/)	Frecuencia
Intervalos de clase	[2, 4>	8
	[4, 6>	6
	[6, 8>	7
	[8, 10>	7
	[10, 12>	5
	[12, 14]	3
		36

Figura 3.7 Tabla de distribución de frecuencias con intervalos.

Procedimiento para la elaboración de los intervalos de clase

Para construir una tabla de distribución de frecuencias con intervalos se deben tener en cuenta los siguientes elementos:

Cantidad de datos o tamaño de la muestra (n): Es el total de unidades estudiadas.

Alcance (A): Es el intervalo cerrado que tiene como extremos el menor y el mayor dato.

$$A = [a; b]$$

Rango (R): Es la diferencia entre el dato mayor y el menor.

$$R = b - a$$

Intervalo de Clase (I_i): Es un subgrupo de los datos y tiene como extremos al límite inferior y superior.

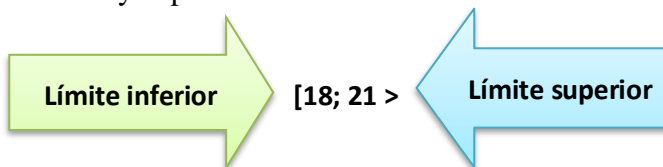


Figura 3.8 El intervalo de clase es semiabierto.

Número de intervalos (K): Para calcular un valor aproximado del número de clases se utiliza la “Regla de Sturges”.

$$k = 1 + 3.3 \log(n)$$

Existen otras formas de calcular el número de intervalos. Una forma distinta de calcular k es utilizando el entero más cercano a la raíz del tamaño de la muestra (Gorgas et al., 2011).

$$k = \sqrt{n}$$

Sin embargo, en los ejercicios de este libro se utilizará la regla de Sturges. Hay que tener en cuenta que la elección de la cantidad de intervalos depende del criterio del investigador y los objetivos de la investigación.

Amplitud o ancho de clase (W): Es la diferencia entre el límite superior e inferior de un intervalo de clase. La amplitud de los intervalos de clase es constante. Se obtiene de la siguiente forma:

$$w = \frac{R}{k}$$

Marca de clase (Xi): Es el punto medio de cada intervalo. Se obtiene a través de la semisuma de los límites en cada intervalo.

Por ejemplo: Para el intervalo [12; 18 >

$$x_i = \frac{12 + 18}{2} = 15$$

Ejemplo: Una empresa de turismo realiza una encuesta online con la finalidad de conocer el gasto (S/) que realizan las personas por un viaje “Full Day” a alguna provincia de Lima. Los datos obtenidos fueron los siguientes:

150	120	160	140	140	130	200	90
120	150	100	120	100	125	135	100
145	200	155	160	170	150	210	140
165	130	135	120	140	145	160	140
150	155	140	140	155	140	140	150

Para elaborar los intervalos de clase primero debemos determinar todos los elementos.

Tamaño de la muestra: $n = 40$, pues hay 40 datos.

Alcance: $A = [90; 210]$, pues 90 es el menor dato y 210 es el mayor dato.

Rango: $R = 210 - 90 = 120$

Número de intervalos: $k = 1 + 3.3 \log(40) = 6.287 \approx 6$

Amplitud o ancho de clase: $w = \frac{120}{6} = 20$

Luego, construimos los 6 intervalos de igual amplitud (20) considerando el 90 como límite inferior del primer intervalo y 210 como el límite superior del último intervalo.

		Gasto (S/)
6 intervalos de clase	}	[90; >
		[; >
		[; >
		[; >
		[; >
		[;210]

Figura 3.9 Observe que el último intervalo de clase es cerrado.

Los demás límites de los intervalos se obtienen sumando el ancho de clase 20 al límite anterior.

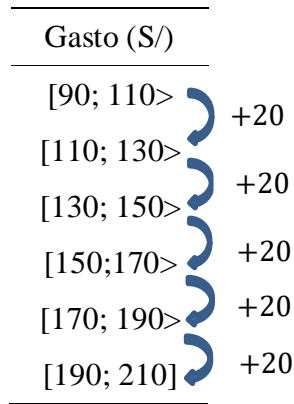


Figura 3.10 El límite inferior de cada intervalo es igual al límite superior del intervalo anterior.

La variable gasto es una variable continua y entre un valor y otro siempre existe otro valor, por ello los intervalos son semiabiertos. Por ejemplo, si en el primer intervalo colocamos [90; 109] y en el segundo intervalo [110; 129] no se podría ubicar el valor 109.5. Este último valor no está entre los datos, pero es un posible valor numérico para la variable gasto.

Una vez contruidos los intervalos se procede a ubicar los datos en su respectivo intervalo y luego se realiza el conteo.

150	120	160	140	140	130	200	90
120	150	100	120	100	125	135	100
145	200	155	160	170	150	210	140
165	130	135	120	140	145	160	140
150	155	140	140	155	140	140	150

Figura 3.11 Los datos han sido señalados de un color diferente para hacer más sencillo el conteo.

En el primer intervalo contamos a los números mayores o iguales a 90 pero menores que 110.

En el segundo intervalo contamos a los números mayores o iguales a 110 pero menores que 130.

En el tercer intervalo contamos a los números mayores o iguales a 130 pero menores que 150.

En el cuarto intervalo contamos a los números mayores o iguales a 150 pero menores que 170.

En el quinto intervalo contamos a los números mayores o iguales a 170 pero menores que 190.

En el sexto intervalo contamos a los números mayores o iguales a 190 hasta el 210.

Tabla 3.9 Frecuencia absoluta por intervalo de clase.

Gasto (S/)	f_i
[90; 110>	4
[110; 130>	5
[130; 150>	15
[150; 170>	12
[170; 190>	1
[190; 210]	3
Total	n=40

Con base en la información de la tabla anterior se puede completar la tabla de distribución de frecuencias.

Tabla 3.10 Distribución de frecuencias del gasto.

Gasto (S/)	f_i	X_i	F_i	h_i	H_i	$h_i\%$	$H_i\%$
[90; 110>	4	100	4	0.100	0.100	10	10
[110; 130>	5	120	9	0.125	0.225	12.5	22.5
[130; 150>	15	140	24	0.375	0.600	37.5	60
[150; 170>	12	160	36	0.300	0.900	30	90
[170; 190>	1	180	37	0.025	0.925	2.5	92.5
[190; 210]	3	200	40	0.075	1	7.5	100
Total	n=40			1		100	

Algunas de las frecuencias de la tabla 3.10 servirán para determinar estadígrafos como las medidas de tendencia central y las medidas de dispersión. Sin embargo, para la presentación inicial de los resultados se puede utilizar una tabla más resumida.

Tabla 3.11 Gasto por viaje turístico “Full Day”.

Gasto (S/)	f_i	$h_i\%$	$H_i\%$
[90; 110>	4	10	10
[110; 130>	5	12.5	22.5
[130; 150>	15	37.5	60
[150; 170>	12	30	90
[170; 190>	1	2.5	92.5
[190; 210]	3	7.5	100
Total	n=40	100	

De esta tabla se pueden extraer algunas interpretaciones. Por ejemplo.

$h_3\%$: El 37.5% de los encuestados gasta como mínimo S/ 130, pero menos de S/ 150.

H₃ %: El 60% de los encuestados gasta menos de S/ 150.

H₄ %: El 90% de los encuestados gasta menos de S/ 170.

3.5. Tabla de contingencia

Es una tabla de doble entrada en cuyo margen superior y margen izquierdo se sitúan las diferentes categorías de las variables. También se le llama tabla cruzada y es utilizada para la presentación de distribuciones bidimensionales, es decir cuando se analizan dos variables en una misma muestra. Por ejemplo, en una muestra de 80 niños podemos obtener datos acerca de su edad y su masa corporal.

En las casillas interiores de la tabla se ubica la frecuencia o número de casos y en las casillas marginales se ubican los subtotales. Se puede agregar la definición de distribución marginal entendida como aquella distribución donde solo se toma en cuenta una variable con respecto al total (Moya, 2007).

Ejemplo: En la siguiente tabla se cruza la información acerca del nivel educativo y el tipo de comida que se consume un grupo de comensales.

Tabla 3.12 Nivel educativo y el tipo de comida consumida.

	Comida rápida	Comida casera	Total
Estudiante	16	40	56
Técnico	32	28	60
Profesional	34	26	60
Total	82	94	176

En la tabla 3.12 se han utilizado frecuencias absolutas, pero la distribución bidimensional también se puede elaborar en base a frecuencias porcentuales.

Tabla 3.13 Nivel educativo y el tipo de comida consumida.

	Comida rápida	Comida casera	Total
Estudiante	9.09	22.73	31.82
Técnico	18.18	15.91	34.09
Profesional	19.32	14.77	34.09
Total	46.59	53.41	100

Interpretaciones:

En la tabla 3.13 se observa que la cantidad de comensales estudiantes, técnicos y profesionales es casi la misma. En cambio, la cantidad de comensales que prefieren consumir comida casera supera a la cantidad de comensales que optan por la comida rápida.

La cantidad de estudiantes que consume comida casera representa un 22.37% del total de comensales y conforma la mayor parte del grupo. En el grupo de las personas que consumen comida casera los técnicos y profesionales están representados el 15.91% y 14.77% del total de comensales.

La cantidad de estudiantes que consume comida rápida representa un 9.09% del total de comensales y conforma la menor parte del grupo. En el grupo de personas que consume comida rápida los técnicos y profesionales están distribuidos en similar proporción con un 18.18% y 19.32% del total de comensales.

De la tabla 3.13 se pueden extraer las distribuciones marginales referentes al nivel educativo y al tipo de comida.

Tabla 3.14 Distribución de la muestra por nivel educativo.

Categorías	Frecuencia porcentual
Estudiante	31.82%
Técnico	34.09%
Profesional	34.09%
Total	100%

Observe que en la distribución marginal de la tabla 3,14 solo se analiza la variable nivel educativo y en la distribución marginal de la tabla 3.15 solo se analiza la variable tipo de comida consumida.

Interpretaciones:

En la tabla 3.14 se observa que la proporción de comensales técnicos y profesionales es la misma 34.09%. En cambio, la cantidad de estudiantes representa una menor parte, alcanzando solo el 31.82%.

Tabla 3.15 Distribución de la muestra según el tipo de comida consumida.

Categorías	Frecuencia porcentual
Comida rápida	46.59%
Comida casera	53.41%
Total	100

Interpretaciones:

En la tabla 3.15 se observa que la cantidad de comensales que prefiere la comida casera representa el 53.41% del total de comensales, superando notoriamente al 46.59% conformado por los comensales que prefieren la comida rápida.

Capítulo 4: Gráficos estadísticos

Contenido

- ❖ Elementos de un gráfico estadístico
- ❖ Tipos de gráficos estadísticos
- ❖ Diagrama de barras
- ❖ Diagrama circular
- ❖ Histograma
- ❖ Polígono de frecuencias
- ❖ Diagrama escalonado
- ❖ Ojiva
- ❖ Gráfica de una distribución bidimensional
- ❖ Diagrama de dispersión
- ❖ Gráfica de líneas
- ❖ Criterios para la interpretación de tablas y gráficos estadísticos

La presentación de la información organizada se puede realizar por medio de tablas estadísticas como se ha visto en la sección anterior. Estas tablas resumen y facilitan la interpretación de los resultados. Sin embargo, La elaboración de gráficos es otra opción para realizar la presentación de los resultados. Algunos autores consideran que los gráficos estadísticos son una forma eficiente de transmitir la información desde la base de datos a la cabeza del lector (Naciones Unidas, 2009).

Los gráficos estadísticos son de mucha utilidad porque permiten visualizar el comportamiento de una muestra con respecto a la variable en estudio y entender los resultados, inclusive, con menos información numérica. Como se verá más adelante, si utilizamos un diagrama de barras y observamos que una barra tiene el doble de longitud que otra barra comprenderemos que el dato que tiene la barra de mayor longitud tiene mayor frecuencia.

Por ejemplo:

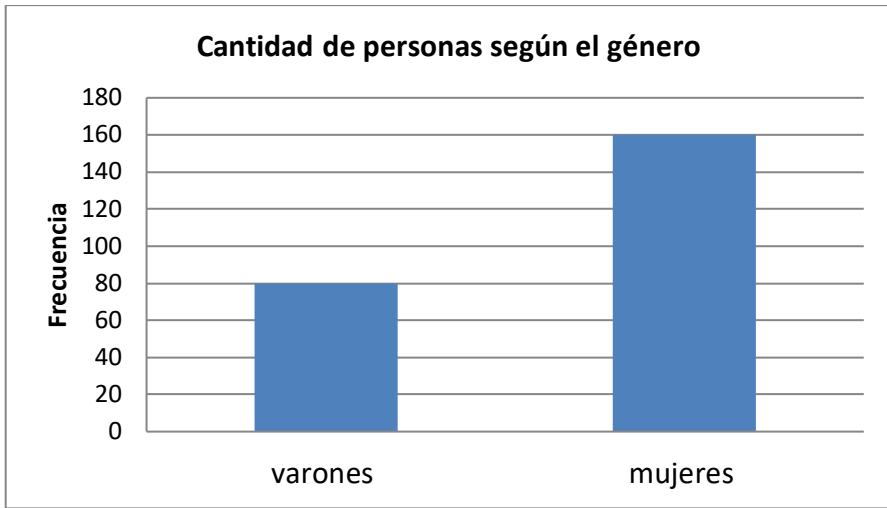


Figura 4.1 Diagrama de barras que permite comparar la cantidad de varones y mujeres.

Observe la figura 4.1, si quitamos los datos del eje de frecuencias, el gráfico resultante aún nos brinda información importante. En la figura 4.2 se puede apreciar que la cantidad de mujeres es el doble de la cantidad de varones.

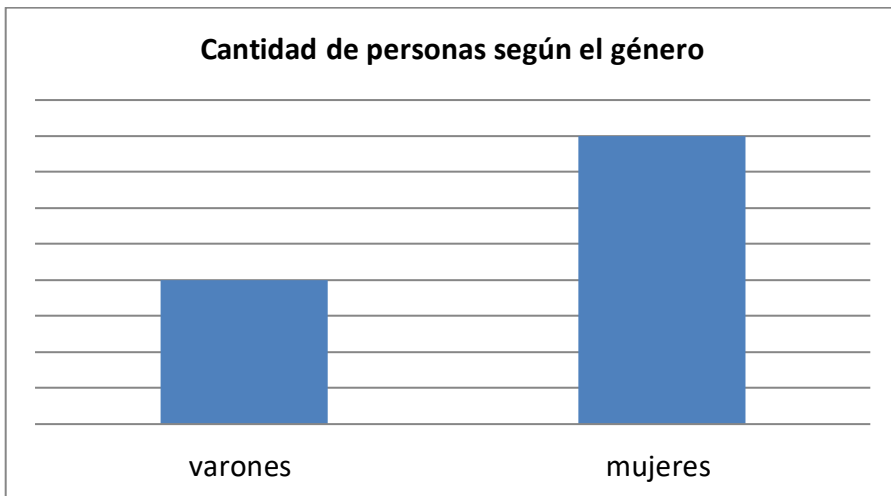


Figura 4.2 Sin usar números se puede concluir que la cantidad de mujeres duplica a la cantidad de varones.

El Instituto Nacional de Estadística e Informática (INEI) ha presentado una serie de orientaciones con la finalidad de estandarizar la presentación de información estadística y define a un gráfico estadístico como una representación visual por medio de puntos, líneas, barras, polígonos o figuras con la finalidad de facilitar la comprensión de los resultados (INEI, 2009)

4.1. Elementos de un gráfico estadístico

Código o número de gráfico: Como en el caso de las tablas estadísticas, es importante enumerar los gráficos para hacer una correcta referencia de estos en la etapa de presentación e interpretación de resultados.

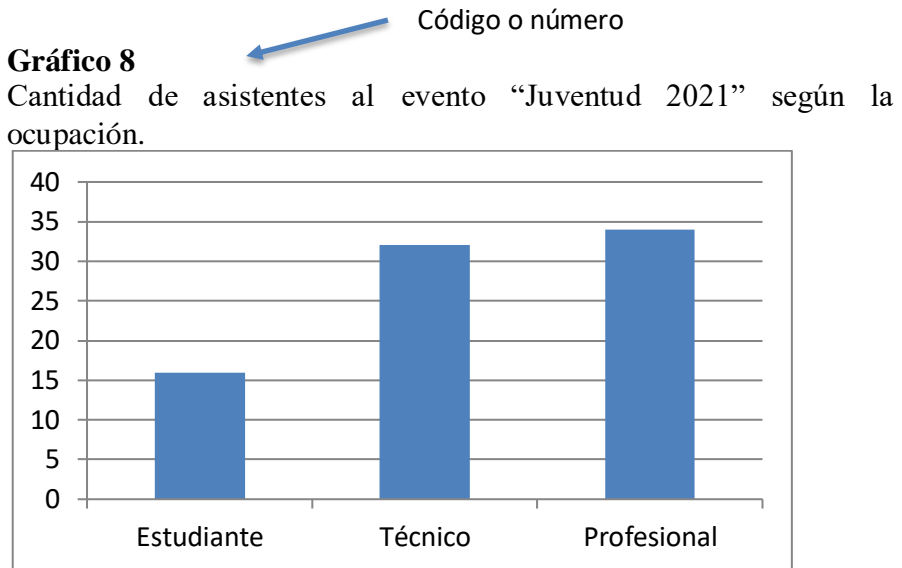


Figura 4.3 Los gráficos o figuras deben tener la numeración respectiva.

Título: Del mismo modo que el título para las tablas estadísticas, el título para un gráfico debe ser claro, preciso y debe responder a las preguntas ¿Qué? ¿Cómo?, ¿Dónde? Y ¿Cuándo? Además, en la redacción del título se debe evitar el uso de verbos.

El código y el título también pueden ubicarse en la parte inferior del gráfico. Esto depende de los protocolos de presentación de la investigación.

Ejemplos:

Cantidad de familias según el tipo de vivienda en Arequipa, 2017.

- ¿Qué? : Cantidad de familias
- ¿Cómo? : Según el tipo de vivienda
- ¿Dónde? : Arequipa
- ¿Cuándo? : 2017

Consumo de frutas y verduras según edades en Lima, 2015-2016.

- ¿Qué? : Consumo de frutas y verduras
- ¿Cómo? : Según edades
- ¿Dónde? : Lima
- ¿Cuándo? : 2015-2016

Cuerpo: Se compone de las figuras geométricas como puntos, líneas, círculos, rectángulos o polígonos asociados con los valores. En algunos casos, como en la elaboración del pictograma, se utilizan imágenes referentes a la temática.

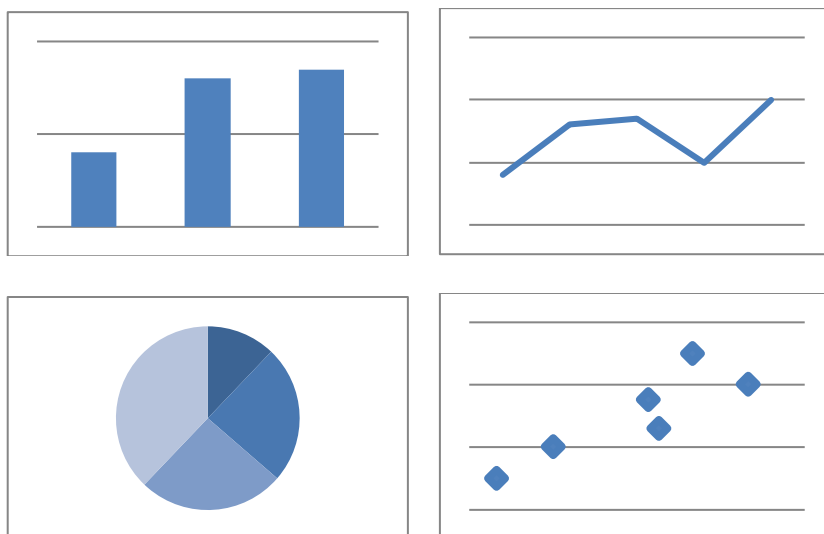


Figura 4.4 Existen diferentes figuras geométricas que se pueden asociar con las frecuencias.

Además, el cuerpo del gráfico presenta un eje horizontal y otro vertical asociado, cada uno de ellos, a las categorías y a las frecuencias respectivamente. De acuerdo con el INEI (2009), estos ejes se denominan eje de conceptos y eje de valores.

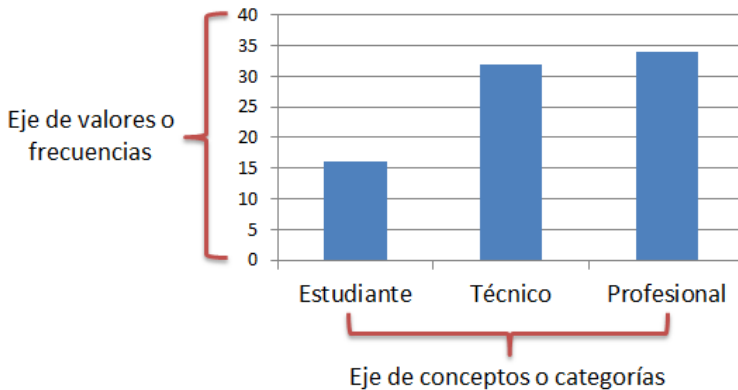


Figura 4.5 Los ejes de conceptos y valores permiten la interpretación del gráfico.

También existen otros recursos como la leyenda y la etiqueta que permiten la lectura más completa del gráfico. Las leyendas expresan el significado de las tramas o colores utilizados en los gráficos. Las etiquetas describen la frecuencia que representa cada línea, barra o sector circular de la figura.

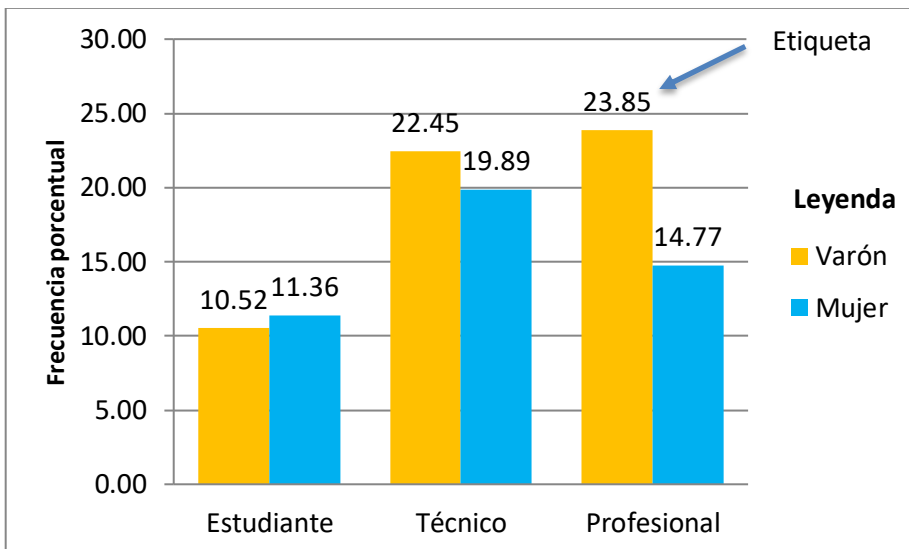
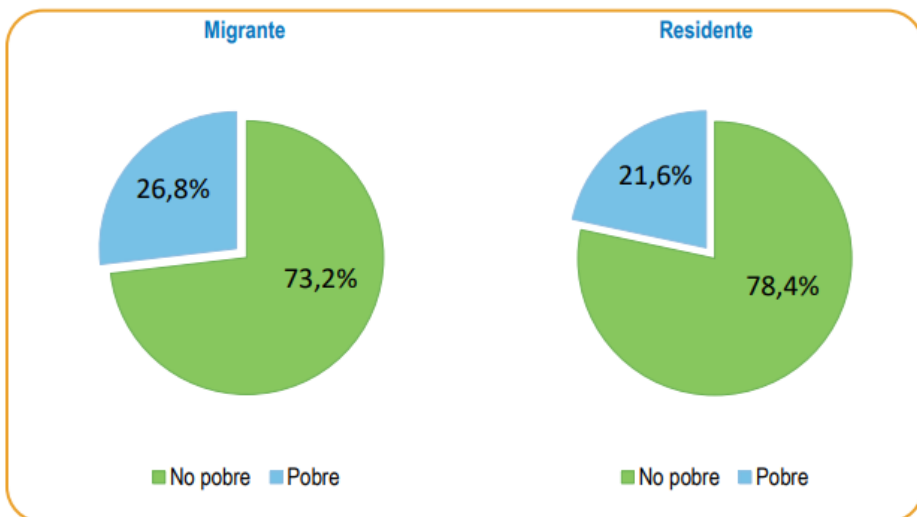


Figura 4.6 Las etiquetas indican la cantidad o proporción asociada a cada columna.

En la figura 4.6 se observa la leyenda que ayuda a identificar el color con el sexo, el color amarillo representa a los varones y el color celeste representa a las mujeres. También se observa las etiquetas, en este caso las etiquetas representan el tanto por ciento de personas en cada categoría puesto que en el eje de valores se indica que la frecuencia es porcentual.

Pie: Se ubica en la parte inferior de la tabla, aquí se pueden colocar las notas y la fuente. Como en el caso de las tablas estadísticas, las notas permiten entender mejor el gráfico y se utilizan cuando es necesario hacer algunas precisiones. La fuente indica el autor u organización que produjo la información.

Gráfico 9
DEPARTAMENTO DE LIMA: CONDICIÓN DE POBREZA MONETARIA DE LA POBLACIÓN
MIGRANTE RECIENTE EN HOGARES, 2017
 (Porcentaje)



Nota: La migración reciente relaciona el lugar de residencia cinco años antes con la residencia actual.
 Fuente: Instituto Nacional de Estadística e Informática - Censos Nacionales 2017: XII de Población y VII de Vivienda.

← Pie

Figura 4.7 En el pie se pueden colocar las notas o la fuente.

En la figura 4.7 se puede observar un gráfico estadístico con los elementos respectivos. De la misma forma que en el caso de las tablas, si no se escribe la fuente del gráfico se asume que es una elaboración propia.

4.2. Tipos de gráficos estadísticos

Existen diferentes gráficos estadísticos y la elección del gráfico adecuado para la presentación de la información depende de muchos factores. Por ejemplo, se debe tener en cuenta los objetivos de la investigación, la escala de medición y el tipo de variable.

Mediante un gráfico se puede hacer una comparación entre las frecuencias de las categorías, se puede conocer la distribución de los datos recolectados, se puede visualizar la proporción para cada componente de las variables, se puede establecer la correlación entre variables. También existen gráficos que permiten observar el comportamiento y la tendencia de una variable a través del tiempo.

4.3. Diagrama de barras

Se representa por medio de rectángulos en el primer cuadrante del plano cartesiano. Este tipo de gráfico se utiliza para el análisis de variables cualitativas. Los rectángulos deben estar separados por una longitud constante.

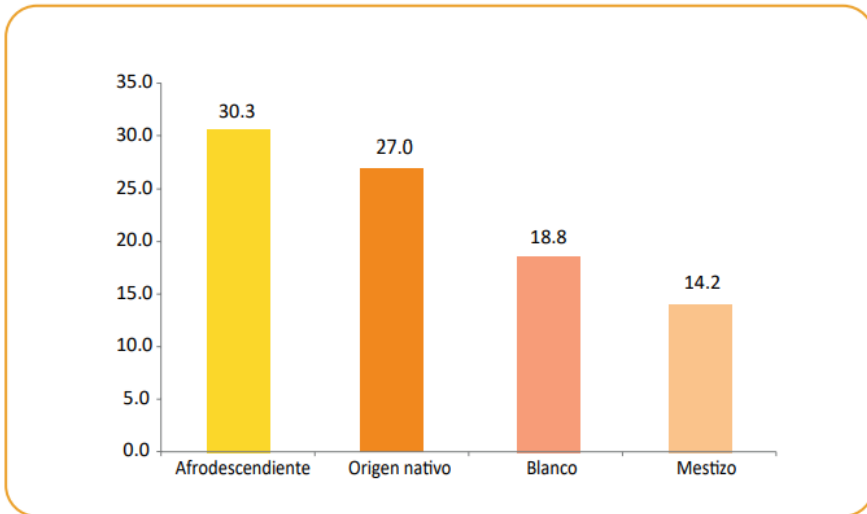
4.3.1. Diagrama de barras verticales

En este caso en el eje horizontal se ubican las categorías y se levantan rectángulos de igual ancho en la base y de altura proporcional a la frecuencia absoluta, relativa o porcentual.

También se le denomina gráfico de columnas. El programa Excel lo incluye dentro de la opción insertar como gráfico de columnas.

Ejemplo:

PERÚ: INCIDENCIA DE LA POBREZA SEGÚN AUTOPERCEPCIÓN ÉTNICA, 2019
(Porcentaje)



Fuente: Instituto Nacional de Estadística e Informática - Encuesta Nacional de Hogares, 2019.

Figura 4.8 Diagrama de barras o columnas verticales.

A continuación, se muestra cómo se obtiene un diagrama de barras a partir de una tabla de distribución de frecuencias referente a la variable tipo de transporte utilizado para trasladarse a la universidad.

Tabla 4.1 Tipo de transporte utilizado para el traslado a la universidad.

Tipo de transporte	Frecuencia
Bus	30
Taxi	40
Motocicleta	10
Bicicleta	15
Otros	25
Total	120

Las categorías de la variable tipo de transporte se ubican en el eje horizontal y las frecuencias en el eje vertical.

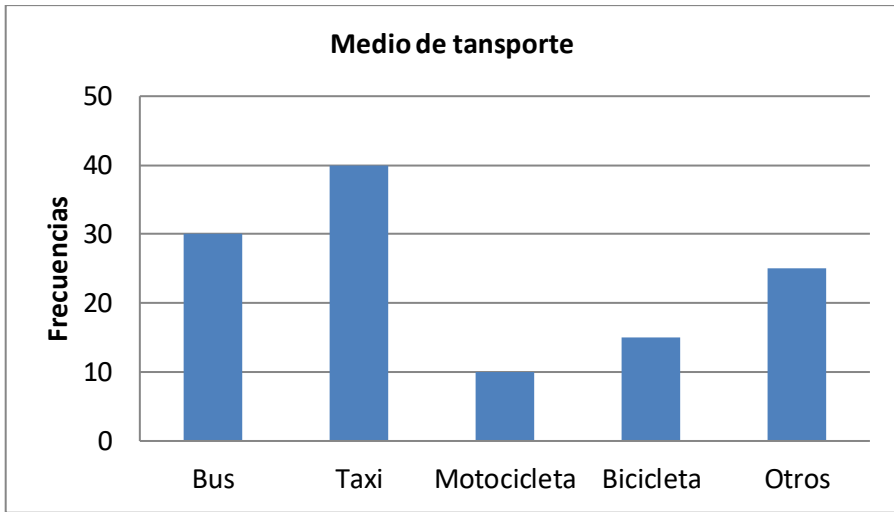


Figura 4.9 En este diagrama las alturas están asociadas a la frecuencia absoluta.

En el caso de las variables cualitativas nominales, se sugiere que la información se presente en forma ordenada de acuerdo a las frecuencias. Es decir, las barras se ubican de menor a mayor frecuencia o de forma inversa. Esta disposición de las barras hace más comprensible la gráfica y centra la atención en los valores máximos y mínimos (Naciones Unidas, 2009).

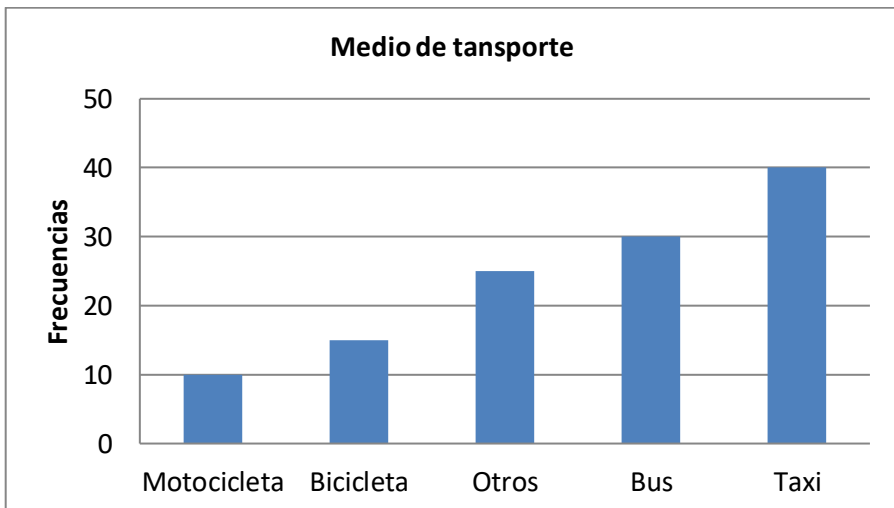


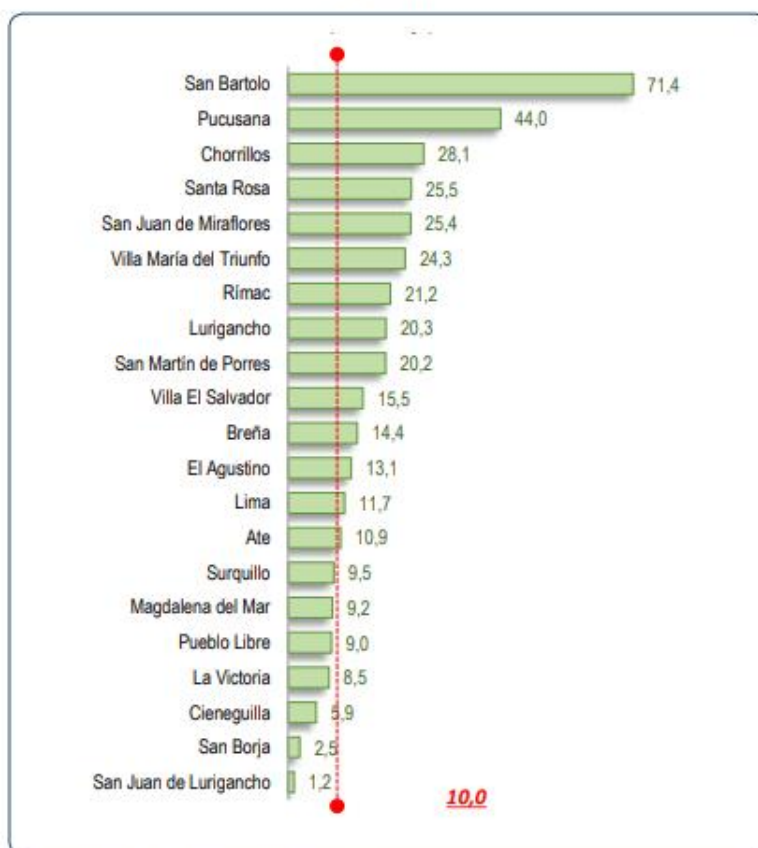
Figura 4.10 Las barras están ordenadas en forma creciente.

En los gráficos de diagramas de barras, el uso de diferentes colores solo tiene una finalidad decorativa. Por ello, se sugiere utilizar el mismo color para las barras.

4.3.2. Diagrama de barras horizontales

En este diagrama se ubican las categorías en el eje vertical y se grafican rectángulos de igual altura y de ancho proporcional a la frecuencia absoluta o relativa. Para variables nominales, se sugiere ordenar las barras en forma decreciente. Este gráfico es utilizado cuando el texto de cada categoría es muy extenso.

Lima Metropolitana: Matrimonios celebrados en forma masiva, según distrito, 2018
(Porcentaje)



Nota: Considera a los distritos con información disponible.

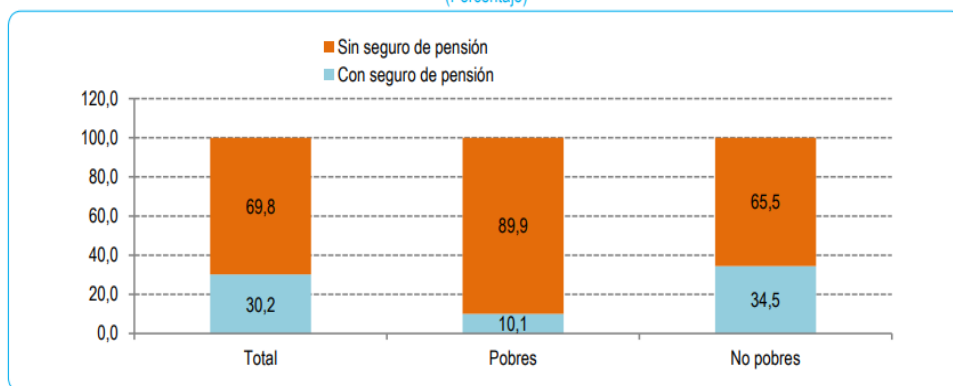
Fuente: Instituto Nacional de Estadística e Informática - Informe estadístico de matrimonio.

Figura 4.11 Observe que el nombre de varias categorías es muy extenso.

4.3.3. Gráfico de barras apiladas

En este tipo de gráfico se comparan dos variables al mismo tiempo utilizando barras superpuestas, una a continuación de la otra. Este tipo de gráfico puede ser elaborado con barras verticales u horizontales y en base a frecuencias absolutas o porcentuales (INEI, 2009).

POBLACIÓN DE 14 Y MÁS AÑOS DE EDAD POR CONDICIÓN DE TENENCIA DE SEGURO DE PENSIÓN, SEGÚN POBREZA, 2019
(Porcentaje)



Fuente: Instituto Nacional de Estadística e Informática - Encuesta Nacional de Hogares, 2019.

Figura 4.12 En este tipo de diagramas se pueden hacer múltiples comparaciones. Observe que cada columna representa el 100% de las categorías pobres, no pobres y el total.

Ejemplo: La siguiente tabla contiene información porcentual de dos variables el sexo y la interpretación del índice de masa corporal de los estudiantes de un centro educativo.

Tabla 4.2 Frecuencias según Sexo e IMC.

	Varones	Mujeres
Saludable	30	40
Sobrepeso	48	24
Obesidad	22	36
Total	100	100

Podemos elaborar un diagrama de barras apiladas colocando la interpretación del índice de masa corporal (saludable, sobrepeso y obesidad) en eje horizontal y las barras superpuestas representan la frecuencia porcentual o absoluta según el sexo.

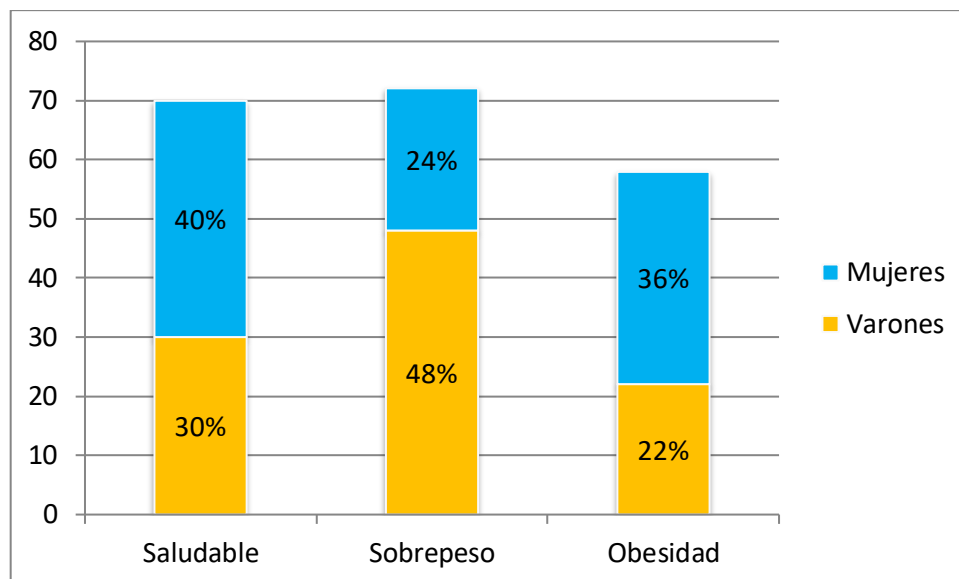


Figura 4.13 En este gráfico cada columna no representa el 100% de cada categoría. Las barras del mismo color en conjunto representan el 100%.

El diagrama de barras apiladas reemplaza a los diagramas de barras que se elaboran para hacer una comparación de dos variables en un mismo gráfico (figura 4.14).

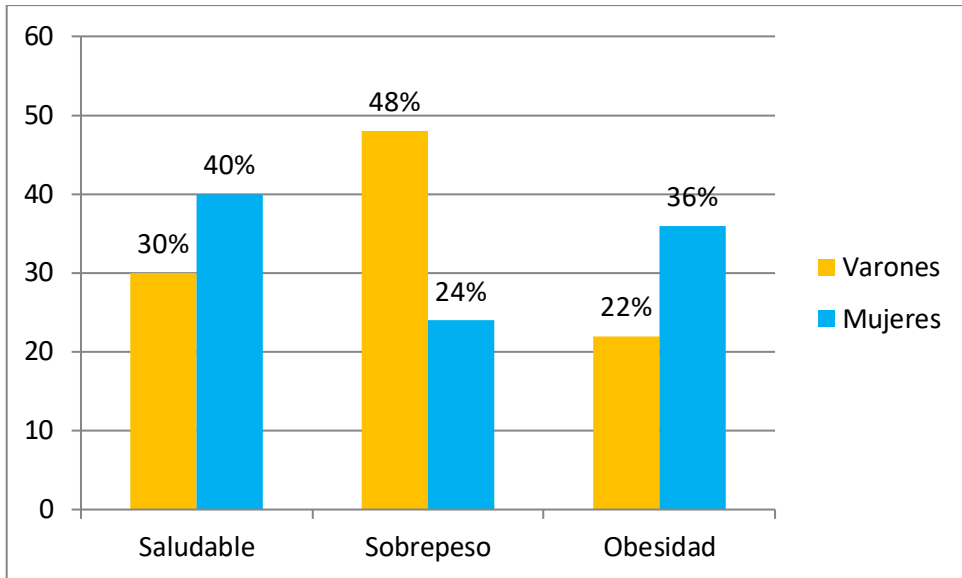


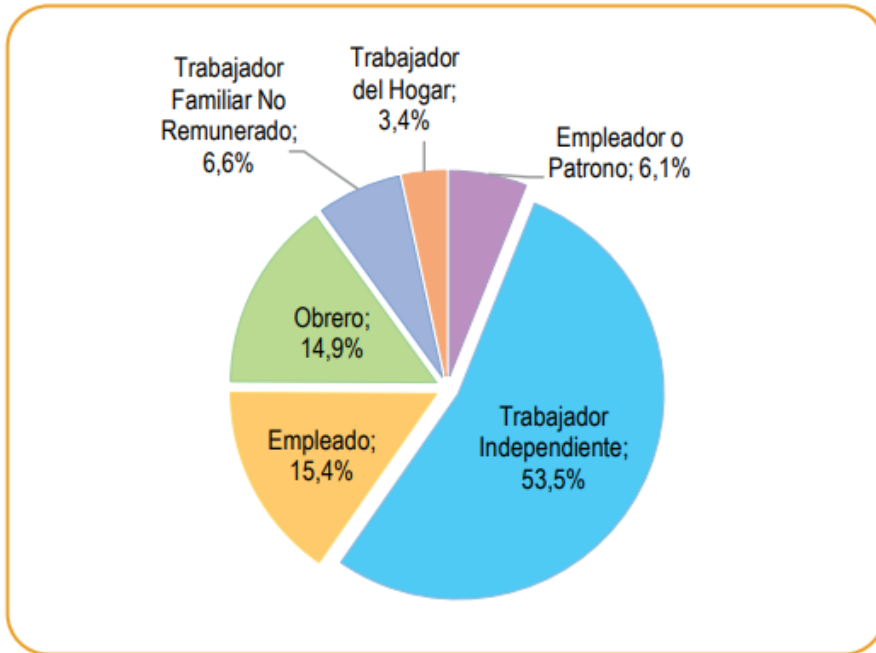
Figura 4.14 Este gráfico presenta la misma información que la figura 4.12.

El diagrama de barras apiladas es más útil para hacer la comparación entre variables. Por ejemplo, en el primer gráfico se puede ver a simple vista que el sobrepeso es la característica más frecuente en los estudiantes. En el segundo gráfico esto no se puede apreciar directamente.

4.4. Diagrama circular

Es la representación gráfica que relaciona la frecuencia relativa porcentual con un conjunto de sectores circulares. Cada sector circular es proporcional a su frecuencia porcentual, y el círculo completo representa el total de elementos (100%). Este gráfico se utiliza en el análisis de variables cualitativas.

PERÚ: DISTRIBUCIÓN DE LA PEA OCUPADA DE 60 AÑOS Y MÁS DE EDAD POR CATEGORÍA DE OCUPACIÓN, 2019
 (Porcentaje respecto al total de PEA ocupada de 60 años y más)



Fuente: Instituto Nacional de Estadística e Informática - Encuesta Nacional de Hogares, 2019.

Figura 4.15 Diagrama circular o diagrama de sectores circulares.

A continuación, se muestra cómo se obtiene un diagrama circular a partir de una tabla de distribución de frecuencias donde la variable es la preferencia en marcas de autos.

Tabla 4.3 Preferencia en marcas de auto.

Marca de auto	Frecuencia	Frecuencia porcentual
Toyota	90	18
Nissan	180	36
Kia	100	20
Mazda	45	9
Otros	85	17
Total	500	100

Observe que el tamaño de cada sector circular es proporcional al tanto por ciento que representa la categoría.

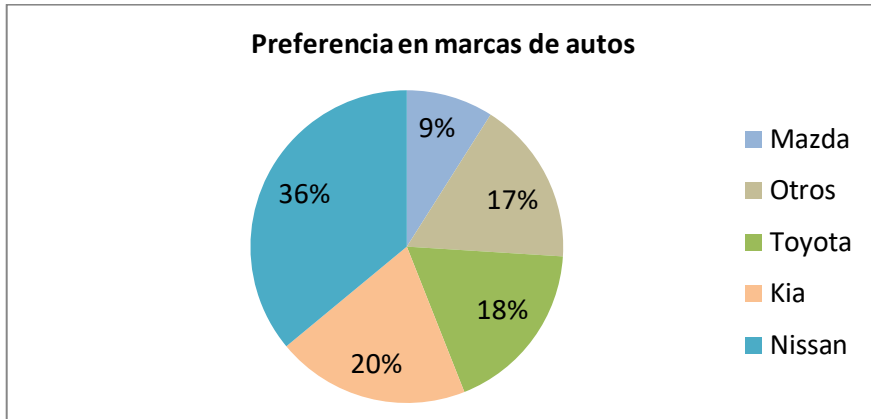


Figura 4.16 El tamaño de cada sector se asocia con la frecuencia porcentual.

En las etiquetas también se pueden colocar las categorías, de esta manera se puede prescindir de la leyenda y permite una lectura más rápida del diagrama.

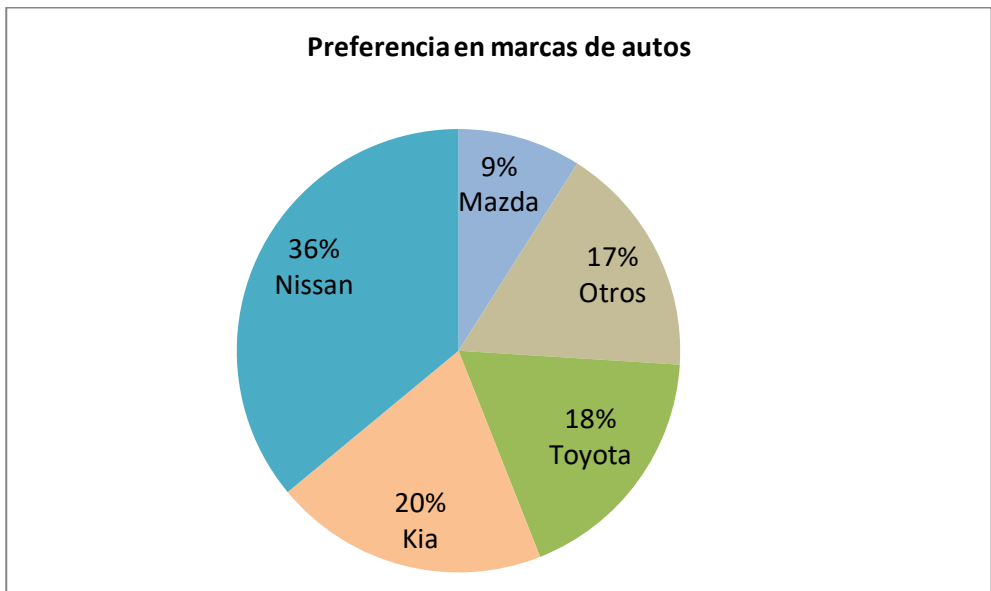


Figura 4.17 Note que toda la información está dentro del diagrama circular.

Al igual que en el diagrama de barras, se sugiere que los sectores circulares que representan a las frecuencias porcentuales estén ordenados en forma ascendente figura (4.17). Además, se sugiere utilizar el formato bidimensional.

Este tipo de gráfico se utiliza cuando la cantidad de categorías es pequeña como en el ejemplo anterior. Se sugiere que el número de categorías para el diagrama circular no sea superior a seis (Naciones Unidas, 2009).

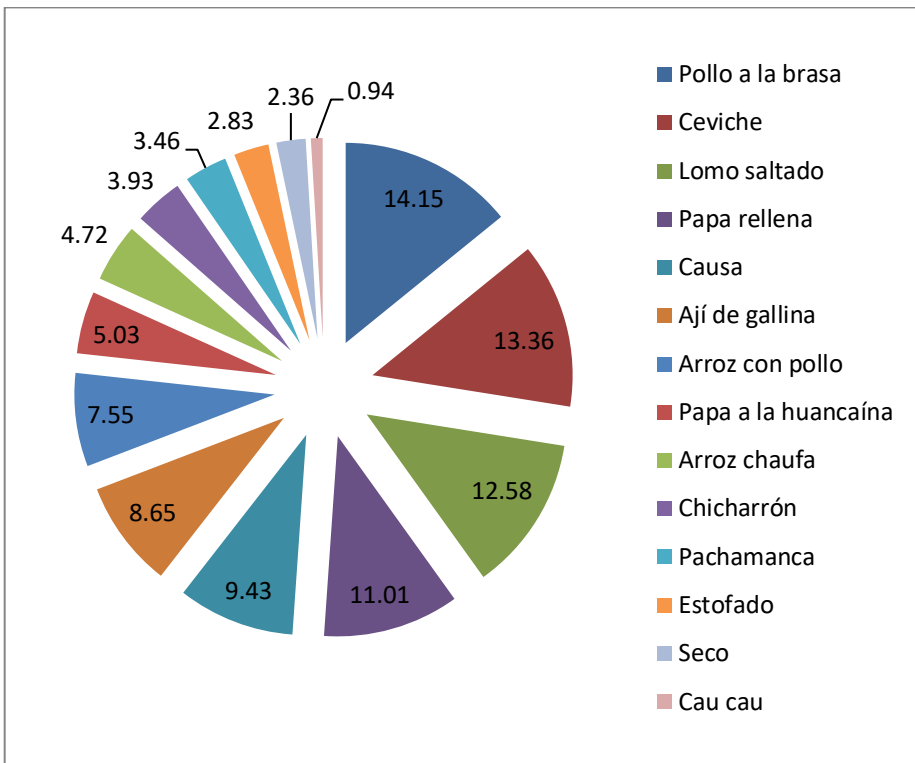


Figura 4.18 El uso del diagrama circular no es adecuado cuando existen muchas categorías.

En la figura 4.18 se puede ver que el diagrama circular contiene demasiadas categorías. La posición de los sectores de menor a mayor medida ayuda a entender el gráfico, pero aun así hay que hacer un esfuerzo para interpretar la información. En este caso se sugiere utilizar un diagrama de barras horizontales.

4.5. Histograma

Diagrama de barras cuyas bases son los intervalos de clase y las alturas son las frecuencias absolutas o relativas porcentuales.

Ejemplo: Se registra el tiempo que utilizan los empleados de una empresa en realizar llamadas telefónicas y se elabora la siguiente tabla:

Tabla 4.4 Duración de las llamadas telefónicas por empleado.

Tiempo	f_i
[0; 10>	2
[10;20>	6
[20;30>	12
[30;40>	10
[40; 50>	6
[50; 60]	4

Para construir el histograma se ubican los intervalos de clase en el eje horizontal y la frecuencia absoluta en el eje vertical.

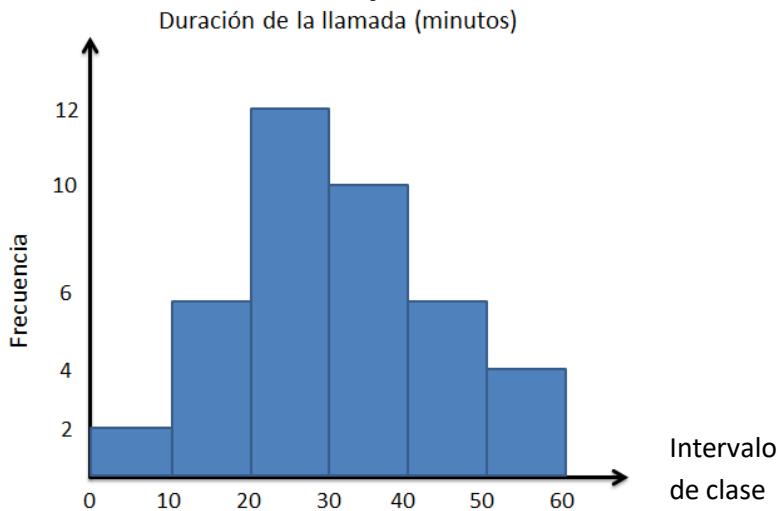


Figura 4.19 Note que la barras están unidas de forma consecutiva.

Si se elabora el histograma utilizando la frecuencia relativa porcentual, el procedimiento es el siguiente:

Tabla 4.5 Duración de las llamadas telefónicas por empleado.

Tiempo	f_i	$h_i\%$
[0; 10>	2	5
[10; 20>	6	15
[20; 30>	12	30
[20; 30>	10	25
[20; 30>	6	15
[20; 30>	4	10

Para construir el histograma se ubican los intervalos de clase en el eje horizontal y la frecuencia porcentual en el eje vertical.

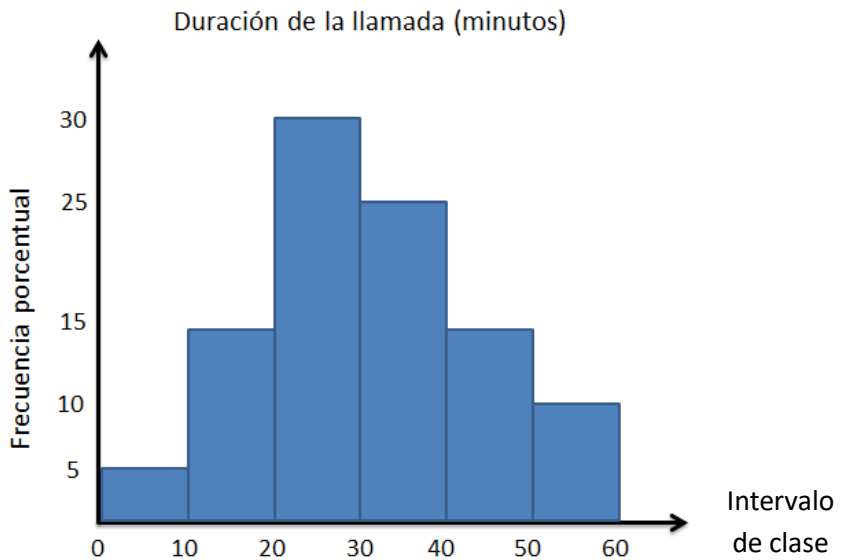


Figura 4.20 En este caso la longitud de las barras está asociada a la frecuencia porcentual.

4.6. Polígono de frecuencias

Para elaborar estos polígonos se unen los puntos medios de las bases superiores del histograma.

Ejemplo: Se mide la masa corporal de un grupo de niños y se organiza la información en la siguiente tabla:

Tabla 4.6 Masa corporal del grupo de niños.

Masa corporal	f_i	$h_i\%$
[12; 14>	10	5
[14; 16>	30	15
[16; 18>	80	40
[18; 20>	40	20
[20; 22>	10	5
[22; 24]	30	15
Total	200	100

Construimos el histograma en base a la frecuencia porcentual.

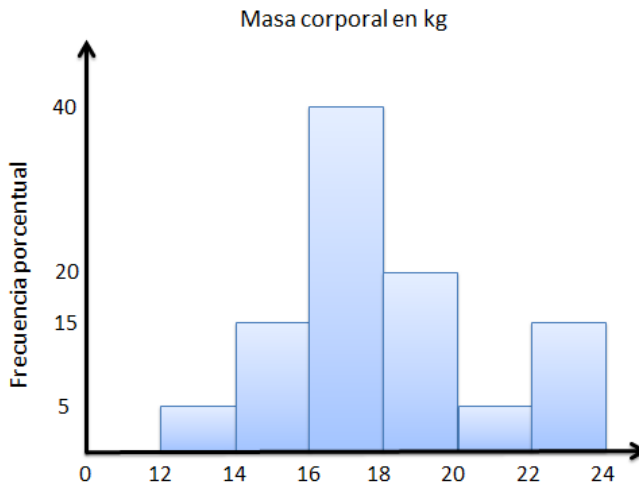


Figura 4.21 El histograma se puede construir usando la frecuencia absoluta o la frecuencia porcentual.

Ubicamos el punto medio de la base superior de cada rectángulo y unimos estos puntos con segmentos de recta, la figura resultante es el polígono de frecuencia.

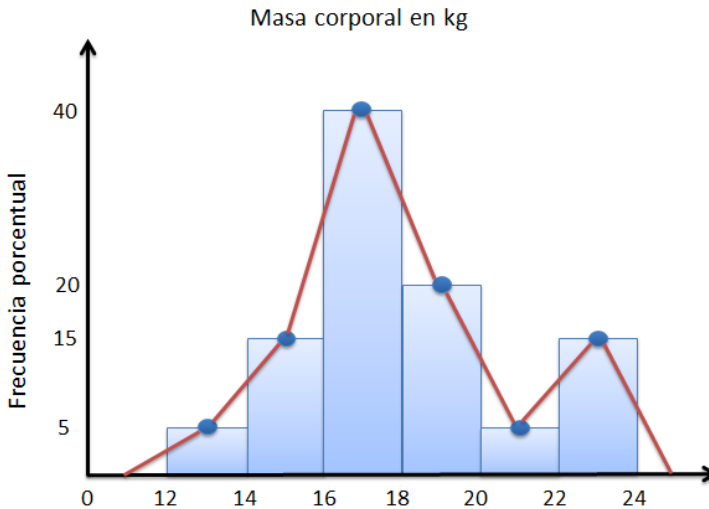


Figura 4.22 El polígono de frecuencias se compone de la unión de los segmentos.

La marca de clase (x_i) representa el punto medio de cada intervalo de clase, por lo tanto, el gráfico se puede simplificar quitando las barras del histograma y colocando en el eje horizontal las respectivas marcas de clase.

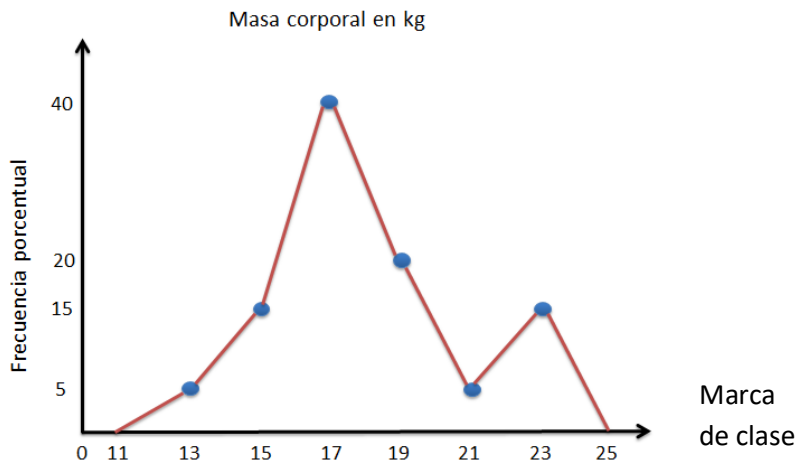


Figura 4.23 Se puede comprobar que el área del polígono formado es igual a la suma de las áreas de los rectángulos del histograma que dio origen al polígono de frecuencias.

Note que se han agregado dos marcas de clase, una antes de la primera marca de clase y otra al luego de la última marca de clase. Esto se realiza con la finalidad de cerrar el polígono.

4.7. Diagrama escalonado

Diagrama de barras cuyas bases son los intervalos de clase y las alturas son las frecuencias absolutas acumuladas o relativas porcentuales acumuladas.

Ejemplo: Utilizando la tabla del ejemplo anterior agregamos la columna de la frecuencia porcentual acumulada.

Tabla 4.7 Masa corporal del grupo de niños.

Masa corporal	f_i	$h_i\%$	$H_i\%$
[12; 14>	10	5	5
[14; 16>	30	15	20
[16; 18>	80	40	60
[18; 20>	40	20	80
[20; 22>	10	5	85
[22; 24]	30	15	100
Total	200	100	

La construcción de este diagrama es similar a la construcción del histograma.

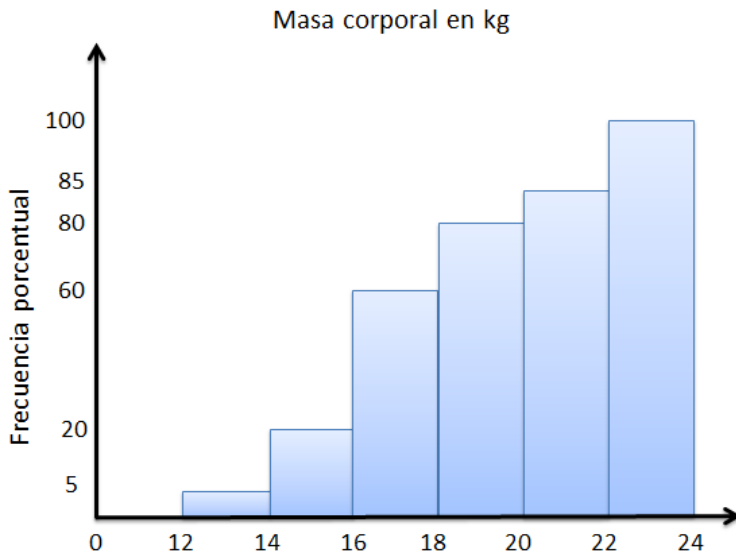


Figura 4.24 La altura de las barras se relaciona con la frecuencia porcentual acumulada.

4.8. Ojiva

Es un polígono de frecuencia que resulta de la unión de los puntos que representan a los límites superiores de cada intervalo de clase del diagrama escalonado.

Ejemplo: Usando el diagrama escalonado del ejemplo anterior unimos los límites superiores de cada intervalo con segmentos de recta.

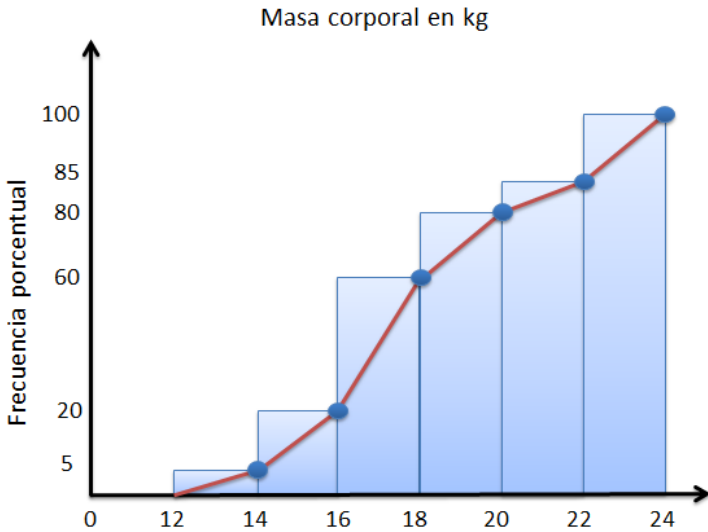


Figura 4.25 Los límites superiores de cada intervalo se unen con segmentos.

El gráfico se puede simplificar quitando las barras del diagrama escalonado.

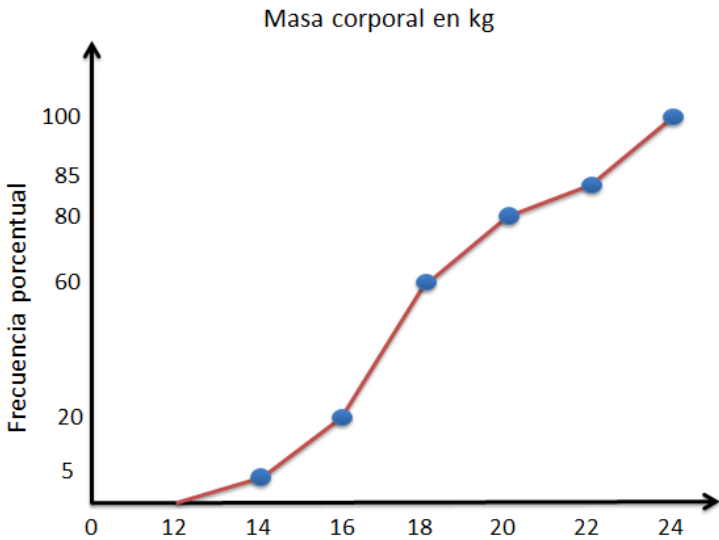


Figura 4.26 La ojiva está compuesta por la unión de los segmentos.

4.9. Grafica de una distribución bidimensional

Anteriormente se ha observado que se pueden medir diferentes variables en una sola muestra. En este caso solo consideraremos el caso donde se analizan dos variables de una sola muestra. La distribución que se obtiene al organizar los datos de estas dos variables es una distribución bidimensional.

Observe la información de la siguiente tabla:

Tabla 4.8 Datos sobre el consumo de azúcar y la actividad física.

Persona	Consumo de azúcar	Actividad física
1	Bajo	Moderada
2	Bajo	Poca
3	Alto	Moderada
4	Alto	Poca
5	Bajo	Mucha
6	Moderado	Mucha
7	Moderado	Poca
8	Moderado	Poca
9	Alto	Poca
10	Alto	Poca
11	Alto	Poca
12	Alto	Moderada
13	Alto	Poca
14	Moderado	Poca
15	Alto	Poca
16	Bajo	Mucha
17	Bajo	Mucha
18	Bajo	Moderada

Los datos recolectados se pueden organizar en una tabla de doble entrada.

Tabla 4.9 Consumo de azúcar y actividad física.

		Consumo de azúcar		
		Bajo	Moderado	Alto
Actividad física	Poca	1	3	6
	Moderada	2	0	2
	Mucha	3	1	0

Finalmente se elabora un gráfico tridimensional donde las alturas de cada barra representan las frecuencias correspondientes a cada combinación entre las categorías de las variables.

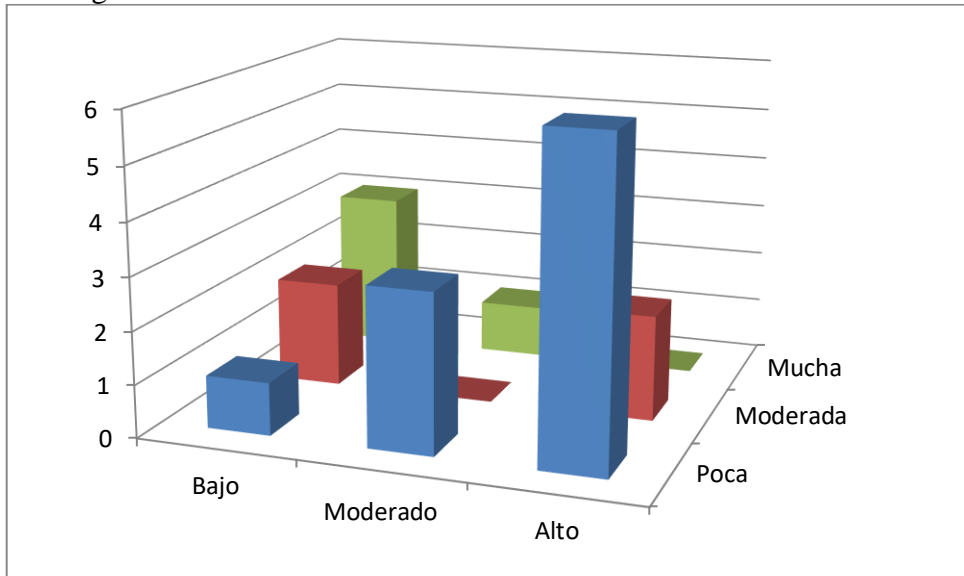


Figura 4.27 Sobre el plano donde se ubican las categorías se eleva el eje perpendicular de las frecuencias.

Interpretaciones

En la figura 4.27 se observa que la mayor frecuencia corresponde a las personas con alto consumo de azúcar y que realizan poca actividad física. No se registra personas con alto consumo de azúcar y gran actividad física. Tampoco se registran personas con un consumo de azúcar y actividad física moderada. De otro lado, se observa que 3 personas consumen poca azúcar y realizan mucha actividad física.

4.10. Diagrama de dispersión

Es la representación gráfica usada para describir el comportamiento conjunto de los datos de dos variables, donde cada caso aparece representado como un punto (x, y) en el plano definido por los valores correspondientes a las variables x e y .

X: Representa las observaciones de una variable.

Y: Representa las observaciones de la otra variable.

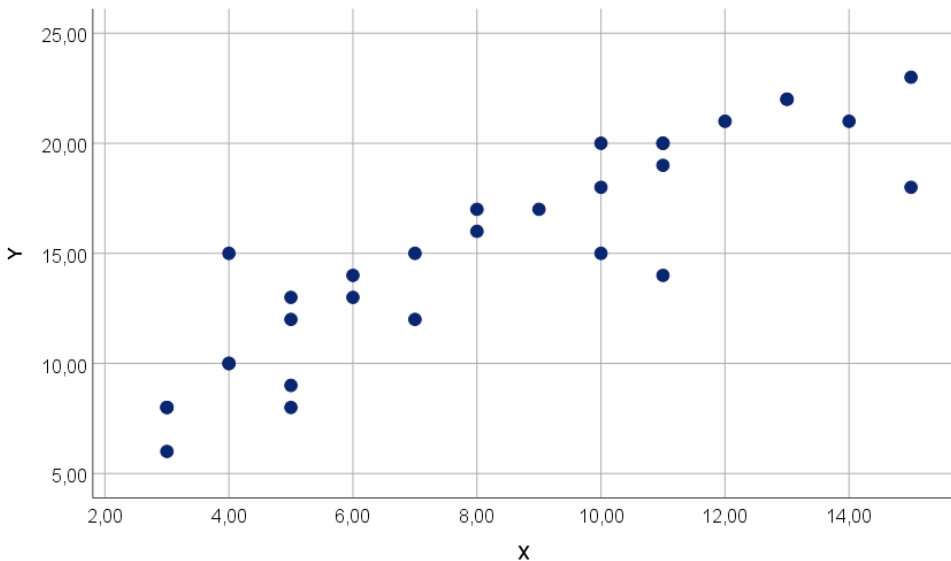


Figura 4.28 El conjunto de puntos representa un diagrama de dispersión.

Aquí se analiza la relación entre dos variables, pero los datos se obtienen de una sola muestra. Por ejemplo, el punto (4, 15) significa que un elemento de la muestra registra 4 unidades de medición en la variable X y 15 unidades de medición en la variable Y.

El diagrama de dispersión se utiliza como procedimiento exploratorio para determinar la correlación entre dos variables. La correlación es el grado de relación entre dos variables. Con este concepto se trata de establecer una ecuación que describa la relación entre las variables (Spiegel y Stephens, 2009, p. 345). Si todos los valores de las variables satisfacen dicha ecuación se dice que las variables están en perfecta correlación.

Existen diferentes tipos de correlación entre variables, sin embargo, para los fines de este libro solo consideraremos una clasificación general: correlación lineal, correlación no lineal y sin correlación.

La correlación lineal es un caso particular en el que los valores de las variables (x; y) satisfacen la ecuación: $y = a + bx$ con cierta aproximación.

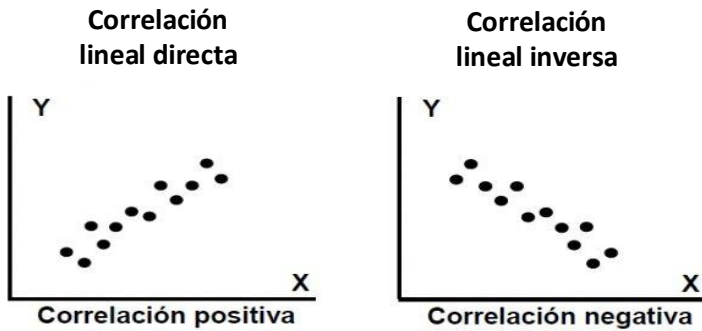


Figura 4.29 Note que en la correlación directa los puntos se agrupan en torno a una línea recta de pendiente positiva. En la correlación los puntos se agrupan en torno a una línea recta de pendiente negativa.

Ejemplo: La siguiente tabla muestra información acerca del consumo diario promedio de calorías y la masa corporal (en kg) de un grupo de estudiantes.

Tabla 4.10 Datos sobre el consumo de calorías y la masa corporal.

Estudiante	Calorías	Masa
1	2000	67
2	2120	68
3	2140	69
4	2140	70
5	2190	70
6	2190	70
7	2200	70
8	2200	70
9	2200	72
10	2200	72
11	2300	72
12	2300	73
13	2300	73
14	2350	75
15	2500	80

Usando el programa SPSS podemos obtener el diagrama de dispersión que nos indica si la tendencia entre las variables es lineal.

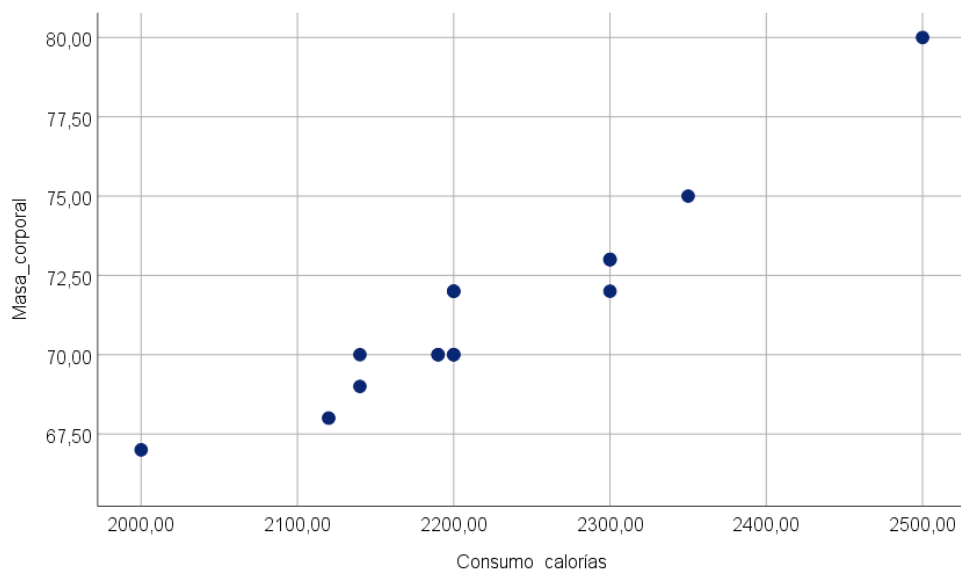


Figura 4.30 Diagrama de dispersión para las variables consumo de calorías y masa corporal.

En el diagrama de dispersión de la figura 4.30 los puntos se ubican conformando una línea recta aproximadamente, por ello se puede afirmar que existe una correlación lineal directa entre las variables masa corporal y consumo de calorías.

Ejemplo: La siguiente tabla muestra información acerca de las horas diarias de uso de redes sociales y el promedio de calificación de los estudiantes de un instituto.

Tabla 4.11 Datos sobre el uso de las redes sociales y la calificación promedio.

Estudiante	Horas	Calificación
1	2	16
2	2	18
3	3	16
4	3	17
5	4	18
6	4	14
7	4	15
8	4	16
9	5	13

10	5	15
11	5	13
12	6	17
13	6	15
14	7	13
15	8	12
16	9	14
17	9	12
18	10	11

Usando el programa SPSS se obtiene el diagrama de dispersión que nos indica si la tendencia entre las variables es lineal.

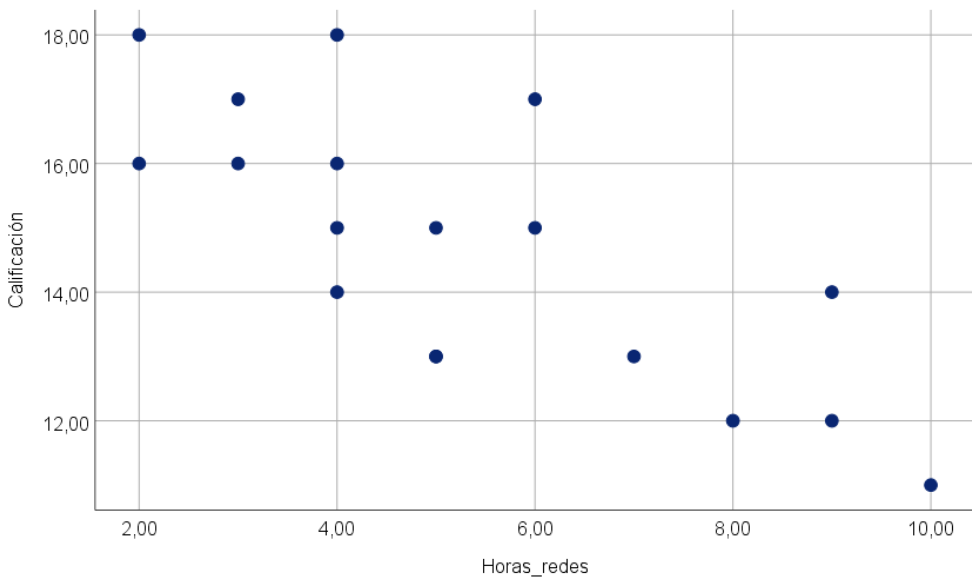


Figura 4.31 Diagrama de dispersión para las variables horas de uso de las redes sociales y promedio de calificación.

En la figura 4.31 se observa el diagrama de dispersión donde los puntos no se ubican conformando precisamente una línea recta, sin embargo, se puede identificar una correlación lineal inversa entre las variables horas de uso de redes sociales y calificación.

4.11. Gráfico de líneas

Este tipo de gráficos se utiliza para variables cuantitativas, generalmente cuando los datos se agrupan en base a un intervalo de tiempo que puede estar dado en días, meses o años. Este tipo de organización de los datos se denomina serie de tiempo. Se sugiere presentar los datos de una serie de tiempo en una gráfica de líneas con el tiempo como eje horizontal para tratar de distinguir un patrón o tendencia (Medenhall et al., 2010).

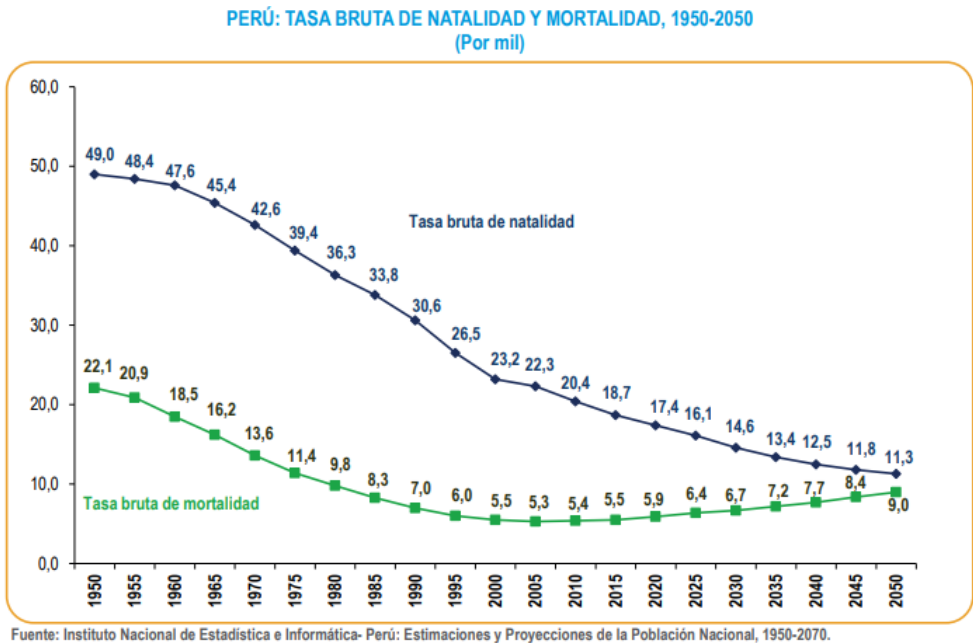
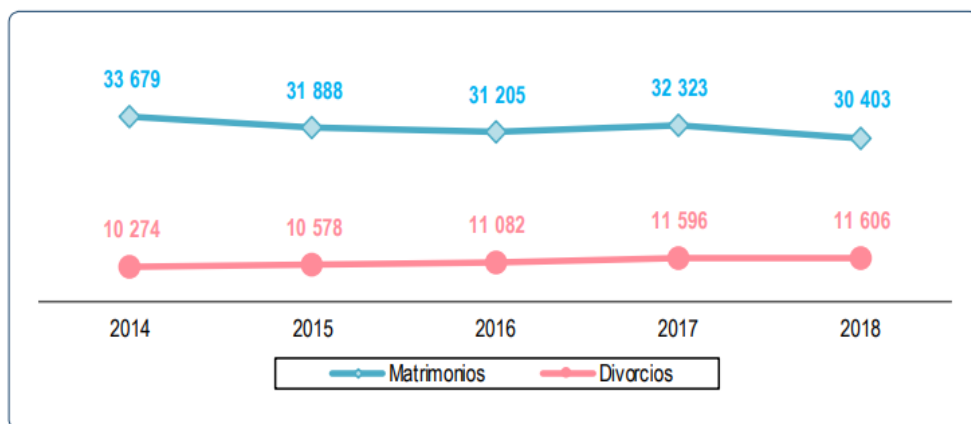


Figura 4.32 Por medio de un solo gráfico se pueden observar las tendencias de dos variables a través del tiempo.

En la figura 4.32 se muestra una serie de tiempo para dos variables la tasa bruta de natalidad y la tasa bruta de mortalidad por años. Note que en el caso de la tasa de natalidad se observa una tendencia hacia los valores bajos, es decir que la tasa se reduce con el transcurrir de los años. En el caso de la tasa de mortalidad la tendencia no es definitiva, hay una disminución en la tasa de mortalidad hasta los años 2000 y 2005, pero en los años posteriores se observa un ligero crecimiento.

Como se puede ver, este tipo de gráfico permite identificar tendencias y es utilizado, también para hacer predicciones y estimaciones. Estos gráficos son utilizados en diferentes campos del conocimiento, aunque es más frecuente encontrarlos en los análisis económicos.

Lima Metropolitana: Matrimonios celebrados y divorcios inscritos, 2014 - 2018



Fuente: Registro Nacional de Identificación y Estado Civil.
Elaboración: Instituto Nacional de Estadística e Informática.

Figura 4.33 En este gráfico se puede observar que la cantidad de divorcios presenta menor variabilidad que la cantidad de matrimonios.

El gráfico de líneas de la figura 4.33 se elaboró a partir de una tabla que organiza la información correspondiente a dos variables.

Tabla 4.12 Matrimonios y divorcios en el periodo 2014-2018.

Año	Matrimonios	Divorcios
2014	33 679	10 274
2015	31 888	10 578
2016	31 205	11 082
2017	32 323	11 596
2018	30 403	11 606

Como se puede observar, en el gráfico es más sencillo apreciar la ligera tendencia creciente para los divorcios y la tendencia irregular para los matrimonios.

4.12. Criterios para la interpretación de tablas y gráficos estadísticos

En la actualidad existe mucha información que se transmite por medio de gráficos y tablas estadísticas. Existe la suposición de que una persona con instrucción básica debe estar en condiciones de poder entender y leer críticamente la información estadística que encuentra en la prensa, Internet, medios de comunicación, y los trabajos de investigación (Batanero et al., 2010), sin embargo, esto no se verifica comúnmente en nuestra realidad. De ahí que, los diferentes países incluyen dentro de sus programas educativos el desarrollo de competencias básicas para la cultura estadística y esta se incorpora de forma general en todos los niveles educativos (Batanero, 2002). Por ello, los investigadores deben estar capacitados para presentar sus resultados en forma clara y precisa, de modo que su aporte en investigación sea significativo.

Las Naciones Unidas por medio de la Comisión Económica para Europa presenta una guía para la presentación de datos estadísticos (Naciones Unidas, 2009). En esta guía se elaboran sugerencias para hacer más entendible la presentación de la información estadística abordando temas como la estructura, la redacción, el contexto y la complejidad del mensaje que se brinda por medio de un cuadro o un gráfico estadístico.

Para el tema de la estructura y coherencia, se sugiere identificar adecuadamente la información específica y la información más general. No se recomienda utilizar demasiados números, solo los necesarios. Los datos que son secundarios se pueden colocar en la parte de anexos y hacer la referencia respectiva.

En el caso de la redacción la guía para la presentación de datos estadísticos sugiere tomar en cuenta los siguientes aspectos.

- Utilizar frases cortas.
- Tratar una idea en cada frase.
- Dividir las frases largas.
- Iniciar cada párrafo con lo principal del mensaje.
- Mantener una escritura concisa.

Sobre la complejidad de la interpretación se sugiere usar un lenguaje sencillo y claro. Esta última característica no le resta rigurosidad académica a la presentación cuando se usa una buena redacción.

También se hace énfasis en el uso adecuado de la tecnología. Sabemos que existen muchos programas estadísticos que nos permiten procesar los datos y obtener los resultados de manera muy rápida. Sin embargo, estos programas son solo herramientas, el trabajo del investigador es dar la interpretación correcta los a resultados obtenidos y llegar a conclusiones válidas.



Figura 4.34 Estos son algunos de los programas estadísticos más utilizados. R es un software estadístico libre.

Se recomienda no explotar los recursos que nos brindan los programas estadísticos de forma indiscriminada (figura 4.34). Se debe evitar la elaboración de gráficos o tablas sobrecargados de datos. El uso innecesario de tablas o gráficos con abundante información no ayuda como sustento en la explicación de los resultados y conclusiones de una investigación, por el contrario, satura la vista y puede llegar a confundir a las personas que acceden a nuestra investigación.

Grupo	Frecuencia	Porcentaje
Joven	90	45.0
Adulto	70	35.0
Anciano	40	20.0
	200	100.0

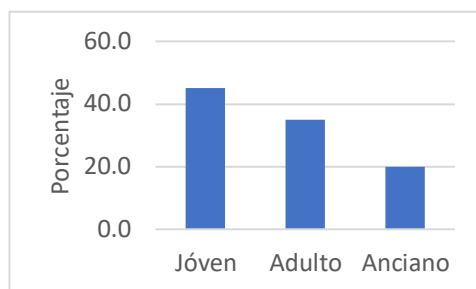


Figura 4.35 Observe que la tabla y el gráfico contienen la misma información. El investigador debe elegir cuál de los dos elementos es el más adecuado para la presentación de sus resultados.

Cabe resaltar que en los trabajos de investigación los resultados estadísticos, generalmente, se presentan de forma textual. Esto se debe a que el uso de las tablas y gráficos sirven como soporte para la interpretación de los resultados y no son la finalidad de la investigación. Es decir, es el investigador quien debe interpretar sus resultados de acuerdo con el contexto y los objetivos de su investigación. Las tablas y gráficos son de utilidad cuando la persona que accede a la investigación desea entender mejor los resultados y verificar coherencia de las conclusiones.

De lo anterior se puede concluir que las tablas y gráficos deben contener información resumida que pueda ser comprensible para las personas que accedan a los resultados de la investigación. Sin embargo, existen tablas que tienen diferente cantidad de información y algunas requieren que los lectores tengan una preparación previa en la lectura de información estadística.

La medida de cuanto esfuerzo tiene que hacer el lector para entender lo que se está tratando de comunicar por medio de una tabla o gráfico se denomina carga cognitiva (Naciones unidas 2009). Esto implica que existen tablas y gráficos con diferente carga cognitiva. Las tablas de baja carga cognitiva son fáciles de entender y recordar mientras que los de elevada carga cognitiva requieren de procesos mentales más complejos para su interpretación y requieren más esfuerzo para recordar la información.

Tabla 4.13 Tabla cruzada sexo y estado nutricional.

		Estado nutricional				Total	
		Bajo peso	Normal	Sobrepeso	Obesidad		
Sexo	Varón	% dentro de Sexo	34,8%	37,0%	15,2%	13,0%	100,0%
		% dentro de Estado nutricional	64,0%	58,6%	46,7%	54,5%	57,5%
		% del total	20,0%	21,3%	8,8%	7,5%	57,5%
Mujer		% dentro de Sexo	26,5%	35,3%	23,5%	14,7%	100,0%
		% dentro de Estado nutricional	36,0%	41,4%	53,3%	45,5%	42,5%
		% del total	11,3%	15,0%	10,0%	6,3%	42,5%
Total		% dentro de Sexo	31,3%	36,3%	18,8%	13,8%	100,0%
		% dentro de Estado nutricional	100,0%	100,0%	100,0%	100,0%	100,0%
		% del total	31,3%	36,3%	18,8%	13,8%	100,0%

La tabla 4.13 presenta una gran carga cognitiva. Es decir, para poder dar lectura a la tabla se tiene que hacer un mayor esfuerzo. Para entender cada número hay que orientarse primero por las categorías, las variables y el total al que hace referencia cada tanto por ciento.

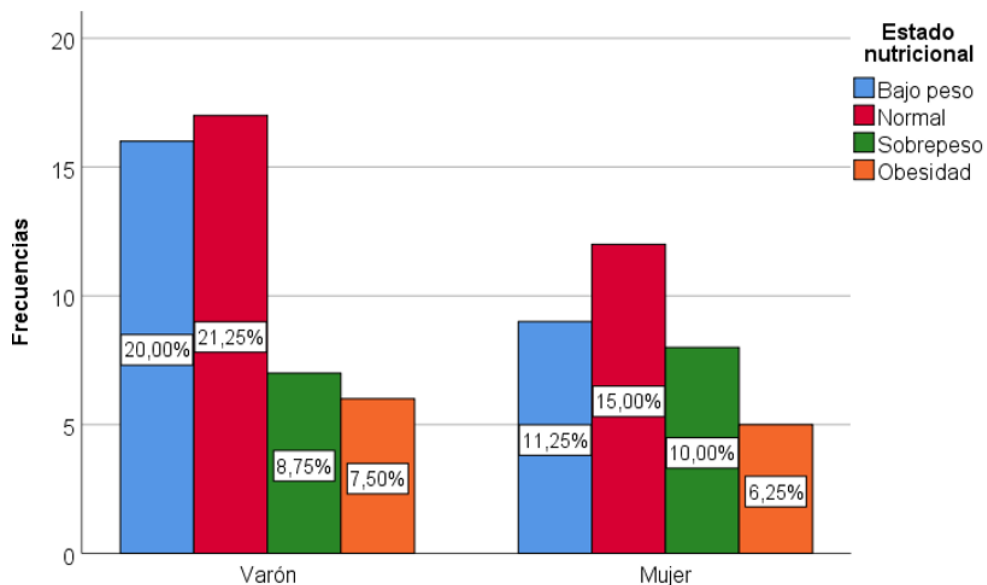


Figura 4.36 El uso de un gráfico ayuda a comprender mejor los resultados de la tabla 4.13.

La figura 4.36 se basa en los datos de la tabla 4.13 y como se puede apreciar la lectura es más sencilla, es decir, requiere una menor carga cognitiva. Por ejemplo, se puede interpretar a simple vista que hay una cantidad menor de personas con obesidad en comparación con las personas que tienen un estado nutricional normal. Por ello, al momento de elaborar la presentación de resultados se debe hacer una elección adecuada entre el uso de tablas o gráficos.

Tanto en las tablas como en los gráficos, el uso adecuado de líneas o sombreado permite a los usuarios orientarse sobre cómo empezar la lectura de una tabla o gráfico estadístico (Naciones Unidas, 2009). Esto es importante al momento de sustentar un trabajo de investigación, puesto que permite que el investigador transmita el análisis de sus resultados de forma más eficiente.

Capítulo 5: Medidas de posición central

Contenido

- ❖ Estadígrafos
- ❖ Media
- ❖ Cálculo de la media para datos no agrupados, datos agrupados y datos agrupados en intervalos. Otros promedios.
- ❖ Mediana
- ❖ Cálculo de la mediana para datos no agrupados, datos agrupados y datos agrupados en intervalos.
- ❖ Moda
- ❖ Cálculo de la moda para datos no agrupados, datos agrupados y datos agrupados en intervalos.

5.1. Estadígrafos

El análisis estadístico se basa generalmente en muestras, no se analiza a individuos en particular ni a toda la población. Debido a esto, se cuenta con un número relativamente grande de datos y es necesario resumir la información por medio de métodos adecuados para presentar los resultados de manera que se pueda facilitar su interpretación (Bologna, 2011).

Lind et al. (2016) indican que cualquier medición basada en una muestra de datos recibe el nombre de estadístico o estadígrafo. En este capítulo se estudia a aquellos estadígrafos que tiene la particularidad de ubicarse en la parte central de una distribución de datos. Estas medidas de posición central que resumen algunas características de una distribución son la media, la mediana y la moda.

5.2. Media

Es el promedio aritmético de los datos. Por lo tanto, el cálculo e interpretación de la media sólo tiene sentido cuando utilizamos datos numéricos como es el caso de las variables discretas, continuas, de escala de medición de intervalo o de razón.

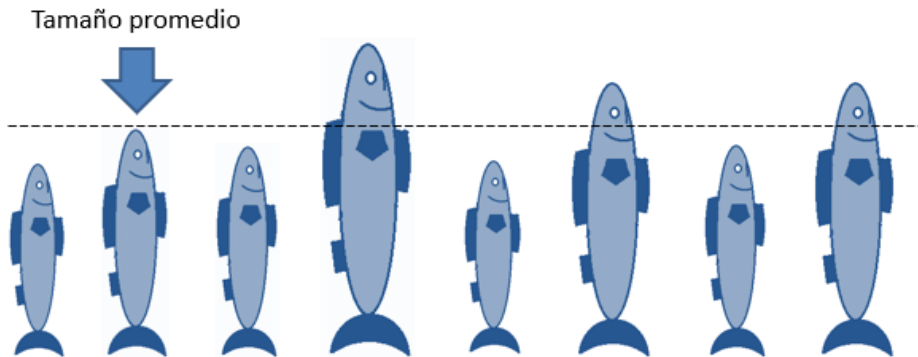


Figura 5.1 Observe que existen peces más pequeños y más grandes que el pez de tamaño promedio.

Ejemplo: Las calificaciones más altas al final de un curso fueron 19.2, 19 y 18.8. Entonces, el promedio de estas calificaciones será:

$$\text{Promedio} = \frac{19.2 + 19 + 18.8}{3} = 19$$

Este promedio de las tres calificaciones representa el conjunto de datos y tiene coherencia puesto que es un valor cercano a cada uno de los datos. Dentro del trabajo estadístico se busca resumir la información, por ello, cuando se tiene un conjunto de datos cuantitativos y se quiere expresar una característica de estos es muy útil utilizar la media. Por ejemplo, para un administrador es útil conocer la ganancia diaria promedio, a un ingeniero industrial le interesa conocer la producción semanal promedio, a un economista le interesa conocer el consumo mensual promedio de las familias en distintas regiones, etc.

Cuando se analiza variables cualitativas no es viable utilizar la media puesto que no se pueden promediar atributos o categorías. Por ejemplo, si analizamos la variable marca de auto favorita y tenemos las marcas Nissan, Toyota y Kia ¿Cómo se podría calcular el promedio? Para las variables cualitativas existen otros estadígrafos más adecuados.

5.2.1. Cálculo de la media para datos sin agrupar

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

\bar{X} : Es la media de los datos

n : Es la cantidad de datos

x_i : Es el valor de cada dato

Ejemplo: Luego de registrar la masa corporal de un grupo de estudiantes se obtuvo los siguientes datos:

Estudiante	1	2	3	4	5	6	7	8
Masa (Kg)	50	55	60	62	60	58	60	55

Para obtener la media debemos sumar todos los datos y dividir el resultado entre la cantidad de datos.

$$\bar{X} = \frac{50 + 55 + 60 + 62 + 60 + 58 + 60 + 55}{8} = 57.5$$

Interpretación: La masa corporal promedio del grupo de estudiantes es de 57.5 kg.

5.2.2. Cálculo de la media para datos agrupados

En la mayoría de investigaciones se trabaja con muestras más grandes que la muestra del ejemplo anterior y se dificulta realizar el mismo procedimiento. Por ello, para muestras más grandes se utiliza la tabla de distribución de frecuencias y para calcular la media se utiliza la siguiente expresión:

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

Ejemplo: Se registra y organiza la masa corporal de un grupo de 50 empleados en la siguiente tabla:

Tabla 5.1 Masa corporal por empleado.

Masa corporal (Kg)	Cantidad de estudiantes
50	6
55	10
60	15
65	13
68	4
70	2

El x_i está representado por los valores de la masa corporal y la frecuencia f_i corresponde a la cantidad de estudiantes que tienen una masa x_i . Entonces, colocamos los símbolos respectivos en la tabla y agregamos una columna para el producto $x_i f_i$.

Tabla 5.2 Elementos para el cálculo de la media.

x_i	f_i	$x_i f_i$
50	6	300
55	10	550
60	15	900
65	13	845
68	4	272
70	2	140

$$\bar{X} = \frac{300 + 550 + 900 + 845 + 272 + 140}{50} = 60.14$$

Interpretación: La masa corporal promedio del grupo de empleados es de 60.14 kg.

5.2.3. Cálculo de la media para datos agrupados en intervalos

El procedimiento para el cálculo de la media es igual al proceso anterior, pero antes se debe calcular la marca de clase x_i .

Ejemplo: En el mes de abril se realizó una encuesta a un grupo de padres de familia para conocer el gasto que realizaron en la compra de útiles escolares. La información recogida se organizó en la siguiente tabla:

Tabla 5.3 Gasto en útiles escolares por padre de familia.

Gasto (S/)	f_i
[60;70>	23
[70; 80>	35
[80; 90>	65
[90; 100>	20
[100; 110]	17
Total	n=160

Debemos calcular la marca de clase para cada intervalo y multiplicar este valor por su respectiva frecuencia.

Tabla 5.4 Proceso para el cálculo de la media.

Gasto (S/)	x_i	f_i	$x_i f_i$
[60; 70>	$(60+70)/2=65$	23	$65 \times 23=1495$
[70; 80>	$(70+80)/2=75$	35	$75 \times 35=2625$
[80; 90>	$(80+90)/2=85$	65	$85 \times 65=5525$
[90; 100>	$(90+100)/2=95$	20	$95 \times 20=1900$
[100; 110]	$(100+110)/2=105$	17	$105 \times 17=1785$
Total		n=160	

Luego se suman los productos calculados en la última columna.

Tabla 5.5 Elementos necesarios para el cálculo de la media.

Gasto (S/)	x_i	f_i	$x_i f_i$
[60; 70>	65	23	1495
[70; 80>	75	35	2625
[80; 90>	85	65	5525
[90; 100>	95	20	1900
[100; 110]	105	17	1785
Total		n=160	$\sum x_i f_i=13330$

Reemplazamos los valores en la fórmula:

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

$$\bar{X} = \frac{13330}{160} = 88.31$$

Interpretación: El gasto promedio por la compra de útiles escolares es de S/ 88.31.

5.2.4. Otros promedios

Media geométrica ($\overline{X_G}$): Si se tienen n datos, la media geométrica se define como la raíz enésima del producto de todos los datos.

$$\overline{X_G} = \sqrt[n]{x_1 x_2 x_3 \dots x_n}$$

Media armónica ($\overline{X_H}$): Si se tienen n datos, la media armónica se define como en cociente entre la cantidad de datos y la suma de los inversos de todos los datos.

$$\overline{X_H} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \dots + \frac{1}{x_n}}$$

Media cuadrática ($\overline{X_Q}$): Si se tienen n datos, la media cuadrática se define como la raíz cuadrada del promedio de todos los cuadrados de los datos.

$$\overline{X_Q} = \sqrt{\frac{x_1^2 + x_2^2 + x_3^2 + \dots + x_n^2}{n}}$$

Ejemplo: Si se toma una muestra de 6 personas y se observa la estatura (en metros) de cada una de ellas se obtienen los siguientes datos.

1.70 1.80 1.76 1.75 1.70 1.74

A continuación, se calcula los cuatro tipos de promedios para establecer una relación entre ellos.

Calculamos la media aritmética

$$\bar{X} = \frac{1.70 + 1.80 + 1.76 + 1.75 + 1.70 + 1.74}{6} = 1.7417$$

Calculamos la media geométrica

$$\bar{X}_G = \sqrt[6]{1.70 \times 1.80 \times 1.76 \times 1.75 \times 1.70 \times 1.74} = 1.7413$$

Calculamos la media armónica

$$\bar{X}_H = \frac{6}{\frac{1}{1.70} + \frac{1}{1.80} + \frac{1}{1.76} + \frac{1}{1.75} + \frac{1}{1.70} + \frac{1}{1.74}} = 1.7410$$

Calculamos la media cuadrática

$$\bar{X}_Q = \sqrt{\frac{1.70^2 + 1.80^2 + 1.76^2 + 1.75^2 + 1.70^2 + 1.74^2}{6}} = 1.7420$$

En los resultados se han utilizado 4 decimales con la finalidad de hacer una comparación más precisa.

$$1.7410 < 1.7413 < 1.7417 < 1.7420$$

En general, existe una relación de orden entre los promedios siempre y cuando se calculen sobre la misma muestra. Se cumple:

$$\bar{X}_H \leq \bar{X}_G \leq \bar{X} \leq \bar{X}_Q$$

Sin embargo, el promedio más utilizado es el promedio aritmético y con base en este promedio se desarrollan otros procesos de la estadística descriptiva e inferencial. El promedio cuadrático se verá más adelante en el tema de las medidas de dispersión.

5.3. Mediana

Es el valor que se encuentra en la mitad de un conjunto de datos ordenados en forma ascendente. Este estadígrafo sirve para analizar variables cuantitativas y cualitativas ordinales.

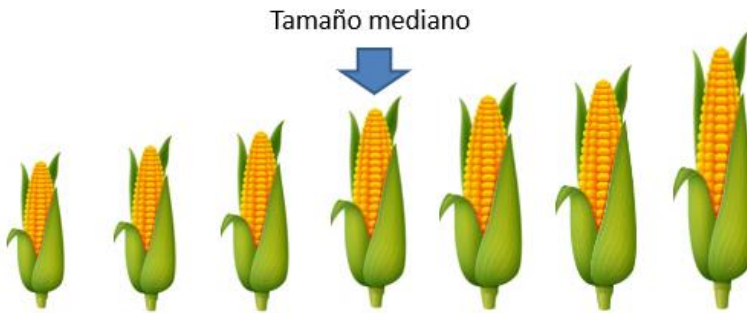


Figura 5.2 Observe que a la izquierda del maíz de tamaño mediano se encuentran los de menor tamaño y a la derecha los de mayor tamaño

Al ubicarse en la parte central, la mediana, divide al conjunto de datos en dos mitades (figura 5.2 y 5.3). La primera mitad contiene a todos los datos cuyos valores son menores o iguales a dicha mediana y la segunda mitad contiene a todos los datos cuyos valores son mayores o iguales a esta.

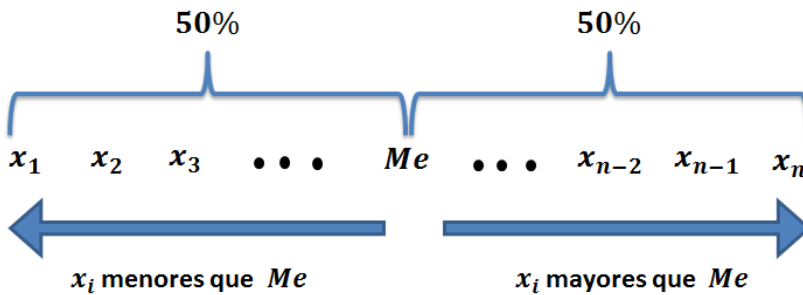


Figura 5.3 50% de los datos son menores o iguales la mediana y el otro 50% de datos son mayores o iguales a esta.

5.3.1. Cálculo de la mediana para datos sin agrupar

Ejemplo: A continuación, se muestra las edades de un grupo de personas.

43 33 50 48 48 40 35 48 52 30 35

Para calcular la mediana de este grupo de datos primero hay que ordenarlos en forma ascendente:

30 33 35 35 40 43 48 48 48 50 52

La mediana es el dato que se encuentra en la parte central, en este caso: $Me=43$

Interpretación: El 50% de las personas tiene como mínimo 43 años. También se puede decir que la mitad de las personas tiene como máximo 43 años.

Cuando la cantidad de datos (n) es impar la posición de la mediana se puede ubicar utilizando la siguiente expresión:

$$Posición = \frac{n + 1}{2}$$

En el ejemplo anterior se puede verificar que $n=11$. Luego, utilizamos la expresión que permite ubicar la posición de la mediana:

$$Posición = \frac{11 + 1}{2} = 6$$

Ubicamos la sexta posición en la fila de datos ordenados y ahí se encuentra la mediana.

Sexta posición

30 33 35 35 40 43 48 48 48 50 52

Pero ¿Qué sucede cuando la cantidad de datos (n) es par? En este caso existen dos datos en la parte central. Por ello se debe efectuar un promedio de estos dos valores.

Ejemplo: A continuación, se muestra el gasto mensual (en S/) en consumo de galletas de un grupo de estudiantes.

20 15 15 28 18 30 32 12 17 12

Ordenamos los datos en forma ascendente:

12 12 15 15 **17 18** 20 28 30 32

Tenemos dos datos en la parte central, por lo tanto:

$$Me = \frac{17 + 18}{2} = 17.5$$

Interpretación: El 50% de las personas gasta más de S/ 17.5 en su consumo mensual de galletas. También se puede decir que la mitad de los estudiantes gasta menos de S/ 17.5 en su consumo mensual de galletas.

5.3.2. Cálculo de la mediana para datos agrupados

El procedimiento anterior nos permite el cálculo de la mediana, pero para el caso en que la muestra es pequeña. Cuando la cantidad de datos es muy grande se utiliza tablas de distribución de frecuencias y el procedimiento es similar al que se ha desarrollado. Primero se ubica la frecuencia mediana (F_m) utilizando:

$$F_m \geq \frac{n}{2}$$

El valor correspondiente a esta frecuencia es la mediana.

Ejemplo: La siguiente tabla agrupa las edades de los alumnos del último ciclo de la universidad.

Tabla 5.6 Edades por estudiante.

Edades	Frecuencia
23	8
24	12
25	35
26	48
27	15

Utilizamos la simbología respectiva y agregamos la frecuencia absoluta acumulada (F).

Tabla 5.7 Edades por estudiante.

Edades	f_i	F_i
23	8	8
24	12	20
25	45	65
26	58	123
27	25	148

El valor de n se obtiene sumando todas las frecuencias absolutas.

$$n = 8 + 12 + 45 + 58 + 25 = 148$$

Luego calculamos la frecuencia mediana.

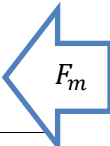
$$F_m \geq \frac{148}{2}$$

$$F_m \geq 74$$

Ahora buscamos en la tabla la primera frecuencia absoluta acumulada cuyo valor es mayor o igual a 74.

Tabla 5.8 Ubicación de la frecuencia mediana.

Edades	f_i	F_i
23	8	8
24	12	20
25	45	65
26	58	123
27	25	148



La frecuencia mediana es 123, porque es el primer valor mayor que 74. Por lo tanto, la mediana es la edad correspondiente a dicha frecuencia.

$$Me = 26$$

Interpretación: El 50% de los estudiantes del último ciclo tiene por lo menos 26 años. También se puede decir que la mitad de los estudiantes del último ciclo tiene como máximo 26 años.

5.3.3. Cálculo de la mediana para datos agrupados en intervalos

Como en el caso anterior, primero se ubica la frecuencia mediana (F_m) utilizando:

$$F_m \geq \frac{n}{2}$$

Luego se utiliza la siguiente expresión:

$$M_e = L_i + w \times \frac{\frac{n}{2} - F_{m-1}}{f_m}$$

Dónde:

M_e : Mediana

w : Ancho de clase o amplitud.

L_i : Límite inferior de la clase

n : cantidad de datos

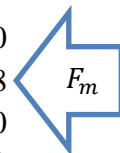
F_{m-1} : Frecuencia acumulada anterior a la F_m

f_m : Frecuencia absoluta mediana.

Ejemplo: En la tabla 5.9 se muestra el ahorro semanal de un conjunto de personas agrupado en intervalos. Con estos datos se calcula e interpreta la mediana.

Tabla 5.9 Ahorro semanal por persona.

Ahorro semanal (S/)	fi	Fi
[50; 60>	4	4
[60; 70>	6	10
[70; 80>	18	28
[80; 90>	12	40
[90; 100]	15	50
Total	n=55	



$$F_m \geq \frac{55}{2}$$

$$F_m \geq 27.5$$

La primera frecuencia absoluta acumulada que es mayor que 27.5 es 28, entonces, $F_m=28$.

El límite inferior de la clase mediana [70; 80> es $L_i = 70$, la amplitud $w=10$, $n/2=27.5$, la frecuencia anterior a la frecuencia mediana $F_{m-1} = 10$ y la frecuencia absoluta mediana $f_m = 18$.

Reemplazamos estos valores en la fórmula:

$$M_e = L_i + w \times \frac{\frac{n}{2} - F_{m-1}}{f_m}$$

$$M_e = 70 + 10 \left(\frac{27.5 - 10}{18} \right)$$

$$M_e = 79.72$$

Interpretación: La mitad de los encuestados tiene un ahorro semanal mínimo de S/ 79.72.

5.4. Moda

Es el dato de mayor frecuencia. La moda es el estadígrafo que se puede utilizar tanto en variables cuantitativas como cualitativas.



Figura 5.4 La camiseta de color rojo es la que se observa en un mayor número de ocasiones.

Por ejemplo, si registramos las edades de un grupo de estudiantes y obtenemos los siguientes datos:

23 22 21 **20** **20** 23 **20** **20** **20** 24

Observamos que el dato que se presenta con mayor frecuencia es el 20, por lo tanto, podemos decir que la moda para este conjunto de edades es de 20 años.

5.4.1. Cálculo de la moda para datos no agrupados

Ejemplo: Se ha registrado la cantidad de platos a la carta que ofrecen los restaurantes de un distrito de Lima. Los resultados son los siguientes:

10 **12** 9 10 15 **12** **12** 16 13 **12** **12** **12** 10 **12**

Se puede ver que el número que aparece con más frecuencia es el 12.

Interpretación: La cantidad de platos a la carta ofrecidos con mayor frecuencia es igual a 12.

Sin embargo, la moda no es necesariamente única. En ocasiones pueden existir dos o más modas.

Ejemplo: En el siguiente conjunto de datos el 7 y el 11 aparecen 5 veces cada uno. Por lo tanto, existen dos modas $M_o=7$ y $M_o=11$. Se puede decir que la distribución de datos es bimodal.

11 4 11 11 7 10 7 6 7 7 7 11 11 8

Cuando la distribución de datos posee tres modas se le conoce como trimodal y si tiene más de tres modas se le denomina multimodal.

5.4.2. Cálculo de la moda para datos agrupados

Cuando la cantidad de datos es muy grande se utiliza tablas de distribución de frecuencias y el procedimiento es el mismo que se ha utilizado. Es decir, debemos identificar el dato con la mayor frecuencia.

Ejemplo: En la siguiente tabla se muestran las edades de un grupo de estudiantes.

Tabla 5.10 Edades de un grupo de estudiantes.

Edad	f_i
18	4
19	6
20	12
21	18
22	10
23	10
Total	60

Se puede ver que la mayor frecuencia es 18 y a esta frecuencia le corresponde el valor de 21 años de edad.

Interpretación: La edad más frecuente entre los estudiantes es de 21 años.

5.4.3. Cálculo de la moda para datos agrupados en intervalos

La clase modal se encuentra en la mayor frecuencia

$$M_o = L_i + w \times \frac{d_1}{d_1 + d_2}$$

Dónde:

M_o : Moda.

w : Ancho de clase o amplitud.

L_i : Límite inferior de la clase.

f_m : Frecuencia modal o mayor frecuencia absoluta.

$d_1 : f_m - f_{m-1}$

$d_2 : f_m - f_{m+1}$

Ejemplo: Usando la información de la tabla 5.11 realizaremos el cálculo e interpretación de la moda.

Tabla 5.11 Ahorro semanal por persona.

Ahorro semanal (S/)	fi	
[50; 60>	4	
[60; 70>	6	
[70; 80>	18	f_m
[80; 90>	12	
[90; 100]	15	
Total	n=55	

La frecuencia modal (f_m) es 18, puesto que es la mayor frecuencia absoluta.

A esta frecuencia modal le restamos la frecuencia anterior:

$$d_1 = 18 - 6 = 12$$

Luego, a la misma frecuencia modal le restamos la frecuencia posterior:

$$d_2 = 18 - 12 = 6$$

El límite inferior de la clase modal $[70; 80>$ es $L_i = 70$ y la amplitud $w=10$.

Reemplazamos los valores en la fórmula:

$$M_o = 70 + 10\left(\frac{12}{12+6}\right)$$

$$M_o = 76.67$$

Interpretación: El ahorro semanal más frecuente entre los encuestados es de S/ 76.67.

Capítulo 6: Medidas de posición no central

Contenido

- ❖ Cuartiles
- ❖ Cálculo de cuartiles para datos no agrupados, para datos agrupados y para datos agrupados en intervalos.
- ❖ Deciles
- ❖ Percentiles

En esta sección se analizan algunos valores que se ubican en ciertas posiciones de la distribución que son de interés para el investigador. Estos estadígrafos se utilizan para tratar variables numéricas y permiten dividir a la distribución en partes iguales.

6.1. Cuartiles

Son tres valores que dividen una distribución de datos en cuatro partes iguales. Se debe tener en cuenta que los datos tienen que estar ordenados en forma ascendente para poder realizar correctamente las interpretaciones. Este estadígrafo sirve para analizar variables cuantitativas y algunas variables ordinales que presentan gran cantidad de categorías.

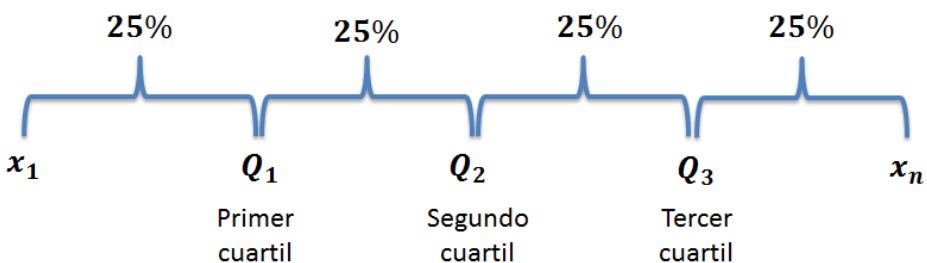


Figura 6.1 Los cuartiles dividen a la distribución de datos en cuatro partes iguales.

6.1.1. Cálculo de cuartiles para datos sin agrupar

Ejemplo: A continuación, se muestra el tiempo de cocción (en minutos) de diferentes variedades de papa.

28 28 35 30 30 30 25 31 24 30 35 32 35 36 30

Para ubicar la posición de los cuartiles utilizamos:

Posición del primer cuartil: $x_{Q_1} = \frac{n+1}{4}$

Posición del segundo cuartil: $x_{Q_2} = 2 \left(\frac{n+1}{4} \right)$

Posición del tercer cuartil: $x_{Q_3} = 3 \left(\frac{n+1}{4} \right)$

Como se puede apreciar $n=15$. Por lo tanto:

$$x_{Q_1} = \frac{15 + 1}{4} = 4$$

$$x_{Q_2} = 2 \left(\frac{15 + 1}{4} \right) = 8$$

$$x_{Q_3} = 3 \left(\frac{15 + 1}{4} \right) = 12$$

Para calcular los cuartiles se debe ordenar los datos en forma ascendente y ubicar las posiciones 4, 8 y 12.

		Posición 4				Posición 8				Posición 12			
24	25	28	28	30	30	30	30	30	31	32	35	35	36
		Primer cuartil				Segundo cuartil				Tercer cuartil			

Figura 6.2 Los cuartiles dividen a la distribución de datos en cuatro partes iguales.

Note que el segundo cuartil coincide con el centro de la distribución. Es decir, el segundo cuartil es la mediana. Por ello, solo interpretaremos

el primer y tercer cuartil, el segundo cuartil se interpreta como la mediana.

interpretaciones:

Q_1 : El 25% de las variedades de papa se cuece en 28 minutos como máximo.

Q_1 : El 75% de las variedades de papa se cuece por lo menos en 28 minutos.

Q_3 : El 25% de las variedades de papa se cuece 35 minutos como mínimo.

Q_3 : El 75% de las variedades de papa se cuece en 35 minutos como máximo.

6.1.2. Cálculo de cuartiles para datos agrupados

Cuando la cantidad de datos es muy grande se utiliza tablas de distribución de frecuencias y el procedimiento es similar al que se ha desarrollado anteriormente para el cálculo de la mediana. Primero se ubica la frecuencia cuartil (F_Q) utilizando:

Frecuencia para el primer cuartil: $F_{Q_1} \geq \frac{n}{4}$

Frecuencia para el segundo cuartil: $F_{Q_2} \geq 2 \left(\frac{n}{4} \right)$

Frecuencia para el tercer cuartil: $F_{Q_3} \geq 3 \left(\frac{n}{4} \right)$

El valor de la variable correspondiente a cada frecuencia es el cuartil.

Ejemplo: Se encuestó a un grupo de amas de casa acerca de la cantidad de veces por mes que hacen compras en el supermercado. Los resultados fueron los siguientes:

Tabla 6.1 Frecuencia de compras en supermercados.

Cantidad de veces	Frecuencia
4	10
5	36
6	24
7	15
8	50
9	15
10	6
Total	n=156

Completamos a tabla con la frecuencia acumulada.

	Cantidad de veces	f_i	F_i	
	4	10	10	
Cuartiles {	Q_1	5	46	F_{Q_1}
		6	70	
	Q_2	7	85	F_{Q_2}
	8	135		F_{Q_3}
	9	15	150	
	10	6	156	
	Total	n=156		

Figura 6.3 Observe que el valor del cuartil se ubica en la columna de categorías.

Frecuencia para el primer cuartil: $F_{Q_1} \geq \frac{156}{4}$

$$F_{Q_1} \geq 39$$

La primera frecuencia acumulada que es mayor o igual a 39 es 46. Por lo tanto, $F_{Q_1} = 46$

$$Q_1 = 5$$

Frecuencia para el segundo cuartil: $F_{Q_2} \geq 2 \left(\frac{156}{4} \right)$

$$F_{Q_2} \geq 78$$

La primera frecuencia acumulada que es mayor o igual a 78 es 85. Por lo tanto, $F_{Q_2} = 85$

$$Q_2 = 7$$

Frecuencia para el tercer cuartil: $F_{Q_3} \geq 3 \left(\frac{156}{4} \right)$

$$F_{Q_3} \geq 117$$

La primera frecuencia acumulada que es mayor o igual a 117 es 135. Por lo tanto, $F_{Q_3} = 135$

$$Q_3 = 8$$

Interpretaciones

Q_1 : El 25% de las amas de casa va de compras al supermercado 5 veces por mes como máximo.

Q_1 : El 75% de las amas de casa va de compras al supermercado 5 veces por mes como mínimo.

Q_3 : El 75% de las amas de casa va de compras al supermercado, como máximo 8 veces por mes.

Q_3 : El 25% de las amas de casa va de compras al supermercado, como mínimo 8 veces por mes.

6.1.3. Cálculo de cuartiles para datos agrupados en intervalos

Como en el caso anterior, primero se ubica la frecuencia cuartil (F_Q) utilizando:

Frecuencia para el primer cuartil: $F_{Q_1} \geq \frac{n}{4}$

Frecuencia para el segundo cuartil: $F_{Q_2} \geq 2 \left(\frac{n}{4}\right)$

Frecuencia para el tercer cuartil: $F_{Q_3} \geq 3 \left(\frac{n}{4}\right)$

$$Q_i = L_i + w \times \frac{\frac{n \times i}{4} - F_{Q_{i-1}}}{f_{Q_i}}$$

Dónde:

Q_i : Cuartil

w : Ancho de clase.

L_i : Límite inferior de la clase

n : cantidad de datos

$F_{Q_{i-1}}$: Frecuencia acumulada anterior a F_{Q_i}

f_{Q_i} : Frecuencia absoluta cuartil i.

Ejemplo: Una empresa distribuidora de balones de gas realizó un estudio para conocer el consumo mensual de gas natural (en soles) por hogar en una determinada zona de Lima. Los resultados se presentan en la siguiente tabla:

Tabla 6.2 Consumo mensual de gas por hogar.

Consumo de gas	fi
[20; 25>	50
[25; 30>	62
[30; 35>	55
[35; 40>	40
[40; 45]	36
Total	n=243

Primero se completa la tabla con la frecuencia absoluta acumulada.

Tabla 6.3 Ubicación del primer y tercer cuartil.

Consumo de gas	fi	Fi	
[20; 25>	50	50	
[25; 30>	62	112	F_{Q_1}
[30; 35>	55	167	
[35; 40>	40	207	F_{Q_3}
[40; 45]	36	243	
Total	n=243		

Para calcular el primer cuartil:

$$F_{Q_1} \geq \frac{243}{4}$$

$$F_{Q_1} \geq 60.75$$

La primera frecuencia acumulada que es mayor o igual a 60.75 es 112. Por lo tanto, $F_{Q_1} = 112$

El límite inferior de la clase primer cuartil [25; 30> es $L_1 = 25$, la amplitud $w=5$, $n/4=60.75$, la frecuencia anterior a la frecuencia primer cuartil $F_{Q_1-1} = 50$ y la frecuencia absoluta correspondiente al primer cuartil $f_{Q_1} = 62$.

Reemplazamos estos valores en la fórmula:

$$Q_1 = L_1 + w \times \frac{\frac{n}{4} - F_{Q_1-1}}{f_{Q_1}}$$

$$Q_1 = 25 + 5 \left(\frac{60.75 - 50}{62} \right)$$

$$Q_1 = 25.87$$

Interpretación:

Q_1 : El 25% de los hogares observados presenta un consumo mensual máximo de S/ 25.87.

Q_1 : El 75% de los hogares observados presenta un consumo mensual mínimo de S/ 25.87.

Para calcular el tercer cuartil:

$$F_{Q_3} \geq \frac{3 \times 243}{4}$$

$$F_{Q_3} \geq 182.25$$

La primera frecuencia acumulada que es mayor o igual a 182.25 es 207. Por lo tanto, $F_{Q_3} = 207$

El límite inferior de la clase tercer cuartil $[35; 40>$ es $L_i = 35$, la amplitud $w=5$, $3n/4=182.25$, la frecuencia anterior a la frecuencia tercer cuartil $F_{Q_3-1} = 167$ y la frecuencia absoluta correspondiente al tercer cuartil $f_{Q_3} = 40$.

Reemplazamos estos valores en la fórmula:

$$Q_3 = L_3 + w \times \frac{\frac{3n}{4} - F_{Q_3-1}}{f_{Q_3}}$$

$$Q_3 = 35 + 5 \left(\frac{182.25 - 167}{40} \right)$$

$$Q_3 = 36.91$$

Interpretación:

Q_3 : El 75% de los hogares observados presenta un consumo mensual máximo de S/ 36.91.

Q_3 : El 25% de los hogares observados presenta un consumo mensual mínimo de S/ 36.91.

6.2. Deciles

Son nueve valores que dividen una distribución de datos en diez partes iguales. Por lo tanto, los deciles no se utilizan para analizar muestras pequeñas, pues no tiene mucho sentido dividir una muestra pequeña en diez partes. Este estadígrafo sirve para analizar variables cuantitativas.

Se debe considerar cada parte generada como un diez por ciento del total, esto nos sirve para hacer las interpretaciones.

6.2.1. Cálculo de los deciles

Cuando se utiliza tablas de distribución de frecuencias el procedimiento es similar al que se ha desarrollado anteriormente para el cálculo de los cuartiles. Primero se ubica la frecuencia Decil (F_D) utilizando:

$$F_{D_i} \geq \frac{in}{10}$$

Donde i es el decil que se quiere conocer.

Frecuencia para el primer decil: $F_{D_1} \geq \frac{n}{10}$

Frecuencia para el segundo decil: $F_{D_2} \geq \frac{2n}{10}$

Frecuencia para el tercer decil: $F_{D_3} \geq \frac{3n}{10}$

·
·
·

Frecuencia para el noveno decil: $F_{D_9} \geq \frac{9n}{10}$

El valor de la variable correspondiente a cada frecuencia es el decil.

Ejemplo: Se encuestó a un grupo de personas acerca de la cantidad horas por semana que dedican a realizar ejercicios. Los resultados fueron los siguientes:

Tabla 6.4 Tiempo dedicado al ejercicio

Horas	Frecuencia
2	20
4	40
6	54
8	40
10	10
12	16
14	8
Total	n=188

Completamos a tabla con la frecuencia acumulada.

Tabla 6.5 Tiempo dedicado al ejercicio

Horas	f_i	F_i
2	20	20
4	40	60
6	54	114
8	40	154
10	10	164
12	16	180
14	8	188
Total	n=188	

Para calcular el segundo decil:

$$F_{D_2} \geq \frac{2 \times 188}{10}$$

$$F_{D_2} \geq 37.6$$

La primera frecuencia acumulada que es mayor o igual a 37.6 es 60. Por lo tanto, $F_{D_2} = 60$ y la cantidad de horas que le corresponde a esta frecuencia es 4.

$$D_2 = 4$$

Para entender lo que representa el segundo decil podemos hacer una representación gráfica.

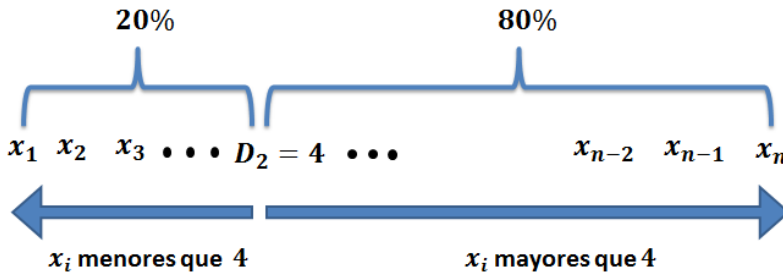


Figura 6.4 Los deciles dividen a la distribución de datos en diez partes iguales.

Interpretaciones

D_2 : El 20% de las personas encuestadas realiza 4 horas de ejercicios por semana como máximo.

D_2 : El 80% de las personas encuestadas realiza 4 horas de ejercicios por semana como mínimo.

Para calcular el séptimo decil:

$$F_{D_7} \geq \frac{7 \times 188}{10}$$

$$F_{D_7} \geq 131.6$$

La primera frecuencia acumulada que es mayor o igual a 131.6 es 154. Por lo tanto, $F_{D_7} = 154$ y la cantidad de horas que le corresponde a esta frecuencia es 8.

$$D_7 = 8$$

Para entender lo que representa el séptimo decil podemos hacer una representación gráfica.

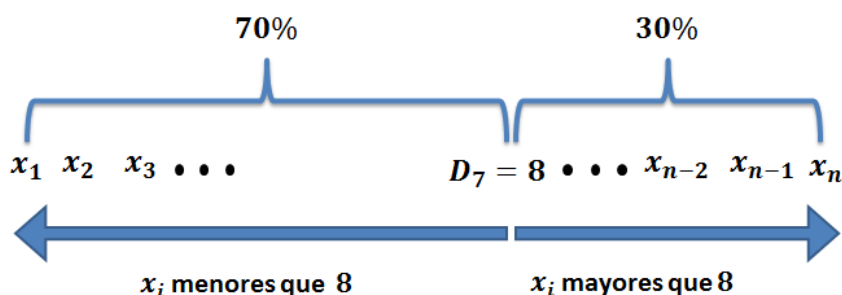


Figura 6.5 El séptimo decil indica que el 70% de los datos no supera su valor.

Interpretaciones

D_7 : El 70% de las personas encuestadas realiza 8 horas de ejercicios por semana como máximo.

D_7 : El 30% de las personas encuestadas realiza 8 horas de ejercicios por semana como mínimo.

Se debe tener en cuenta que el quinto decil es equivalente a la mediana dado que ambos se encuentran en el centro de la distribución de datos.

6.3. Percentiles

Son 99 valores que dividen una distribución de datos en 100 partes iguales. Como en el caso de los deciles, es recomendable el cálculo de percentiles solo cuando la muestra es grande. El proceso de cálculo de los percentiles es similar al cálculo de los deciles y se utiliza para analizar variables cuantitativas.

6.3.1. Cálculo de los percentiles

Primero se ubica la frecuencia percentil (F_P) utilizando:

$$F_{P_i} \geq \frac{in}{100}$$

Donde i es el percentil que se quiere conocer.

Frecuencia para el primer percentil: $F_{P_1} \geq \frac{n}{100}$

Frecuencia para el segundo percentil: $F_{P_2} \geq \frac{2n}{100}$

Frecuencia para el tercer percentil: $F_{P_3} \geq \frac{3n}{100}$

·
·
·

Frecuencia para el noveno percentil: $F_{P_{99}} \geq \frac{99n}{100}$

El valor de la variable correspondiente a cada frecuencia es el percentil buscado.

Anderson et al. (2012) señalan que el percentil p es un valor tal que por lo menos “ p ” por ciento de las observaciones es menor o igual que este valor y el tanto por ciento restantes de las observaciones es mayor o igual que este valor.

Ejemplo: Para conocer la cantidad anual de libros que leen los jóvenes de una ciudad se realizó una encuesta y los datos se organizaron en la siguiente tabla:

Tabla 6.6 Cantidad de libros leídos por estudiante.

Cantidad de libros	Frecuencia
0	8
1	50
2	74
3	60
4	46
5	30
6	16
7	10
Total	n=304

Completamos a tabla con la frecuencia acumulada.

Tabla 6.7 Cantidad de libros leídos por estudiante.

Cantidad de libros	f_i	F_i
0	8	8
1	60	68
2	74	142
3	60	202
4	46	248
5	30	278
6	16	294
7	10	304
Total	$n=304$	

Para calcular el percentil 35:

$$F_{P_{35}} \geq \frac{35 \times 304}{100}$$

$$F_{P_{35}} \geq 106.4$$

La primera frecuencia acumulada que es mayor o igual a 106.4 es 142. Por lo tanto, $F_{P_{35}} = 142$ y la cantidad de libros que le corresponde a esta frecuencia es 2.

$$P_{35} = 2$$

Interpretaciones

P_{35} : El 35% de los jóvenes encuestados lee 2 libros por año como máximo.

P_{35} : El 65% de los jóvenes encuestados lee 2 libros por año como mínimo.

Para calcular el percentil 85:

$$F_{P_{85}} \geq \frac{85 \times 304}{100}$$

$$F_{P_{85}} \geq 258.4$$

La primera frecuencia acumulada que es mayor o igual a 258.4 es 278. Por lo tanto, $F_{P_{85}} = 278$ y la cantidad de libros que le corresponde a esta frecuencia es 5.

$$P_{85} = 5$$

Interpretaciones

P_{85} : El 85% de los jóvenes encuestados lee 5 libros por año como máximo.

P_{85} : El 15% de los jóvenes encuestados lee 5 libros por año como mínimo.

Para resumir ciertos procesos de cálculo se pueden considerar algunas equivalencias entre los valores de la mediana, los cuartiles, los deciles y los percentiles. Por ejemplo:

$$P_{25} = Q_1$$

$$P_{50} = Q_2 = D_5 = Me$$

$$P_{75} = Q_3$$

$$P_{10} = D_1$$

$$P_{20} = D_2$$

$$P_{30} = D_3$$

$$P_{40} = D_4$$

$$P_{60} = D_6$$

$$P_{70} = D_7$$

$$P_{80} = D_8$$

$$P_{90} = D_9$$

Capítulo 7: Medidas de dispersión

Contenido

- ❖ Rango
- ❖ Cálculo del rango para datos sin agrupar, datos agrupados y datos agrupados en intervalos
- ❖ Rango intercuartil
- ❖ Desviación estándar
- ❖ Cálculo de la desviación estándar para datos sin agrupar, datos agrupados y datos agrupados en intervalos
- ❖ Varianza
- ❖ Coeficiente de variación

Las medidas de tendencia central nos permiten conocer aquellos valores que se ubican en el centro de una distribución, pero en muchas ocasiones esa información, no es muy completa, y en algunos casos nos pueden llevar a cometer errores en su interpretación. Por ello, se requiere de otros estadísticos que permitan conocer más detalles de una distribución de datos numéricos.

Por ejemplo, en la tabla 7.1 se observan dos grupos de masas corporales y el respectivo promedio para cada grupo.

Tabla 7.1 Masa corporal promedio.

Grupo	A	B
Masa corporal	61	80
	59	48
	59	50
	61	62
Media	60	60

Los promedios calculados para cada grupo son iguales. Sin embargo, en el grupo A los valores son muy parecidos a la media 60, en cambio en el grupo B los valores no son tan parecidos a la media 60. Por lo tanto, la media no es suficiente para conocer las características de una muestra. En el grupo A se

puede decir que la media es representativa y en el grupo B la media no es representativa, puesto que existe una mayor variabilidad.

7.1. Rango

Es la diferencia entre el mayor y el menor valor de un conjunto de datos. Entonces, se entiende que el rango solo se utiliza cuando trabajamos con variables cuantitativas.

7.1.1. Cálculo del rango para datos no agrupados

Ejemplo: Luego de un examen de conocimientos se obtienen las siguientes calificaciones:

10 12 **9** 10 15 12 18 **19** 13 16 16 18 14 12

Observe que la menor calificación es 9 y la mayor calificación es 19.

$$R = 19 - 9 = 10$$

7.1.2. Cálculo del rango para datos agrupados

Cuando se utilizan las tablas de distribución de frecuencias, los valores de las clases están ordenados en forma ascendente. Por lo tanto, solo debemos restar el último valor de la última clase con el valor de la primera clase.

Tabla 7.2 Edad por estudiante.

Edad	f_i
18	4
19	6
20	12
21	18
22	10
23	10
Total	60

$$R = 23 - 18 = 5$$

7.1.3. Cálculo del rango para datos agrupados en intervalos

En este caso los valores de los intervalos de clase están ordenados en forma ascendente. Por lo tanto, solo debemos restar el límite superior de la última clase con el límite inferior de la primera clase.

Tabla 7.3 Edad por persona.

Masa	f_i
[50; 54>	10
[54; 58>	20
[58; 62>	25
[62; 66>	15
[66; 70>	15
[70; 74]	10
Total	95

$$R = 74 - 50 = 24$$

El rango como medida de dispersión permite comparar diferentes grupos de datos referentes a una misma variable.

Ejemplo: Se analiza la cantidad de contactos que registran los celulares de los estudiantes de una universidad para conocer qué tan variable es la cantidad de contactos entre varones y mujeres. Se obtienen los siguientes resultados:

Tabla 7.4 Valores máximo y mínimos para la cantidad de contactos.

Grupo	Varones	Mujeres
Valor mínimo	60	80
Valor máximo	210	320

Para el grupo de varones:

$$R = 210 - 60 = 150$$

Para el grupo de mujeres:

$$R = 320 - 80 = 240$$

Se puede concluir que la cantidad de contactos es más dispersa en el grupo de mujeres en comparación con el grupo de varones, puesto que presenta el mayor rango.

Sin embargo, esta medida de dispersión es relativa ya que se conoce poca información sobre los datos. Más adelante veremos otros estadígrafos que ofrecen una mejor perspectiva para determinar qué tan dispersos están los datos de una muestra.

7.2. Rango intercuartil

Es la diferencia entre el tercer y el primer cuartil de un conjunto de datos.

$$R_{IQ} = Q_3 - Q_1$$

Los datos que se ubican antes del primer cuartil son los que tienen valores más bajos y los datos que se ubican luego del tercer cuartil son los que tienen los valores más altos. El rango intercuartil deja fuera las observaciones con valores extremos y solo considera el 50% de datos con valores centrales.

Ejemplo: Se analiza la cantidad de platos que saben cocinar los chefs en el distrito de Chorrillos. Luego de hacer un tratamiento estadístico se obtiene el primer y tercer cuartil:

$$Q_1 = 32$$

$$Q_3 = 45$$

Con esta información se puede concluir que el 50% de los chefs sabe cocinar de 32 a 45 platos.

$$R_{IQ} = 45 - 32 = 13$$

La interpretación del rango intercuartil como medida de dispersión depende directamente de la magnitud del valor obtenido. Es decir, mientras mayor sea el rango intercuartil mayor dispersión presentarán los datos de la muestra. Esta última interpretación es un tanto relativa, ya que depende de

las variables que se esté analizando, la interpretación es muy útil cuando comparamos grupos de datos referentes a la misma variable.

Ejemplo: Se analiza el consumo mensual de galletas de los estudiantes de un colegio en los niveles de primaria y secundaria. Para conocer que tan variable es el consumo se obtiene el primer y tercer cuartil de cada nivel.

Tabla 7.5 Cuartiles por nivel educativo.

Primaria	Secundaria
$Q_1 = 10$	$Q_1 = 8$
$Q_3 = 30$	$Q_3 = 20$

Para el nivel primario:

$$R_{IQ} = 30 - 10 = 20$$

Para el nivel secundario:

$$R_{IQ} = 20 - 8 = 12$$

Se puede concluir que el consumo de mensual de galletas es más disperso en el nivel de educación primaria en comparación con el nivel de educación secundaria.

7.3. Desviación estándar

La desviación estándar o típica es sin duda una medida de dispersión muy importante, ya que además sirve como medida previa al cálculo de otros valores estadísticos (estadística inferencial).

La desviación típica se define como la raíz cuadrada de la media de los cuadrados de las desviaciones con respecto a la media de la distribución.

Hopkins et al. (1997) resaltan que la desviación estándar es de mucha utilidad para describir la variabilidad de un conjunto de datos y que tiene la ventaja de expresarse en las mismas unidades de la variable analizada (en comparación con otro estadígrafo llamado varianza). Cuanto mayor sea la desviación estándar, mayor será la variabilidad del conjunto de datos. Sobre las unidades, si los datos están expresados en Kg, entonces la unidad de la desviación estándar para esos datos también se expresa en kg.

Ejemplo: En la siguiente tabla se observa el gasto mensual en S/ por la compra de galletas que realizan dos grupos de personas:

Tabla 7.6 Consumo mensual de galletas por persona.

Grupo	A	B
Gastos	24	20
	30	18
	15	20
	25	18
	10	20
	30	19
Desviación estándar	8.16	0.98

En la tabla 7.6 note que el grupo A presenta una mayor variabilidad en los gastos en comparación con los gastos del grupo B. Para hacer esta comparación no es necesario observar cada uno de los gastos registrados, sino que basta con comparar las desviaciones estándar. 8.16 es mayor que 0.98, por lo tanto, se puede decir que el grupo A presenta gastos con valores más dispersos.

En cuanto a la interpretación de la desviación estándar, Moya (2007) señala que existe dificultad para realizar una interpretación intuitiva de este estadígrafo. Sin embargo, se puede decir que la desviación es el promedio de dispersión de los datos con respecto a la media.

7.3.1. Cálculo de la desviación estándar para datos no agrupados

La desviación estándar se calcula usando la siguiente expresión:

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

Dónde:

S: Desviación estándar

x_i : Datos

\bar{X} : Media aritmética

n : Tamaño de la muestra.

También se puede usar la siguiente expresión equivalente.

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

Ejemplo: Calcula la desviación estándar del siguiente conjunto de datos:

8 6 8 7 9 5

Tenemos 6 datos ($n=6$) con los cuales calculamos la media

$$\bar{X} = \frac{8 + 6 + 8 + 7 + 9 + 5}{6} = 7.17$$

Reemplazamos los valores en la primera fórmula:

$$S = \sqrt{\frac{(8 - 7.17)^2 + (6 - 7.17)^2 + (8 - 7.17)^2 + \dots + (5 - 7.17)^2}{6 - 1}}$$

$$S = 1.47$$

Si utilizamos la segunda fórmula:

$$\sum_{i=1}^n x_i^2 = 8^2 + 6^2 + 8^2 + 7^2 + 9^2 + 5^2 = 319$$

Para tener una mejor aproximación utilizamos 3 decimales para la media.

$$\bar{X} = 7.167$$

Reemplazamos los valores:

$$S = \sqrt{\frac{319 - 6 \times 7.167^2}{6 - 1}} = 1.47$$

7.3.2. Cálculo de la desviación estándar para datos agrupados

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

Dónde:

S : Desviación estándar

x_i : Valores de la variable X

f_i : Frecuencia absoluta

\bar{X} : Media aritmética

n : Tamaño de la muestra.

k : Número de intervalos

También se puede usar la siguiente expresión equivalente.

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

Ejemplo: En el siguiente cuadro se registra el consumo diario de proteínas de un grupo de niños.

Tabla 7.7 Consumo de proteína por niño.

Consumo (gr)	f_i
21	6
22	10
23	12
24	8
25	20
26	14
27	10
n=80	

Para calcular la desviación estándar primero debemos calcular la media. Por ello, primero multiplicamos el valor de cada consumo x_i con la frecuencia correspondiente.

Tabla 7.8 Consumo de proteína por niño.

Consumo (x_i)	f_i	$x_i f_i$
21	6	126
22	10	220
23	12	276
24	8	192
25	20	500
26	14	364
27	10	270
n=80		$\sum x_i f_i = 1948$

Luego dividimos la sumatoria de los productos con el tamaño de muestra.

$$\bar{X} = \frac{1948}{80} = 24.35$$

Para calcular la desviación estándar elevamos al cuadrado la diferencia de cada consumo con la media y multiplicamos el resultado con la frecuencia respectiva.

Tabla 7.9 Elementos necesarios para el cálculo de la desviación estándar.

Consumo (x_i)	f_i	$x_i - \bar{X}$	$(x_i - \bar{X})^2 f_i$
21	6	-3.35	67.335
22	10	-2.35	55.225
23	12	-1.35	21.87
24	8	-0.35	0.98
25	20	0.65	8.45
26	14	1.65	38.115
27	10	2.65	70.225
n=80		$\sum (x_i - \bar{X})^2 f_i = 262.2$	

Luego reemplazamos los valores en la primera fórmula para el cálculo de la desviación estándar:

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

$$S = \sqrt{\frac{262.2}{80 - 1}} = 1.82$$

Si se quiere utilizar la segunda fórmula para el cálculo de la desviación estándar, se debe multiplicar el cuadrado de cada consumo con la frecuencia respectiva.

Tabla 7.10 Elementos necesarios para el cálculo de la desviación estándar.

Consumo (x_i)	f_i	$x_i^2 f_i$
21	6	2646
22	10	4840
23	12	6348
24	8	4608
25	20	12500
26	14	9464
27	10	7290
n=80		$\sum x_i^2 f_i = 47696$

Luego reemplazamos en la los valores en la fórmula:

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

$$S = \sqrt{\frac{47696 - 80 \times 24.35^2}{80 - 1}} = 1.82$$

Interpretación: El promedio de dispersión de las cantidades de proteínas consumidas con respecto a la media es de 1.82 gr.

También se puede comparar la desviación estándar 1.82 con los valores registrados que están en el intervalo de 21 a 27. Obviamente 1.87 es mucho menor en comparación con los valores del intervalo, por lo tanto, se puede decir que existe poca variabilidad.

7.3.3. Cálculo de la desviación estándar para datos agrupados en intervalos

En este caso el procedimiento es igual al anterior, la única diferencia está en que primero se debe calcular la marca de clase x_i de cada intervalo.

Ejemplo: En el siguiente cuadro se registra el gasto mensual en bebidas gaseosas de un grupo de jóvenes.

Tabla 7.11 Gasto mensual en bebidas gaseosas por estudiante.

Masa	f_i
[22; 24>	4
[24; 26>	6
[26; 28>	8
[28; 30>	12
[30; 32>	10
[32; 34>	10
n=50	

Calculamos la marca de clase de cada intervalo.

Tabla 7.12 Elementos necesarios para el cálculo de la desviación estándar.

Masa	x_i	f_i	$x_i f_i$	x_i^2	$f_i x_i^2$
[22; 24>	23	4	92	529	2116
[24; 26>	25	6	150	625	3750
[26; 28>	27	8	216	729	5832
[28; 30>	29	12	348	841	10092
[30; 32>	31	10	310	961	9610
[32; 34>	33	10	330	1089	10890
n=50		$\sum x_i f_i = 1446$		$\sum f_i x_i^2 = 42290$	

Luego calculamos la media.

$$\bar{x} = \frac{1446}{50} = 28.92$$

Finalmente reemplazamos los valores en la segunda fórmula para el cálculo de la desviación estándar:

$$S = \sqrt{\frac{42290 - 50 \times 28.92^2}{49}} = 3.103$$

Interpretación: El promedio de dispersión del gasto mensual en bebidas gaseosas con respecto a la media es de $S/3.103$.

7.4. Varianza

La varianza se define como el cuadrado de la desviación estándar, es decir la varianza está determinada por la media aritmética del cuadrado de las desviaciones con respecto a la media (Moya, 2007). Algunos autores primero establecen el cálculo de la varianza y luego definen a la desviación estándar como la raíz cuadrada positiva de dicha varianza.

Por ejemplo, si la desviación estándar de los ingresos semanales de un grupo de trabajadores es $S/12.80$. Entonces, la varianza representada por S^2 es:

$$S^2 = 12.80^2 = 163.84$$

La varianza es un elemento importante en el análisis estadístico inferencial, es de mucha utilidad en pruebas de comparación y en los diseños experimentales.

7.5. Coeficiente de variación

En muchas ocasiones se requiere comparar la dispersión entre dos muestras. Pero, si la unidad de medida no es la misma en ambas muestras, en la comparación se pueden presentar errores de interpretación de los resultados.

Ejemplo: En la tabla 7.13 se tiene dos muestras sobre los resultados de dos pruebas con diferente forma de puntuación.

Tabla 7.13 Pruebas con diferente forma de puntuación.

Prueba	A	B
Puntaje mínimo establecido	0	0
Puntaje máximo establecido	10	50
Puntaje obtenido	7; 6; 10; 2	40; 42; 44; 48

Calculamos la desviación estándar para cada muestra.

$$S_1 = 3.30$$

$$S_2 = 3.41$$

Las desviaciones estándar de las muestras son muy parecidas. Sin embargo, no podemos concluir que ambas muestras presentan la misma dispersión, puesto que, los datos de la primera muestra evidentemente están más dispersos que los datos de la segunda muestra.

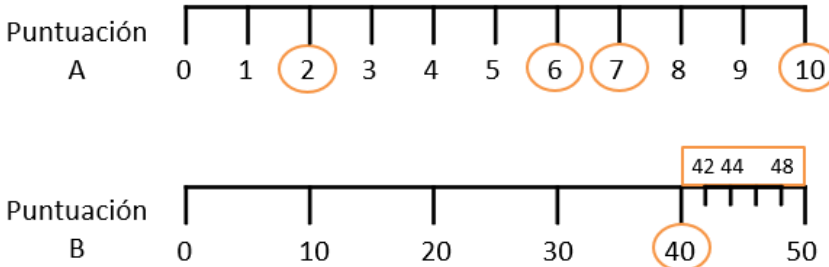


Figura 7.1 Observe que en el grupo con puntuación B hay menor dispersión, pues los valores están agrupados en el intervalo superior de 40 a 48 puntos.

El coeficiente de variación surge de la necesidad de contar con una medida de dispersión para comparar dos muestras que tienen diferente unidad de medida.

El coeficiente de variación permite conocer la magnitud relativa de la desviación estándar con respecto a la media de la distribución (Kasmier y Díaz, 1991). Entonces, el coeficiente de variación se usa para comparar conjuntos de datos pertenecientes a muestras distintas y para determinar la variabilidad de una sola muestra.

El coeficiente de variación (CV) se calcula con la siguiente fórmula:

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

Dónde:

\bar{X} : Media aritmética y S : Desviación estándar.

El coeficiente de variación se calcula por medio del cociente de la desviación estándar con la media y no presenta unidades ya que estas se anulan al efectuar la división. El CV se interpreta como el grado de variabilidad de un conjunto de datos y es independiente de la unidad de medición de la variable.

Martínez (2007) nos ofrece una forma de interpretar los valores que puede asumir el coeficiente de variación:

- $CV \leq 30\%$, los datos de la muestra son relativamente homogéneos. Por lo tanto, la media es representativa.
- $30\% < CV \leq 70\%$, los datos de la muestra son heterogéneos. Por lo tanto, la media no es representativa.
- $CV > 70\%$, los datos de la muestra son muy variables o heterogéneos. Por lo tanto, la media no es representativa.

Ejemplo: Calcula e interpreta el coeficiente de variación para los siguientes datos: 4, 5, 8; 16 y 20.

Primero debemos calcular la media y la desviación estándar.

$$\bar{X} = \frac{4 + 5 + 8 + 16 + 20}{5} = 10.6$$

$$S = \sqrt{\frac{(-6.6)^2 + (-5.6)^2 + (-2.6)^2 + 5.4^2 + 9.4^2}{5 - 1}} = 7.057$$

$$CV = \frac{7.057}{10.6} \times 100\% = 66.575\%$$

Interpretación: Como el CV es mayor al 30%, verificamos que la muestra es heterogénea y se concluye que la media no es representativa.

Ejemplo: Calcula e interpreta el coeficiente de variación para una distribución de costos donde la media es igual a S/ 450 y la desviación estándar es igual a S/ 64.

Reemplazamos los datos directamente en la fórmula:

$$CV = \frac{64}{450} \times 100\% = 14.222\%$$

Interpretación: Como el CV es menor al 30%, verificamos que la muestra es relativamente homogénea y se puede concluir que la media es representativa.

Capítulo 8: Medidas de forma de la distribución

Contenido

- ❖ Asimetría o sesgo
- ❖ Coeficiente de asimetría
- ❖ Curtosis
- ❖ Coeficiente de curtosis
- ❖ Diagrama de cajas

Cuando se analiza un conjunto de datos, en muchas ocasiones se recurre a los gráficos estadísticos para tener un acercamiento más simple a los resultados obtenidos. Las distribuciones de frecuencias mostradas en los gráficos adoptan diferentes formas y poseen significados particulares cuya interpretación también se puede obtener a través de otro grupo de estadísticos llamados coeficiente de asimetría y coeficiente de curtosis.

Por ejemplo, si se analiza la cantidad de libros que han leído dos grupos de estudiantes del nivel secundario de educación básica, se puede hacer una comparación gráfica de los resultados. En la figura 8.1 se presentan 2 escenarios. En el primer diagrama de barras se observa que las mayores frecuencias corresponden a los estudiantes que han leído una cantidad relativamente pequeña de libros. En el segundo diagrama de barras se observa que las frecuencias están distribuidas de manera más uniforme.

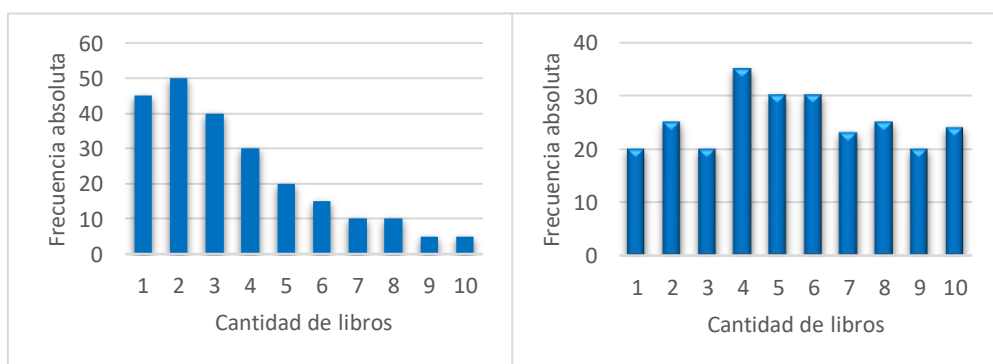


Figura 8.1 Cantidad de libros leídos por los estudiantes del nivel secundario.

8.1. Asimetría o sesgo

Es el grado en el que la distribución de datos se aleja de la simetría (Murray y Stephens, 2009). Una distribución simétrica es una distribución ideal donde se observa que las frecuencias equidistantes son iguales.

Moya (2007) define la asimetría como la deformación horizontal que presenta una distribución de frecuencias. Esta deformación surge de la comparación con la distribución simétrica.

A continuación, se muestran 3 casos generales que pueden observarse en una distribución gráfica de frecuencias. Tomemos el caso en que se analiza la calificación de los estudiantes en un examen introductorio.

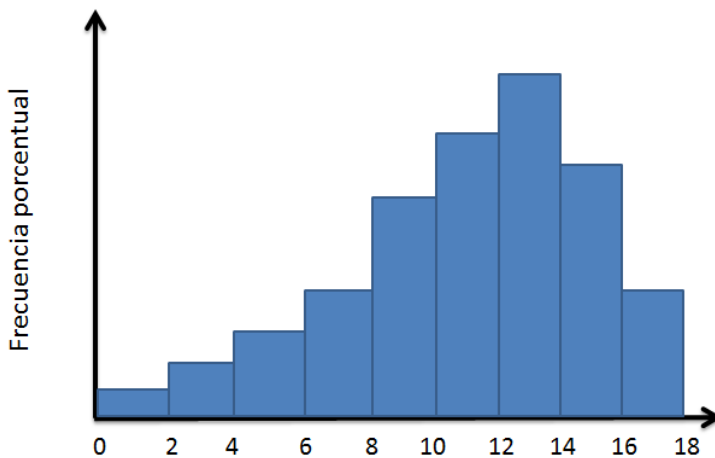


Figura 8.2 Histograma con alargamiento hacia la izquierda.

En la figura 8.2 se observa que las mayores frecuencias se ubican en la parte derecha de la distribución. Estamos frente a una distribución de datos con sesgo a la izquierda (asimetría negativa), puesto que gráficamente existe un alargamiento hacia los menores valores.

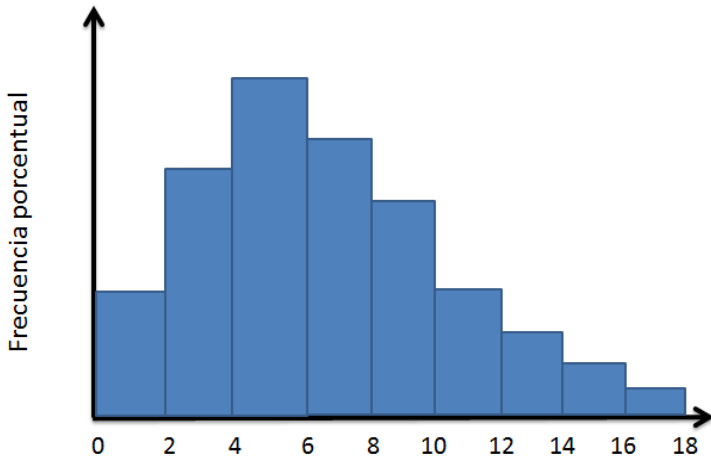


Figura 8.3 Histograma con alargamiento hacia la derecha.

En la figura 8.3 se observa que las mayores frecuencias se ubican en la parte izquierda de la distribución. Estamos frente a una distribución de datos con sesgo a la derecha (asimetría positiva), puesto que gráficamente existe un alargamiento hacia los valores mayores.

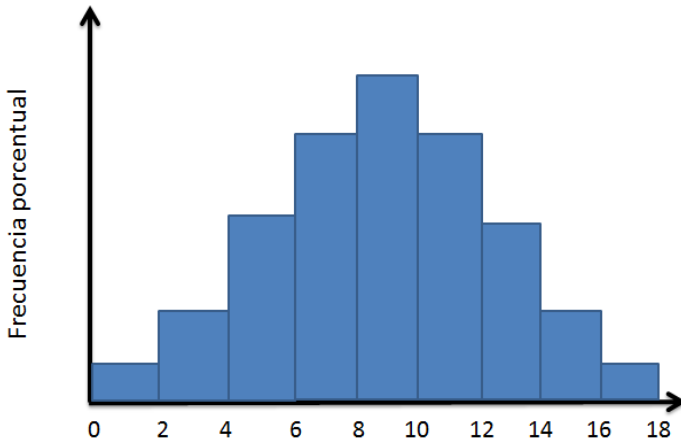


Figura 8.4 La distribución es simétrica.

En este último histograma de la figura 8.4 se observa que las columnas equidistantes son iguales, este caso corresponde a la denominada

distribución simétrica. Entonces, se puede decir que la distribución de datos no tiene sesgo.

La simetría también se puede observar utilizando polígonos de frecuencia.

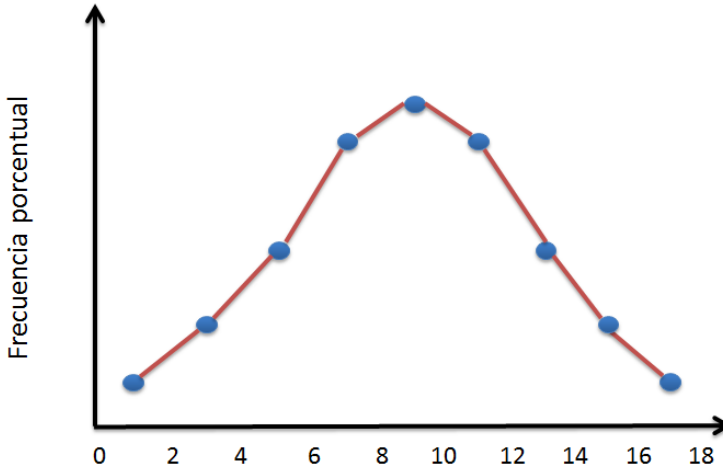


Figura 8.5 Recuerde que el polígono de frecuencias se construye con base en el histograma.

A continuación, se representan los polígonos de frecuencias suavizados, estas figuras también reciben el nombre de curvas de frecuencias.

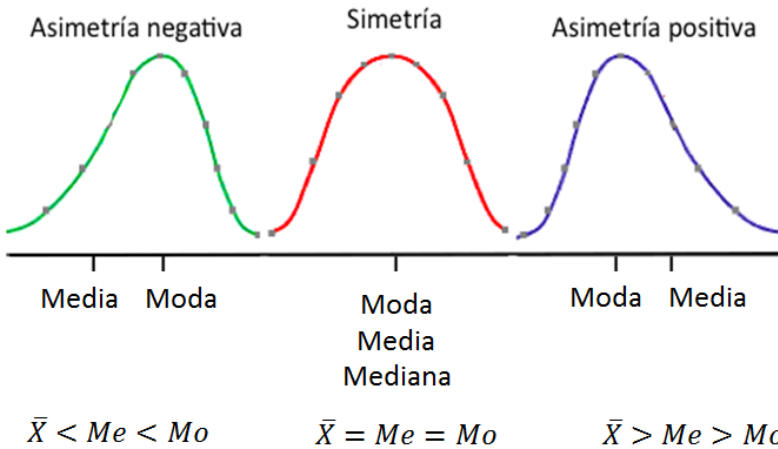


Figura 8.6 Una distribución puede presentar simetría o asimetría.

8.1.1. Coeficiente de asimetría

Para conocer si una distribución es simétrica o asimétrica se puede recurrir a un procedimiento alternativo al método gráfico. El coeficiente de asimetría es un valor cuyo signo determina el tipo de asimetría o la ausencia de ella utilizando estadígrafos ya conocidos como la media, la moda o la mediana.

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

Este coeficiente se puede interpretar de la siguiente manera:

$CA_S < 0$: Distribución asimétrica negativa.

$CA_S = 0$: Distribución simétrica.

$CA_S > 0$: Distribución asimétrica positiva.

Se debe resaltar que este coeficiente de asimetría (de Pearson) se utiliza en distribuciones de datos unimodales, es decir, con una sola moda.

Existe otra expresión equivalente para el cálculo del coeficiente de asimetría:

$$CA_S = \frac{3(\bar{X} - Me)}{S}$$

Ejemplo: Se realizó una encuesta a un grupo de padres de familia para determinar la cantidad de veces a la semana que consumen productos ultra procesados. Los resultados se muestran en la siguiente tabla:

Tabla 8.1 Consumo de productos ultra procesados.

Cantidad de veces	f_i
0	5
1	8
2	12
3	15
4	20

5	22
6	10
7	8
n=100	

Para calcular el coeficiente de asimetría utilizaremos la fórmula del coeficiente de asimetría de Pearson:

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

La mayor frecuencia es 22 por lo tanto $Mo = 5$. Para calcular la media y la desviación estándar debemos completar la tabla.

Tabla 8.2 Proceso para el cálculo de la media y la desviación estándar.

x_i	f_i	$x_i f_i$	$x_i^2 f_i$
0	5	0	0
1	8	8	8
2	12	24	48
3	15	45	135
4	20	80	320
5	22	110	550
6	10	60	360
7	8	56	392
n=100		$\sum x_i f_i = 383$	$\sum x_i^2 f_i = 1813$

Calculamos la media:

$$\bar{X} = \frac{383}{100} = 3.83$$

Calculamos la desviación estándar:

$$S = \sqrt{\frac{1813 - 100 \times 3.83^2}{100 - 1}} = 1.87$$

Reemplazando en la fórmula:

$$CA_S = \frac{3.83 - 5}{1.87} = -0.63$$

Interpretación: El signo del coeficiente nos permite concluir que la distribución de datos es sesgada a la izquierda, es decir, presenta asimetría negativa.

Existen programas estadísticos que facilitan los procesos para determinar el coeficiente de asimetría y la representación gráfica de esta. Por ejemplo, si utilizamos SPSS el programa nos indica el valor del estadígrafo y también la distribución de forma gráfica.

Tabla 8.3 Estadísticos descriptivos utilizando SPSS.

Estadístico	Valor
n	100
Media	3,8300
Moda	5,00
Desviación estándar	1,86978
Asimetría	-0,24

En la tabla 8.3 se pueden verificar los resultados del ejemplo anterior, pero el valor del coeficiente de asimetría es diferente. Esto se debe a que existen otras formas de calcular el coeficiente de asimetría.

La siguiente es la fórmula del coeficiente de asimetría de Fisher:

$$CA_S = \frac{\sum_i^n (x_i - \bar{X})^3 f_i}{n \times S^3}$$

También se puede utilizar la fórmula del coeficiente de asimetría de Bowley:

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

No nos debe preocupar la elección del coeficiente de asimetría adecuado, pues los programas estadísticos nos brindan el valor exacto. Es cierto que el

coeficiente de asimetría de Pearson es más sencillo de calcular, pero se utiliza en distribuciones con poca asimetría.

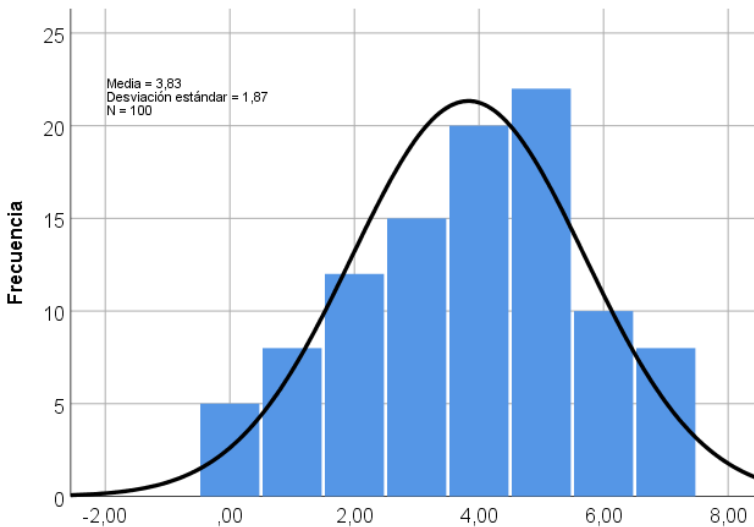


Figura 8.7 Distribución del consumo de productos ultra procesados.

En la figura 8.7 se puede observar un sesgo hacia la izquierda. Entonces, de manera gráfica se verifica la asimetría negativa determinada en el ejemplo.

Ejemplo: Si analizamos la variable masa corporal de un grupo conformado por 30 estudiantes. El programa SPSS nos brinda la siguiente tabla con el valor del coeficiente de asimetría:

Tabla 8.4 Estadísticos descriptivos utilizando SPSS.

Estadísticos	
Variable	Masa corporal
n	30
Asimetría	1.363
Error estándar de asimetría	0.427

Interpretación: El signo del coeficiente 1.363 nos permite concluir que la distribución de datos es sesgada a la derecha, es decir, presenta asimetría positiva.

La gráfica que nos brinda el programa nos permite verificar y orientarnos para comprender el tipo de asimetría que presenta la distribución de datos. En la figura 8.8 se observa el sesgo o estiramiento hacia la derecha.

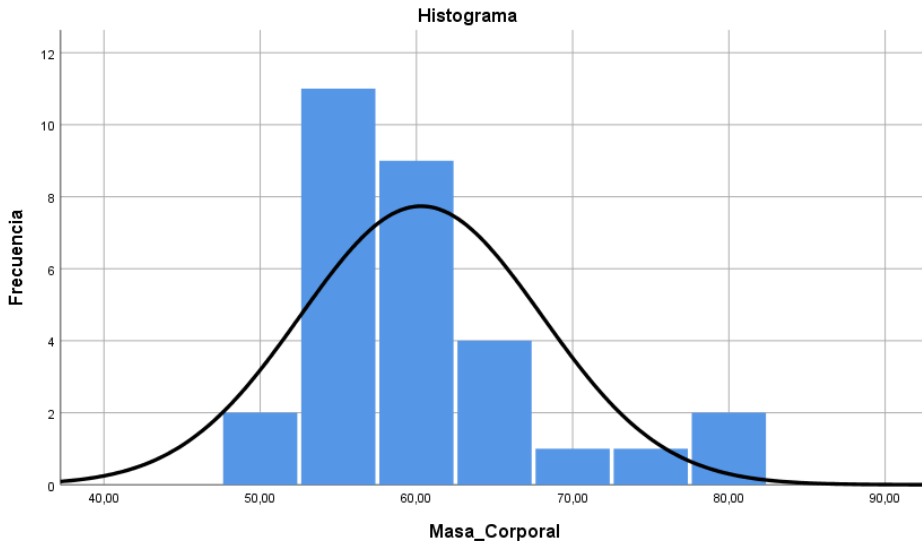


Figura 8.8 Distribución de la masa corporal.

8.2. Curtosis

Moya (2007) define la curtosis como la deformación vertical que presenta una distribución de frecuencias. Esta deformación surge de la comparación con la distribución simétrica.

A continuación, se muestran 3 casos generales que pueden observarse en una distribución gráfica de frecuencias. Así como en la asimetría, tomemos el caso en que se analiza la calificación de los estudiantes en un examen introductorio.

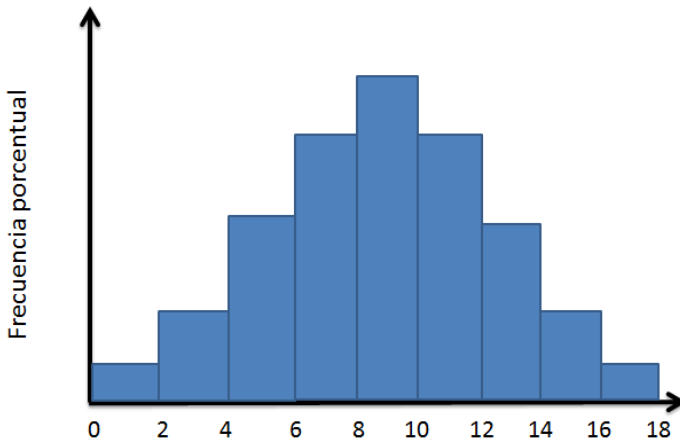


Figura 8.9 Distribución normal.

La figura 8.9 se toma como referencia para ilustrar la distribución normal que tiene forma de campana.

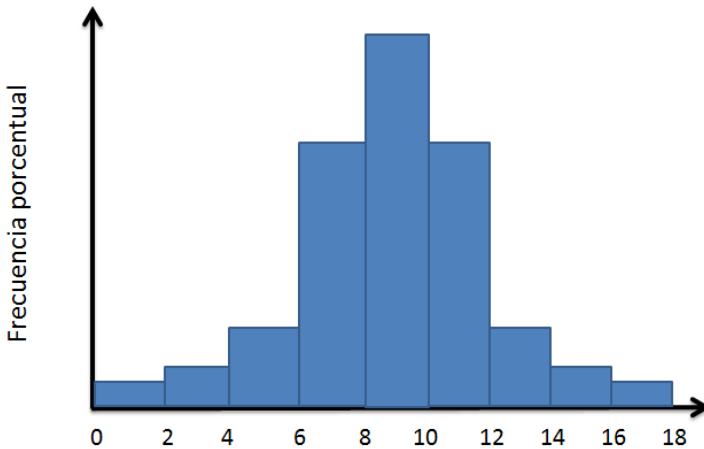


Figura 8.10 Observe que las frecuencias de la parte central presentan mayor altura.

En la figura 8.10 se observa que la mayor cantidad de datos se ubica en la parte central, es decir, son muy pocos los estudiantes que obtienen calificaciones muy altas o muy bajas. También se puede decir que la curva de frecuencia presenta una parte central muy pronunciada.

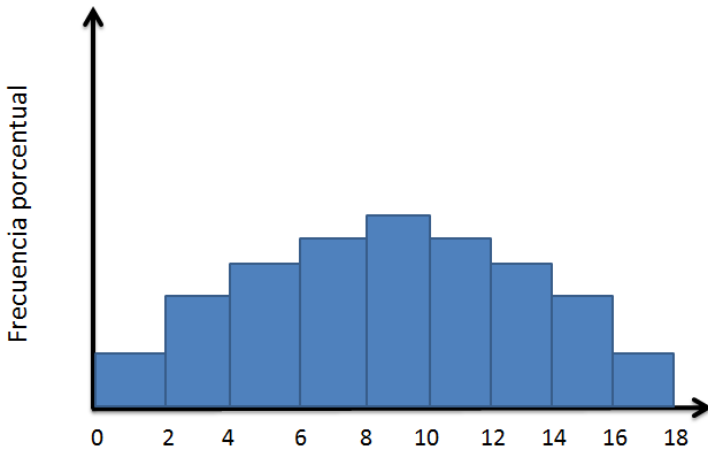


Figura 8.11 Observe que la parte central presenta un aplastamiento en comparación con las distribuciones de las figuras 8.8 y 8.9.

En la figura 8.11 se observa que los datos se distribuyen de manera más uniforme, es decir, no hay una concentración extraordinaria en ninguno de los intervalos. Se puede decir que la curva de frecuencia no es muy pronunciada en su parte central.

A continuación, se presentan los polígonos de frecuencias suavizados y la denominación de cada tipo de curtosis.

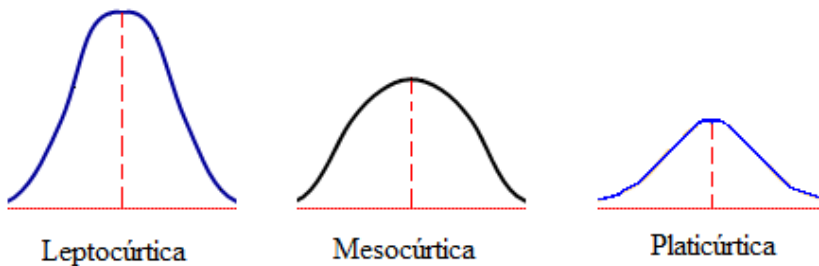


Figura 8.12 Tipos de curtosis.

La distribución normal es la distribución mesocúrtica. En la distribución leptocúrtica la gráfica es presenta un mayor apuntamiento en comparación con la distribución normal. En la distribución platicúrtica la gráfica presenta un mayor achatamiento en comparación con la distribución normal (Spiegel y Stephens, 2009).

Gorgas et al. (2011) describen a la curtosis como la forma del agrupamiento de los datos en torno al valor central. En ese sentido, una distribución leptocúrtica presenta una mayor cantidad de datos en los valores centrales en comparación con la distribución normal. Una distribución platicúrtica presenta una menor cantidad de datos en los valores centrales en comparación con la distribución normal.

8.2.1. Coeficiente de curtosis

Para conocer el grado de deformación vertical se pueden utilizar diferentes expresiones. El coeficiente de curtosis es un valor que determina el tipo de curtosis utilizando estadígrafos ya conocidos como la media, la desviación estándar, los momentos, los cuartiles y los percentiles.

Para calcular la curtosis en función de los momentos:

$$K_1 = \frac{\sum f_i(x_i - \bar{X})^4}{n S^4} = \frac{M_4}{M_2^2}$$

$$K_2 = \frac{M_4}{S^4} - 3$$

Para calcular la curtosis en función de los cuartiles y percentiles:

$$K_3 = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

Para realizar el siguiente ejemplo utilizaremos la tercera fórmula. Por ello, la interpretación de este coeficiente es la siguiente (Moya, 2007):

$K_3 < 0.263$: Distribución platicúrtica.

$K_3 = 0.263$: Distribución mesocúrtica.

$K_3 > 0.263$: Distribución leptocúrtica.

Ejemplo: Utilizaremos la información de la siguiente tabla que hace referencia a la frecuencia de consumo semanal de alimentos ultra procesados.

Tabla 8.5 Frecuencia de consumo de alimentos ultra procesados.

Cantidad de veces	f_i	F_i
0	10	10
1	24	34
2	20	54
3	22	76
4	30	106
5	15	121
6	10	131
7	2	133
n=133		

Para calcular la curtosis K3 debemos calcular Q_3, Q_1, P_{90} y P_{10} .

Para calcular el primer cuartil:

$$F_{Q_1} \geq \frac{133}{4}$$

$$F_{Q_1} \geq 33.25$$

La primera frecuencia acumulada que es mayor o igual a 33.25 es 34. Por lo tanto, $Q_1 = 1$

Para calcular el tercer cuartil:

$$F_{Q_3} \geq \frac{3 \times 133}{4}$$

$$F_{Q_3} \geq 99.75$$

La primera frecuencia acumulada que es mayor o igual a 99.75 es 106. Por lo tanto, $Q_3 = 4$

Para calcular el percentil 10:

$$F_{P_{10}} \geq \frac{10 \times 133}{100}$$

$$F_{P_{10}} \geq 13.3$$

La primera frecuencia acumulada que es mayor o igual a 13.3 es 34. Por lo tanto, $P_{10} = 1$

Para calcular el percentil 90:

$$F_{P_{90}} \geq \frac{90 \times 133}{100}$$

$$F_{P_{90}} \geq 119.7$$

La primera frecuencia acumulada que es mayor o igual a 119.7 es 121. Por lo tanto, $P_{90} = 5$

Reemplazando los cuartiles y percentiles calculados en la tercera fórmula:

$$K_3 = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

$$K_3 = \frac{4 - 1}{2(5 - 1)} = 0.375$$

Interpretaciones: Se puede utilizar cualquiera de las siguientes interpretaciones:

El valor calculado 0.375 es mayor que 0.263. Por lo tanto, la distribución de datos es leptocúrtica. Es decir, la distribución presenta un alargamiento considerable en la parte central.

El coeficiente de curtosis es mayor que 0.263. por lo tanto, la distribución presenta un mayor apuntamiento en comparación con la distribución normal.

El coeficiente de curtosis es mayor que 0.263, esto indica que una gran cantidad de datos se ubica en la parte central de la distribución.

Ejemplo: Si analizamos la variable masa corporal de un grupo conformado por 30 estudiantes. El programa SPSS nos brinda la siguiente tabla:

Tabla 8.6 Estadísticos descriptivos utilizando SPSS.

Estadísticos	
Variable	Masa corporal
n	30
Curtosis	1.903
Error estándar de curtosis	0.833

En el programa SPSS (IBM Knowledge Center, s.f.) se utiliza la siguiente interpretación:

- ❖ Si el coeficiente de curtosis es igual a 0 la distribución es normal.
- ❖ Si el coeficiente de curtosis es menor que 0 la distribución presenta menos valores atípicos extremos que una distribución normal.
- ❖ Si el coeficiente de curtosis es mayor que 0 la distribución presenta más valores atípicos extremos que una distribución normal.

Interpretación: Según la tabla 8.6 el coeficiente de curtosis 1.903 indica que la distribución presenta más valores atípicos extremos que una distribución normal.

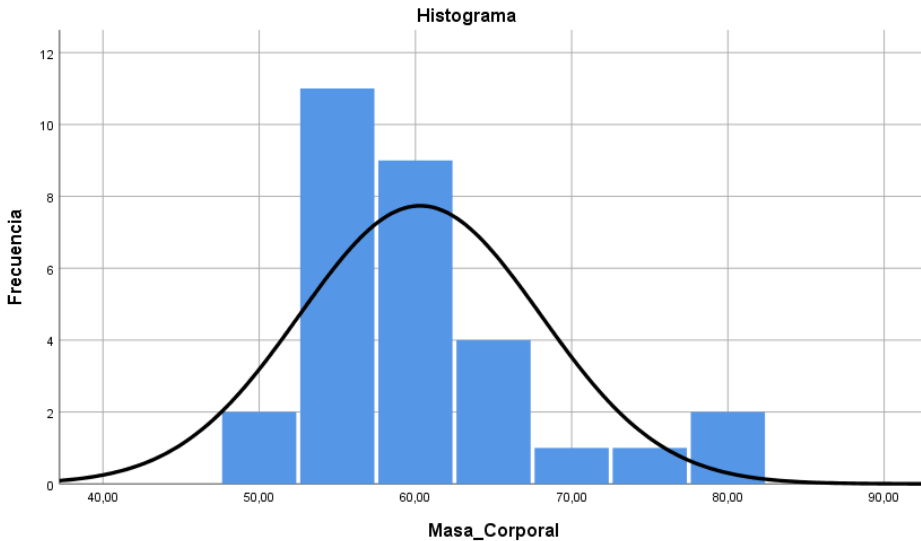


Figura 8.13 Distribución de la masa corporal.

La distribución de masas corporales mostrada en la figura 8.13 presenta un mayor apuntamiento que la curva normal.

Ejemplo: Si analizamos la variable edad de un grupo conformado por 98 individuos. El programa SPSS nos brinda la siguiente tabla:

Tabla 8.7 Estadísticos descriptivos utilizando SPSS.

Estadísticos	
Variable	Edad
n	98
Curtosis	-1.027
Error estándar de curtosis	0.483

Interpretación: Según la tabla 8.7, el coeficiente de curtosis es negativo, por lo tanto, la distribución de datos es platicúrtica. Es decir, distribución presenta menos valores atípicos extremos que una distribución normal.

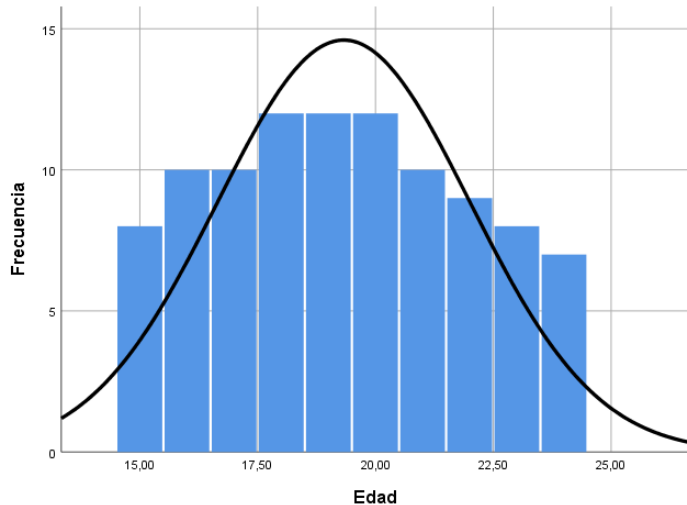


Figura 8.14 Distribución de las edades.

La distribución de edades mostrada en la figura 8.14 presenta un menor apuntamiento que la curva normal.

8.3. Diagrama de cajas

Es una representación semi gráfica de una distribución construida para mostrar sus características principales (Moya, 2007). En este diagrama se pueden observar los valores atípicos, estos son aquellos valores que son muy diferentes a los demás. Considere que el diagrama de cajas se utiliza para el análisis de variables numéricas.

Anderson et al. (2012) señalan que un diagrama de cajas es un resumen gráfico de los datos con base en el resumen de cinco números: el valor mínimo, el valor máximo, el primer cuartil, la mediana y el tercer cuartil.

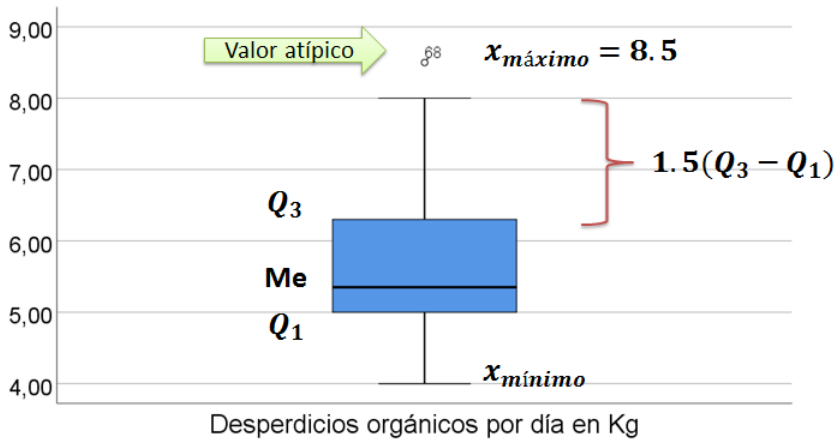


Figura 8.15 Diagrama de cajas elaborado con SPSS.

Moya (2007) indica que para elaborar un diagrama de cajas se debe seguir el siguiente procedimiento:

- 1) Del conjunto de datos se obtiene el valor mínimo, el valor máximo, el primer cuartil, la mediana y el tercer cuartil.
- 2) Sobre un eje de referencia se grafica un rectángulo cuyos extremos coinciden con los valores del primer y tercer cuartil. Dentro del rectángulo se traza un segmento cuya ubicación coincide con el valor de la mediana.

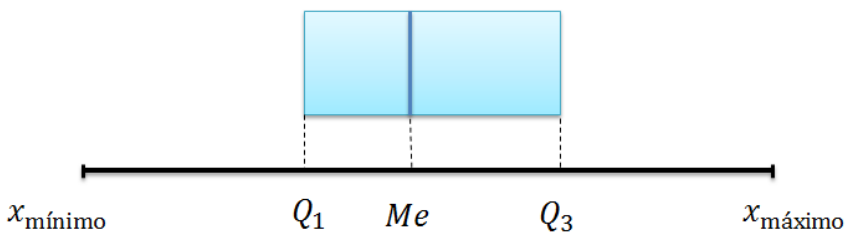


Figura 8.16 Ubicación de la caja en la distribución.

- 3) Se calculan los límites admisibles inferior y superior.

$$\text{Límite inferior} = Q_1 - 1.5(Q_3 - Q_1)$$

$$\text{Límite superior} = Q_3 + 1.5(Q_3 - Q_1)$$

- 4) Se grafican segmentos que van desde los extremos del rectángulo hasta los límites inferior y superior.

Al realizar este procedimiento se pueden dar los siguientes casos:

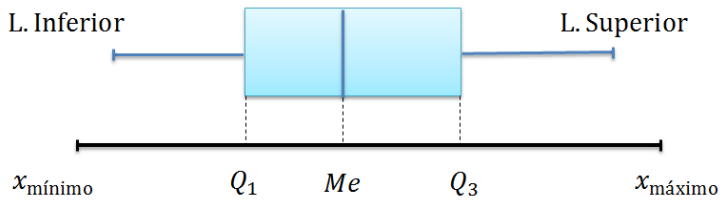


Figura 8.17 Diagrama de cajas con valores atípicos en ambos extremos de la distribución.

En la figura 8.17 se observa que existen valores menores al límite inferior y valores mayores al límite superior, es decir existen valores atípicos en los extremos de la distribución.

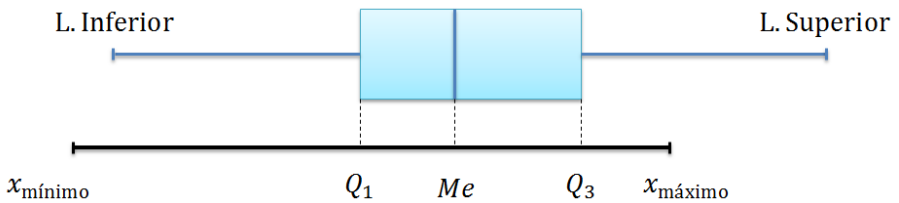


Figura 8.18 Observe que el valor mínimo es menor que el límite inferior.

En la figura 8.18 se observa que existen valores menores al límite inferior pero no existen valores mayores al límite superior, es decir, solo existen valores atípicos en la parte inferior de la distribución.

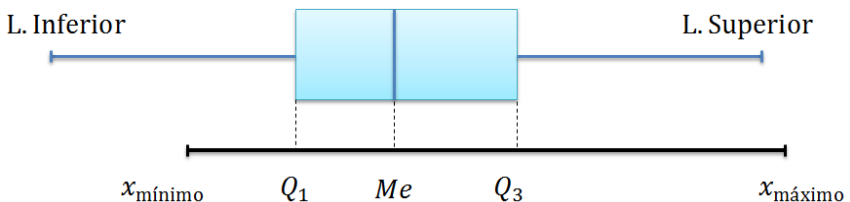


Figura 8.19 Observe que el valor máximo es mayor que el límite superior.

En la figura 8.19 se observa que existen valores mayores al límite superior pero no existen valores menores al límite inferior, es decir, solo existen valores atípicos en la parte superior de la distribución.

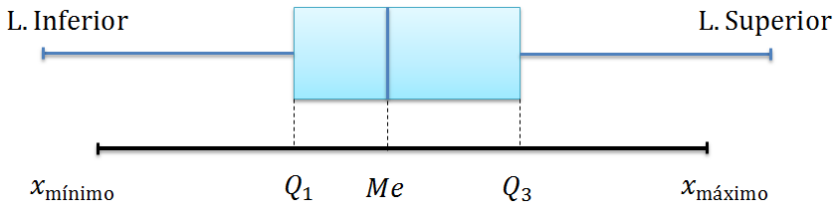


Figura 8.20 Diagrama de cajas sin valores atípicos.

En la figura 8.20 se observa que no existen valores menores al límite inferior ni valores mayores al límite superior, es decir no existen valores atípicos en la distribución.

- 5) Los valores que son menores que el límite inferior o mayores que el límite superior deben ser resaltados como valores atípicos.

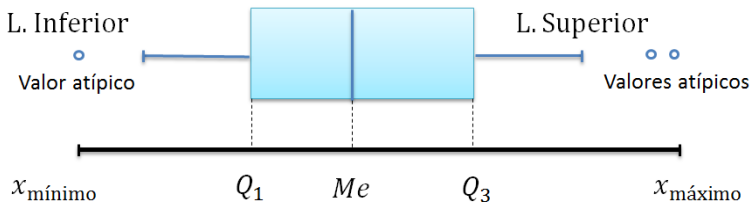


Figura 8.21 Observe como se resalta la ubicación de los valores atípicos.

Ejemplo: Se realiza una encuesta para conocer las edades de las personas que consumen, con mayor frecuencia, los productos de una pastelería. Los resultados son los siguientes:

30	32	45	25	24	24	24	25	20	20
26	30	32	30	24	25	24	20	18	32
28	30	30	25	24	24	24	24	20	28
25	24	23	22	22	23	22	25	20	25

Estos datos se pueden ordenar en una tabla:

Tabla 8.8 Distribución de las edades.

Edad	f_i	F_i	
18	1	1	
20	5	6	
22	3	9	
23	2	11	F_{Q_1}
24	10	21	F_m
25	7	28	
26	1	29	
28	2	31	F_{Q_3}
30	5	36	
32	3	39	
45	1	40	
Total	40		

Observe que el valor mínimo y el valor máximo para las edades son 18 y 45.

Para calcular los cuartiles y la mediana hacemos

Frecuencia para el primer cuartil: $F_{Q_1} \geq \frac{40}{4}$

$$F_{Q_1} \geq 10$$

Frecuencia para el segundo cuartil (Mediana): $F_{Q_2} \geq 2 \left(\frac{40}{4} \right)$

$$F_{Q_2} \geq 20$$

Frecuencia para el tercer cuartil: $F_{Q_3} \geq 3 \left(\frac{40}{4} \right)$

$$F_{Q_3} \geq 30$$

Luego, tenemos los valores $Q_1 = 23$ $Me = 24$ y $Q_3 = 28$

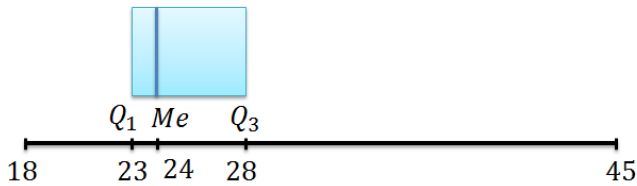


Figura 8.22 Ubicación de la caja en la distribución.

Se calculan los límites admisibles inferior y superior.

$$\text{Límite inferior} = 23 - 1.5(28 - 23) = 15.5$$

$$\text{Límite superior} = 28 + 1.5(28 - 23) = 35.5$$

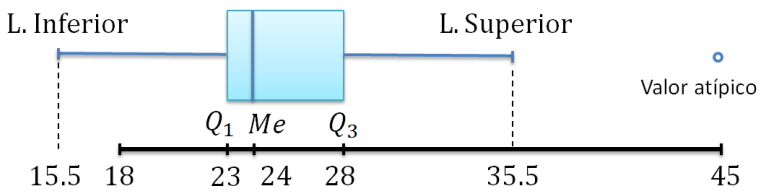


Figura 8.23 Diagrama de cajas con la información completa.

Se grafican los segmentos que van desde los extremos del rectángulo hasta los límites inferior y superior, luego se resalta el valor atípico. Observe que en la distribución solo existe un valor que es mayor que el límite superior. Este valor atípico es el 45, es decir, en la distribución de edades existe una persona cuya edad es muy superior a las edades de los demás integrantes del grupo.

La presentación del diagrama de cajas sobre un eje horizontal no es la única forma en que se puede utilizar este esquema. También se puede utilizar un diagrama de cajas vertical como se muestra en el siguiente gráfico:

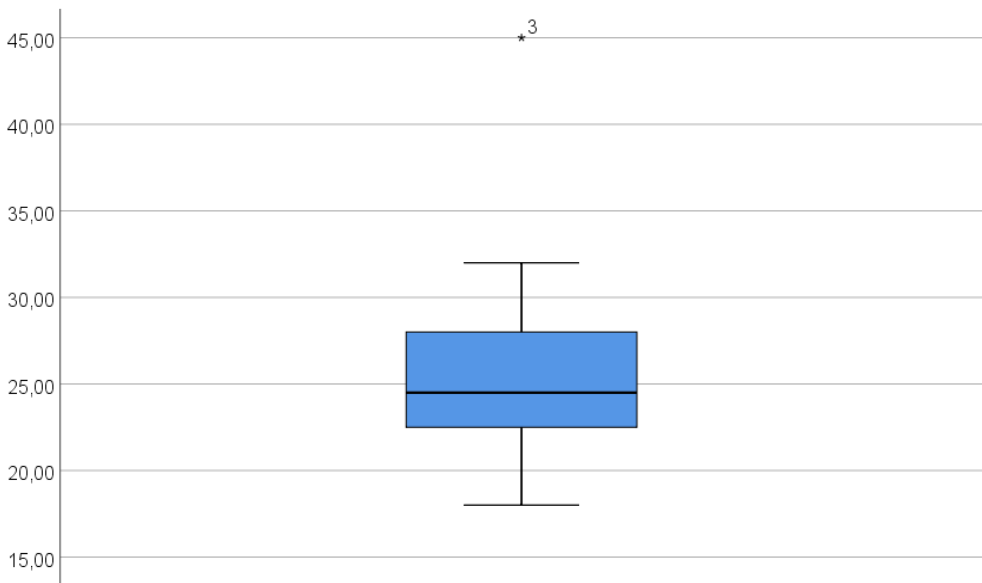


Figura 8.24 Diagrama de cajas con orientación vertical.

Observe que en el gráfico 8.24 se muestra un solo valor atípico señalado con el número 3. Este valor atípico es el 45, el número 3 indica la posición en que se encuentra este valor en el cuadro inicial de datos.

El diagrama de cajas es muy útil cuando se quiere hacer comparaciones entre grupos de datos, permite hacer una comparación visual y aporta otros elementos específicos como los cuartiles y la concentración de datos entre el primero y tercero de ellos.

Ejemplo: Un docente evalúa a sus estudiantes por medio de una prueba de conocimientos. Pero quiere hacer una comparación entre los resultados de dos aulas distintas. Los resultados de la prueba de conocimientos fueron los siguientes:

Tabla 8.9 Calificaciones de los estudiantes del aula “A” y el aula “B”.

Aula A		Aula B	
Estudiante	Calificación	Estudiante	Calificación
1	18	1	12
2	14	2	14
3	15	3	17
4	18	4	16
5	15	5	15
6	17	6	15
7	13	7	14
8	11	8	15
9	11	9	11
10	14	10	16
11	13	11	16
12	14	12	17
13	13	13	15
14	14	14	18
15	14	15	16
16	12	16	17
17	14	17	18
18	12	18	16
19	11		
20	10		

Utilizando el programa estadístico SPSS se obtiene el gráfico con los diagramas de cajas para hacer la comparación entre las distribuciones de datos.

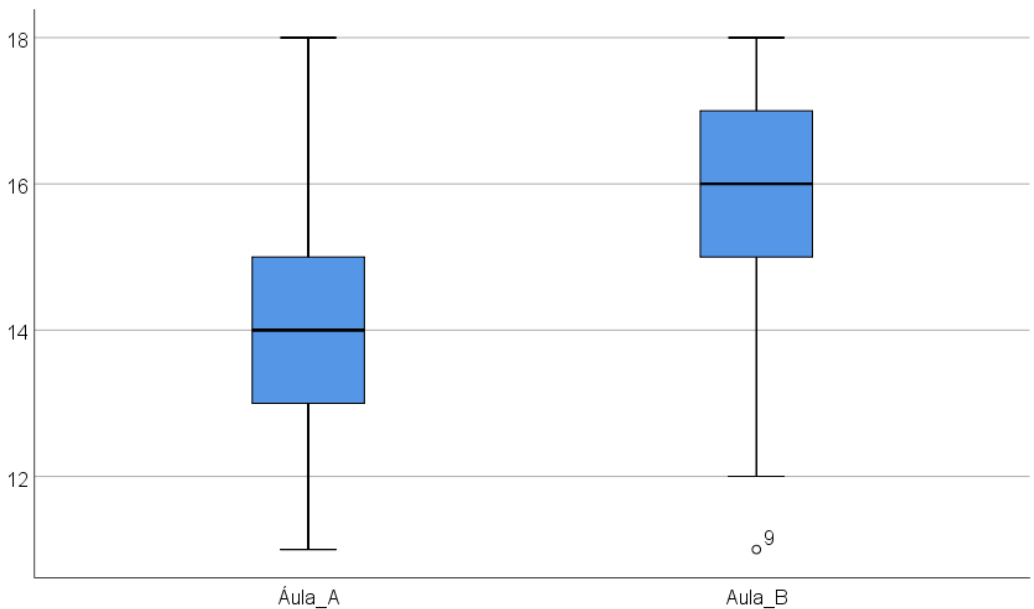


Figura 8.25 Gráfico comparativo de dos distribuciones.

Interpretaciones:

En la figura 8.25 se puede apreciar que la mediana de las calificaciones en el grupo A es menor que la mediana de las calificaciones en el grupo B. Es decir, la mitad de los estudiantes del Aula A tienen como mínimo 14 de calificación mientras que la mitad de los estudiantes del grupo B tienen como mínimo 16 de calificación.

Si analizamos el tercer cuartil se observa que el 25% de los estudiantes del grupo A tiene por lo menos 15 de calificación mientras que el 25% de los estudiantes del grupo B tienen por lo menos 17 de calificación. Además, se puede ver que en el grupo A los valores son más homogéneos, es decir, no se observan valores atípicos, en cambio en el grupo B si se registra un valor atípico en la posición 9 de los datos de la tabla inicial. Este valor atípico es el 11.

Capítulo 9: Análisis descriptivo de los resultados de la investigación

Contenido

- ❖ ¿Qué resultados se deben presentar en el informe?
- ❖ Criterios básicos para el análisis descriptivo.
- ❖ Análisis descriptivo de variables cualitativas
- ❖ Análisis descriptivo de variables cuantitativas

En el siguiente apartado se presentan diferentes ejemplos de resultados de investigación con datos elegidos al azar, es decir, los datos mostrados no corresponden a un estudio previo o a una base de datos extraída de alguna institución en particular. Con los ejemplos se muestra al investigador los elementos que puede seleccionar para presentar sus resultados dentro del informe de investigación. Recuerde que, en el informe, los resultados estadísticos se utilizan para sustentar la discusión, las conclusiones y recomendaciones.

En los capítulos anteriores se ha definido y ejemplificado los estadígrafos, tablas y gráficos que se utiliza en la estadística descriptiva. Corresponde al investigador decidir qué elemento antes señalado es más pertinente para presentar sus resultados. Cabe resaltar que el análisis descriptivo es necesario en toda investigación cuantitativa pues representa un acercamiento inicial al comportamiento de las variables de estudio.

9.1. ¿Qué resultados se deben presentar?

Lerma (2011) resalta que los resultados se deben presentar en orden lógico, usando tablas, figuras, estadígrafos y las pruebas estadísticas respectivas. En este texto no se aborda el tema de las pruebas estadísticas, solo analizaremos la presentación de los resultados a nivel descriptivo. El orden lógico hace referencia a la presentación secuencial de los resultados de tal manera que estos se relacionen con los objetivos de la investigación.

Se debe evitar la inclusión de gráficos y tablas de forma innecesaria, estos solo se utilizan cuando ayudan a una mejor comprensión del problema analizado. La información que es complementaria se puede redactar de manera resumida y sus respectivas tablas o gráficos pueden ubicarse en la parte de los anexos.

Hernández, Ramos et al., (2018) indican que el análisis de los resultados implica un examen particular de cada elemento de la información obtenida para dar respuesta a las interrogantes de la investigación. Esto último es importante, puesto que la información recabada no solo tiene que ser útil para responder al problema general de la investigación, sino también, debe disgregarse para responder a los problemas específicos.

Entonces, los resultados que se muestran deben tener relación directa con los objetivos de la investigación de forma general y específica. La lectura de los elementos estadísticos resultantes debe hacerse de manera clara en concordancia con la estructura de la investigación.

Los resultados que se presentan también dependen del tipo de variable o el nivel de medición. En general, dentro de los resultados se muestran los siguientes elementos con la interpretación respectiva:

- ❖ Tablas.
- ❖ Gráficos.
- ❖ Estadígrafos.
- ❖ Estimación de parámetros.
- ❖ Pruebas de hipótesis.

En este libro solo trabajaremos con tablas, gráficos estadísticos y estadígrafos. La estimación de parámetros y la prueba de hipótesis son parte de la estadística inferencial.

9.2. Criterios básicos para el análisis descriptivo

En la siguiente tabla se sugiere las frecuencias y los gráficos que se pueden utilizar según el tipo de variable.

Tabla 9.1 Frecuencias y gráficos según el tipo de variable.

Variable	Frecuencias para la tabla	Gráficos
Cualitativa	F. Absoluta F. Relativa F. Porcentual	Diagrama circular Diagrama de barras
Cuantitativa	F. Absoluta F. Relativa F. Porcentual F. Absoluta acumulada F. Relativa acumulada	Diagrama de barras
Cuantitativa (tabla con intervalos)	Todas las frecuencias	Histograma Polígono de frecuencias Diagrama escalonado Ojiva

Se debe elegir entre utilizar una tabla o un gráfico cuando la información contenida en ambos es la misma. Tiene sentido utilizar una tabla con un gráfico cuando la información de este último complementa la información contenida en la tabla. Para las tablas deben seleccionarse las frecuencias necesarias para explicar los resultados de forma precisa. En el caso de los gráficos se debe elegir aquel gráfico que resuma y oriente mejor la lectura de los resultados.

La siguiente tabla es una adaptación del cuadro elaborado por Juárez, Villatoro y López (2002), en él se indican los estadígrafos adecuados para cada tipo de variable.

Tabla 9.2 Estadísticos adecuados según la escala de medición o el tipo de variable.

Escala de medición	Tipo de variable	Estadísticos
Nominal	Nominal	Moda y tanto por ciento
Ordinal	Ordinal	Moda, tanto por ciento y mediana
Intervalo	Discreta o continua	Moda, tanto por ciento y mediana, media, cuartiles, deciles, percentiles, desviación estándar, varianza, coeficiente de variación, coeficiente de curtosis, coeficiente de asimetría y diagrama de cajas.
Razón o proporción		

9.3. Análisis descriptivo para variables cualitativas

9.3.1. Análisis descriptivo para variable nominal

Ejemplo No. 1: Se realiza una encuesta con la finalidad de conocer si existe asociación entre el consumo de café y la jornada de trabajo de los docentes de la universidad “V”. Para ello, se realizó una encuesta y los resultados fueron los siguientes:

Tabla 9.3 Datos referentes al consumo de café y la jornada de trabajo.

No.	Jornada	Café	No.	Jornada	Masa	No.	Jornada	Café	No.	Jornada	Café
1	TP	Si	24	TP	No	47	TP	No	70	TC	Si
2	TP	Si	25	TP	No	48	TP	No	71	TC	Si
3	TP	Si	26	TP	No	49	TP	No	72	TC	Si
4	TP	Si	27	TP	No	50	TP	No	73	TC	Si
5	TP	Si	28	TP	No	51	TP	No	74	TC	Si
6	TP	Si	29	TP	No	52	TP	No	75	TC	Si
7	TP	Si	30	TP	No	53	TP	No	76	TC	Si
8	TP	Si	31	TP	No	54	TP	No	77	TC	Si
9	TP	Si	32	TP	No	55	TP	No	78	TC	Si
10	TP	Si	33	TP	No	56	TP	No	79	TC	Si
11	TP	Si	34	TP	No	57	TC	Si	80	TC	Si
12	TP	Si	35	TP	No	58	TC	Si	81	TC	Si
13	TP	Si	36	TP	No	59	TC	Si	82	TC	Si
14	TP	Si	37	TP	No	60	TC	Si	83	TC	No
15	TP	Si	38	TP	No	61	TC	Si	84	TC	No
16	TP	Si	39	TP	No	62	TC	Si	85	TC	No
17	TP	Si	40	TP	No	63	TC	Si	86	TC	No
18	TP	Si	41	TP	No	64	TC	Si	87	TC	No
19	TP	No	42	TP	No	65	TC	Si	88	TC	No
20	TP	No	43	TP	No	66	TC	Si	89	TC	No
21	TP	No	44	TP	No	67	TC	Si	90	TC	No
22	TP	No	45	TP	No	68	TC	Si	91	TC	No
23	TP	No	46	TP	No	69	TC	Si	92	TC	No

Consideraciones metodológicas

Problema de investigación: ¿Existe asociación entre el consumo de café y la jornada de trabajo de los docentes de la universidad “V”?

Objetivo de la investigación: Conocer si existe asociación entre el consumo de café y la jornada de trabajo de los docentes de la universidad “V”.

Hipótesis de la investigación: Existe asociación entre el consumo de café y la jornada de trabajo de los docentes de la universidad “V”.

Variable 1: Consumo de café.

Variable 2: Jornada de trabajo

Escala de medición: Ambas variables están dentro de la escala de medición nominal.

Tipo de investigación: Básica, cuantitativa, descriptiva, transversal y no experimental.

Resultados

Tabla de distribución de frecuencias

Tabla 9.4 Tabla cruzada sobre el consumo de café y la jornada de trabajo.

		Consumo de café		
		Frecuente	Poco frecuente	Total
Jornada	Tiempo parcial	18	38	56
	Tiempo completo	26	10	36
Total		44	48	92

Interpretación: En la tabla 9.4 se puede observar que en el grupo de 56 docentes a tiempo parcial 18 docentes consumen café de manera frecuente y 38 docentes presentan un consumo poco frecuente de café. En el grupo de los 36 docentes a tiempo completo se observa que 26 presentan un consumo frecuente de café y solo 10 presentan un consumo poco frecuente de café.

También se puede interpretar los resultados haciendo una comparación entre las frecuencias obtenidas. Por ejemplo, en la tabla 9.4 se puede observar que en el grupo de docentes a tiempo parcial los docentes que tienen un consumo poco frecuente (18) sobrepasan al doble de la cantidad de docentes que presentan un consumo frecuente de café (38). De manera contraria, en el grupo de docentes a tiempo completo la cantidad de docentes que tienen un consumo frecuente (26) supera notablemente a la cantidad de docentes que presentan un consumo poco frecuente de café (10).

Además, se pueden interpretar las frecuencias marginales. Por ejemplo, la cantidad de docentes a tiempo parcial es mucho mayor que la cantidad de docentes a tiempo completo, sin embargo, entre la cantidad de docentes que presentan un consumo frecuente y la cantidad de docentes que presentan un consumo poco frecuente no se observa mucha diferencia.

Existe la opción de utilizar tablas con frecuencias porcentuales:

Tabla 9.5 Tabla cruzada sobre el consumo de café y la jornada de trabajo.

		Consumo de café		
		Frecuente	Poco frecuente	Total
Jornada	Tiempo parcial	32.14	67.86	100
	Tiempo completo	72.22	27.78	100

Interpretación: En la tabla 9.5 se puede observar que en el grupo de docentes a tiempo parcial el 32.14% de los docentes presenta un consumo frecuente de café y el 67.86% de los docentes presenta un consumo poco frecuente de café. En el grupo de los docentes a tiempo completo se observa el 72.22% presenta un consumo frecuente de café y el 27.78% presenta un consumo poco frecuente de café.

Tabla 9.6 Tabla cruzada sobre el consumo de café y la jornada de trabajo.

		Consumo de café	
		Frecuente	Poco frecuente
Jornada	Tiempo parcial	40.91	79.17
	Tiempo completo	59.09	20.83
Total		100	100

Interpretación: En la tabla 9.6 se puede observar que en el grupo de docentes que presentan un consumo frecuente de café el 40.91% se compone de docentes a tiempo parcial y el 59.09% se compone de docentes a tiempo completo. En el grupo de docentes que presentan un consumo poco frecuente de café el 79.17% se compone de docentes a tiempo parcial y el 20.83% se compone de docentes a tiempo completo.

Las dos últimas tablas se pueden agrupar en una sola, sin embargo, su presentación y lectura se hace un poco más compleja. Se debe resaltar que no todas las tablas mostradas se deben incluir en la presentación de resultados, es el investigador quien debe seleccionar la tabla que brinde el mejor soporte a la descripción de los resultados.

Gráficos estadísticos

Como opción alterna a la presentación de tablas, se puede utilizar un diagrama de barras agrupadas para hacer la comparación entre el consumo de café entre los docentes a tiempo parcial y los docentes a tiempo completo (figura 9.1). Las interpretaciones son iguales a las de la tabla de frecuencias.

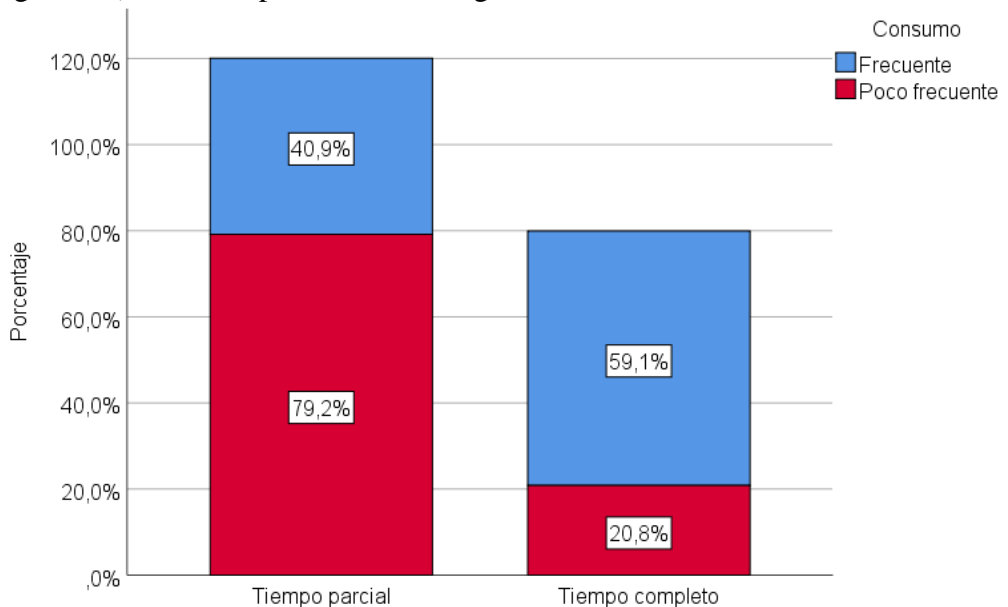


Figura 9.1 Consumo de café según la jornada de trabajo.

También se puede utilizar un diagrama de barras (figura 9.2) donde se presentan las dos variables por separado para hacer la comparación entre el consumo de café entre los docentes a tiempo parcial y los docentes a tiempo completo.

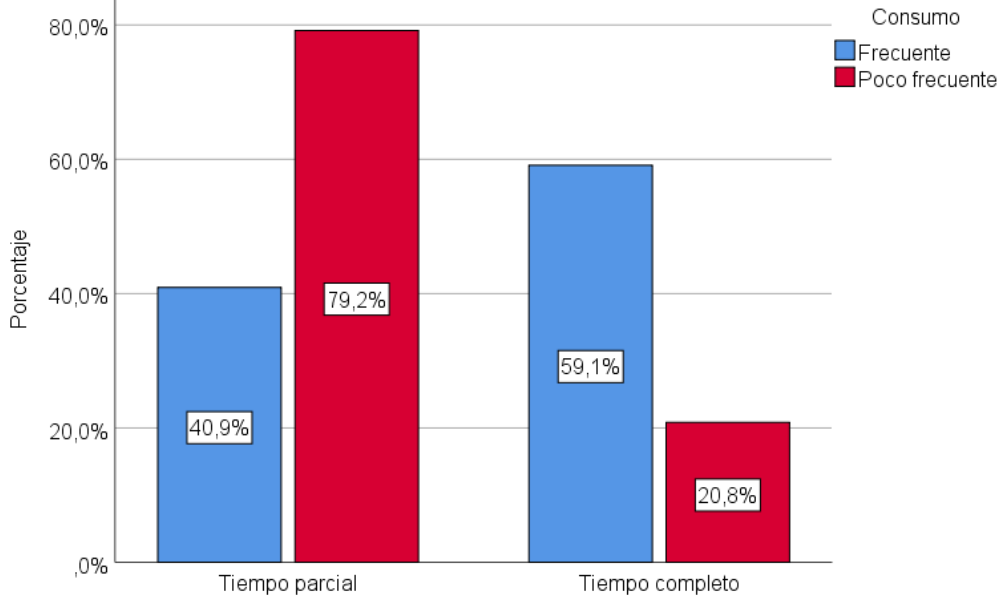


Figura 9.2 Diagrama de barras referente al consumo de café y la jornada de trabajo.

Puesto que la variable está dentro de la escala de medición nominal, se puede utilizar el diagrama circular para hacer una comparación entre el consumo de café entre los docentes.

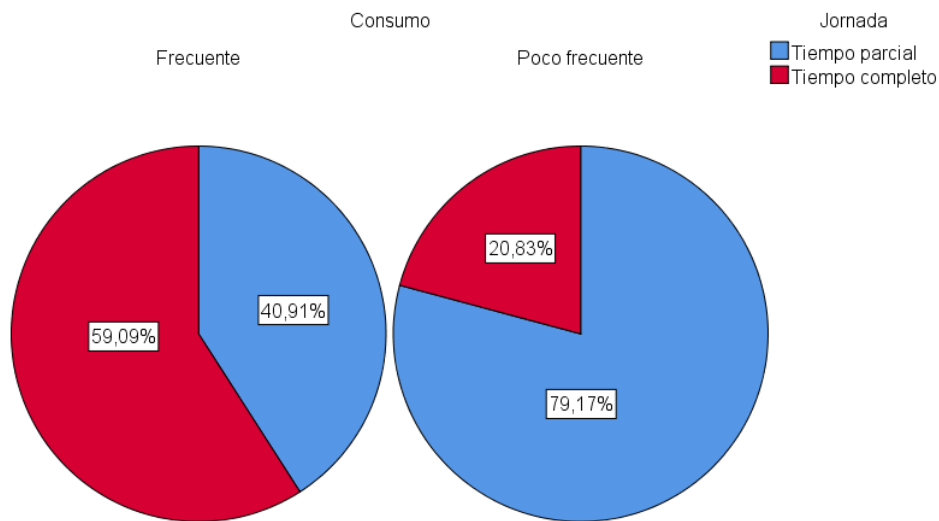


Figura 9.3 Diagrama circular referente al consumo de café y la jornada de trabajo.

La lectura de la figura 9.3 se puede iniciar observando el nivel de consumo, luego los sectores circulares y el tipo de jornada laboral que representa. Por ejemplo, se puede decir que dentro del grupo de docentes que presentan un consumo frecuente de café la mayor parte (59.09%) se compone de docentes a tiempo completo.

Otra forma de lectura podría ser, por ejemplo, que solo una pequeña parte (20.83%) de los docentes que presentan un consumo poco frecuente son docentes a tiempo parcial.

Aquí concluye el análisis descriptivo de los resultados. Posteriormente, se procede a seleccionar la prueba que permita determinar la asociación entre las variables en estudio.

9.3.2. Análisis descriptivo para variable ordinal

Ejemplo No. 1: Un grupo de estudiantes elabora y comercializa un nuevo tipo de galleta reforzada con maca. Luego de un tiempo se desea conocer la opinión de los consumidores con respecto al producto. Por ello, se encuesta a consumidores con edades entre los 18 y 22 años con las preguntas:

- ❖ ¿Qué te parece el sabor de las galletas?
- ❖ ¿Qué te parece el tamaño de las galletas?
- ❖ ¿Qué te parece el valor nutritivo de la galleta?
- ❖ La presentación del empaque de galletas ¿te parece atractiva?
- ❖ ¿Qué te parece el precio del paquete de galletas?

Los resultados se resumen en el siguiente cuadro:

Tabla 9.7 Datos referentes a la opinión de los consumidores.

Sujeto	Sabor	Tamaño	Nutrientes	Presentación	Precio
1	Bueno	Adecuado	Regular	Muy buena	Adecuado
2	Bueno	Adecuado	Regular	Muy buena	Adecuado
3	Muy bueno	Pequeño	Alto	Muy buena	Muy alto
4	Bueno	Adecuado	Regular	Muy buena	Muy alto
5	Regular	Pequeño	Regular	Buena	Adecuado
6	Bueno	Muy pequeño	Alto	Regular	Muy alto
7	Muy bueno	Pequeño	Alto	Muy buena	Muy alto
8	Muy bueno	Muy pequeño	Regular	Buena	Muy Barato
9	Muy bueno	Adecuado	Alto	Muy buena	Muy alto
10	Muy bueno	Pequeño	Regular	Buena	Muy alto
11	Malo	Adecuado	Regular	Muy buena	Adecuado
12	Bueno	Muy pequeño	Alto	Regular	Muy alto
13	Muy bueno	Pequeño	Regular	Buena	Muy Barato
14	Regular	Adecuado	Alto	Muy buena	Adecuado
15	Regular	Adecuado	Muy alto	Muy buena	Muy Barato
16	Muy bueno	Pequeño	Alto	Regular	Muy alto
17	Bueno	Pequeño	Muy alto	Buena	Adecuado
18	Bueno	Adecuado	Alto	Muy buena	Muy Barato
19	Bueno	Muy pequeño	Regular	Regular	Adecuado
20	Muy malo	Muy pequeño	Muy alto	Muy buena	Adecuado
21	Bueno	Adecuado	Regular	Buena	Muy alto
22	Bueno	Pequeño	Muy alto	Muy buena	Muy alto
23	Bueno	Pequeño	Muy alto	Muy buena	Adecuado
24	Regular	Pequeño	Muy alto	Buena	Adecuado
25	Regular	Pequeño	Regular	Buena	Muy Barato

Consideraciones metodológicas

Problema de investigación: ¿Cuál es la opinión de los consumidores con respecto a la galleta reforzada con maca?

Objetivo de la investigación: Conocer la opinión de los consumidores con respecto a la galleta reforzada con maca.

Hipótesis de la investigación: No se ha planteado hipótesis.

Variable: Opinión de los consumidores.

Escala de medición: Ordinal.

Tipo de investigación: Básica, cuantitativa, descriptiva, transversal y no experimental.

Resultados

Tablas de distribución de frecuencias

Tabla 9.8 Opinión de los encuestados con respecto al sabor de la galleta.

Sabor	Frecuencia	Porcentaje
Muy Bueno	7	28,0
Bueno	11	44,0
Regular	5	20,0
Malo	1	4,0
Muy malo	1	4,0
Total	25	100,0

Interpretación: En la tabla 9.8 se observa que, en cuanto al sabor, la galleta tiene gran aceptación, puesto que, el 44.0% de los consumidores encuestados opina que el sabor de la nueva galleta es bueno y el 28.0% opina que el sabor es muy bueno. Solo dos personas afirmaron que el sabor de la galleta les parece malo o muy malo.

Tabla 9.9 Opinión de los encuestados con respecto al tamaño de la galleta.

Tamaño	Frecuencia	Porcentaje
Adecuado	9	36,0
Pequeño	11	44,0
Muy pequeño	5	20,0
Total	25	100,0

Interpretación: En la tabla 9.9 se observa que, con respecto al tamaño, la galleta no satisface las expectativas de los consumidores, puesto que, solo el 36.0% de los consumidores encuestados opina que el tamaño de la nueva galleta es adecuado. En cambio, el 44.0% opina que la galleta es pequeña y el 20% afirma que la galleta es muy pequeña.

Tabla 9.10 Opinión de los encuestados con respecto al valor nutricional de la galleta.

Nutrientes	Frecuencia	Porcentaje
Muy alto	6	24,0
Alto	8	32,0
Regular	11	44,0
Total	25	100,0

Interpretación: En la tabla 9.10 se observa que, con respecto al valor nutritivo, gran parte de los consumidores consideran que la galleta no tiene un valor nutritivo aceptable, puesto que, el 44% opina que el valor nutritivo de la galleta no es alto o muy alto. Sin embargo, 24% de los encuestados opinan que el valor nutritivo es muy alto y el 32% opina que el valor nutritivo es alto.

Tabla 9.11 Opinión de los encuestados con respecto a la presentación de la galleta.

Presentación	Frecuencia	Porcentaje
Muy Buena	13	52,0
Buena	8	32,0
Regular	4	16,0
Total	25	100,0

Interpretación: En la tabla 9.11 se observa que, con respecto a la presentación de la galleta, el 52% de los consumidores considera que el empaque de la galleta es muy bueno y solo el 16% considera que el empaque de la galleta tiene una presentación regular.

Tabla 9.12 Opinión de los encuestados con respecto al precio de la galleta.

Precio	Frecuencia	Porcentaje
Muy alto	10	40,0
Adecuado	10	40,0
Muy Barato	5	20,0
Total	25	100,0

Interpretación: En la tabla 9.12 se verifica que, con respecto al precio de la galleta, el 40% de los consumidores considera que es adecuado, otro 40% considera que es muy alto, en contraste con el 20% que considera que el precio es muy bajo.

Es necesario resaltar que estas tablas, elaboradas para analizar los resultados de cada pregunta, no se pueden resumir en una sola tabla general (tabla 9.13), debido a que las categorías para cada pregunta están muy diferenciadas. Si se hubiesen utilizado las mismas categorías para cada pregunta, la información se podía resumir en una sola tabla como se muestra a continuación:

Tabla 9.13 Opinión de los encuestados con respecto al nuevo tipo de galleta.

Sabor	Tamaño	Nutrientes	Presentación	Precio
Frecuencias				
Muy Bueno				
Bueno				
Regular				
Malo				
Muy malo				
Total				

Gráficos estadísticos

Como opción alterna al uso de tablas se puede utilizar el diagrama de barras verticales (figura 9.4). La interpretación de los gráficos es similar a la interpretación de las tablas.

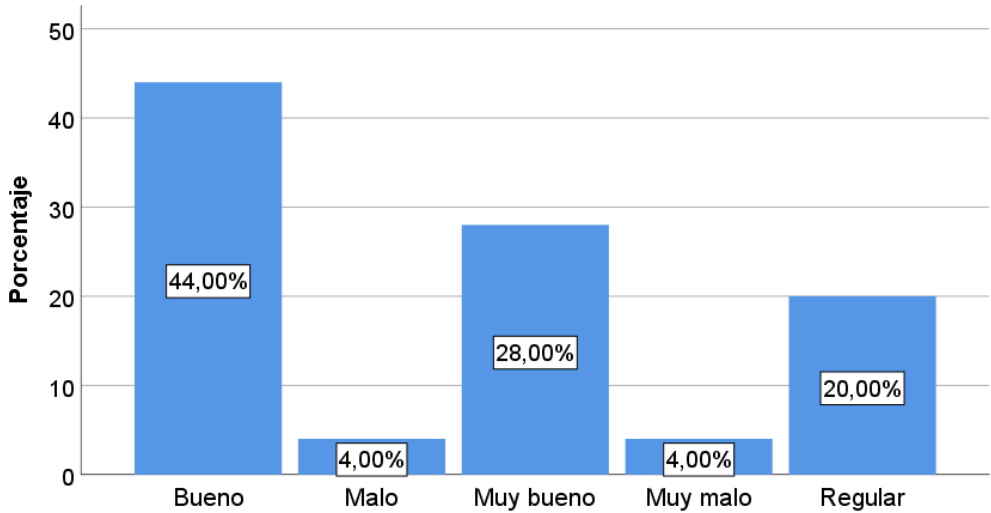


Figura 9.4 Opinión de los encuestados con respecto al sabor de la galleta.

Aquí concluye el análisis descriptivo de los resultados. No es necesario el uso de la estadística inferencial porque no se ha planteado ninguna hipótesis de investigación.

Ejemplo No. 2: Un estudiante de la carrera de ingeniería ambiental realiza una investigación para conocer el nivel de conocimientos sobre contaminación en los estudiantes del primer ciclo de su facultad. Para ello, diseña un instrumento de medición compuesto de diferentes preguntas cuyas respuestas generan puntajes que finalmente se acumularán para obtener el puntaje total. El puntaje total obtenido por cada integrante de la muestra se interpreta de la siguiente manera:

Tabla 9.14 Asignación de categorías de acuerdo al puntaje.

Puntaje total	Nivel de conocimiento
0-10	Bajo (Ba)
11-22	Regular (R)
23-34	Bueno (Bu)
35-40	Óptimo (O)

Se encuesta a 60 estudiantes de sexo masculino (M) y femenino (F), y los resultados son los siguientes:

Tabla 9.15 Datos referentes al sexo y al nivel de conocimiento.

No.	Sexo	Nivel	No.	Sexo	Nivel	No.	Sexo	Nivel	No.	Sexo	Nivel
1	M	O	16	M	R	31	M	Bu	46	F	R
2	F	R	17	M	O	32	M	R	47	F	R
3	F	Bu	18	F	Bu	33	F	Bu	48	F	R
4	F	Bu	19	M	Bu	34	M	R	49	F	R
5	M	Ba	20	F	R	35	F	Bu	50	M	Bu
6	F	Bu	21	F	Ba	36	M	Ba	51	M	R
7	F	Bu	22	F	Bu	37	F	O	52	F	O
8	F	Ba	23	M	Ba	38	M	Ba	53	F	Bu
9	M	O	24	F	Bu	39	F	R	54	F	Bu
10	M	Ba	25	M	R	40	F	O	55	F	Ba
11	M	R	26	M	Ba	41	M	Ba	56	M	Bu
12	M	R	27	M	Ba	42	F	R	57	M	R
13	M	R	28	F	Bu	43	F	Bu	58	M	Ba
14	M	Bu	29	F	R	44	M	R	59	M	R
15	M	Ba	30	F	Bu	45	F	R	60	M	R

Consideraciones metodológicas

Problema de investigación: ¿Cuál es el nivel de conocimientos sobre contaminación en los estudiantes del primer ciclo de la facultad de ingeniería?

Objetivo de la investigación: Conocer el nivel de conocimientos sobre contaminación en los estudiantes del primer ciclo de la facultad de ingeniería.

Hipótesis de la investigación: Los estudiantes del primer ciclo de la facultad de ingeniería presentan un óptimo nivel de conocimientos sobre la contaminación.

Variable: Nivel de conocimientos sobre contaminación.

Escala de medición: Ordinal.

Tipo de investigación: Básica, cuantitativa, descriptiva, transversal y no experimental.

Resultados

La escala de medición es ordinal, por lo tanto, se puede elaborar una tabla de distribución de frecuencias utilizando a frecuencia relativa, relativa porcentual y la relativa porcentual acumulada.

Tabla 9.16 Nivel de conocimientos sobre contaminación.

	Frecuencia	Porcentaje	Porcentaje acumulado
Bajo	13	21,7	21,7
Regular	22	36,7	58,3
Bueno	19	31,7	90,0
Óptimo	6	10,0	100,0
Total	60	100,0	

Interpretaciones:

En la tabla 9.16 se puede observar que solo el 10% de los estudiantes encuestados presenta un nivel óptimo de conocimiento sobre contaminación, el 31.7% presenta un buen nivel, el 36% de los estudiantes presenta un nivel regular y el 21.7% presenta un bajo nivel de conocimientos sobre contaminación.

Podemos interpretar la moda indicando que el nivel de conocimientos más frecuente entre los estudiantes es el nivel regular.

También se puede decir que más de la mitad (58%) de los estudiantes presenta un bajo o regular nivel de conocimientos sobre contaminación.

Gráfico estadístico

Otra opción para presentar los resultados es el uso del diagrama de barras verticales con la frecuencia porcentual o la frecuencia absoluta. En la figura 9.5 se puede apreciar que se ha utilizado la frecuencia porcentual.

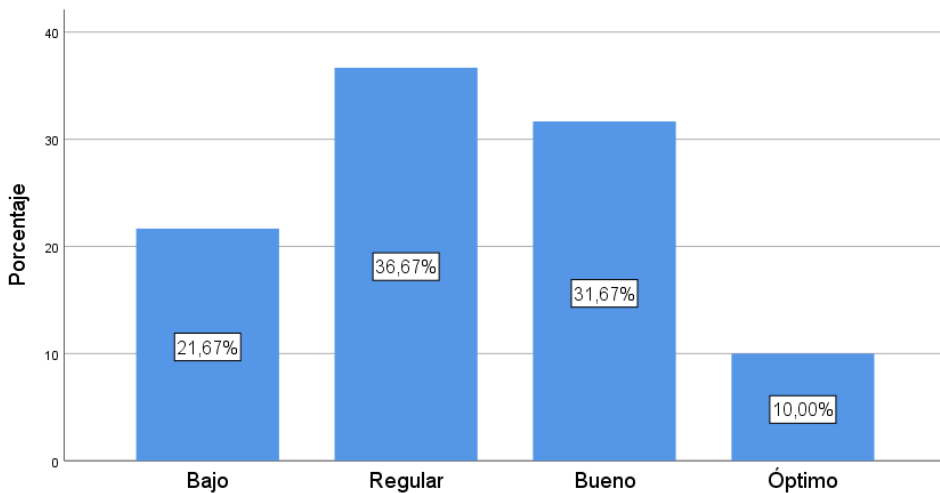


Figura 9.5 Nivel de conocimientos sobre contaminación.

De acuerdo con los objetivos se puede utilizar o prescindir de una tabla cruzada que permita comparar el nivel de conocimientos teniendo en cuenta el sexo del estudiante. En este caso, el sexo es información que también se ha recolectado así que se puede utilizar para hacer la comparación.

Tabla 9.17 Nivel de conocimientos sobre contaminación según el sexo.

		Sexo		Total
		Femenino	Masculino	
Nivel de conocimiento	Bajo	3	10	13
	Regular	10	12	22
	Bueno	14	5	19
	Óptimo	3	3	6
Total		30	30	60

Interpretación: En la tabla 9.17 se puede observar que la muestra se compone de un 50% de mujeres y un 50% de varones. La cantidad de mujeres que presentan un nivel bajo de conocimientos es mucho menor que la cantidad de varones en este nivel. En el nivel regular la diferencia entre las cantidades no es muy significativa en cambio, en el nivel bueno la cantidad de mujeres es casi el triple que la cantidad de varones. En el nivel óptimo se observa la misma cantidad de varones y mujeres.

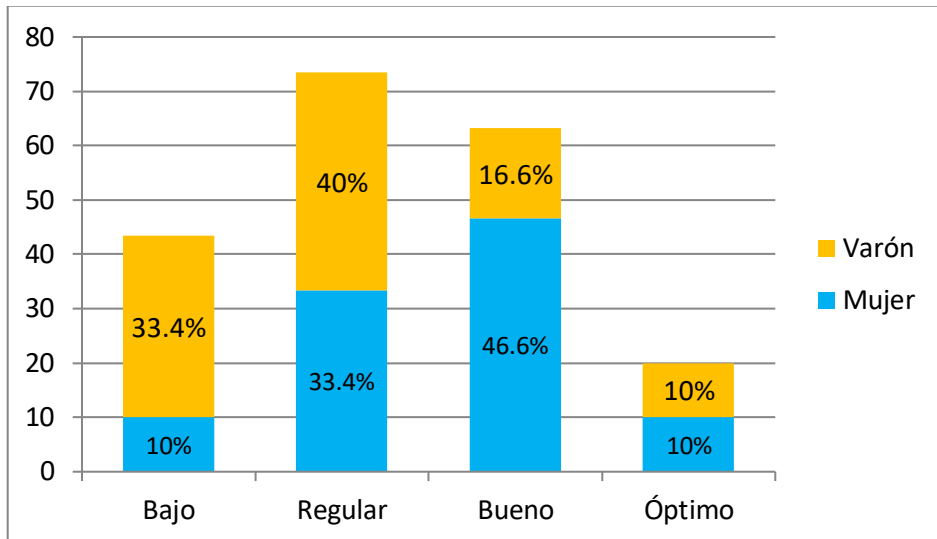
Gráfico estadístico

Figura 9.6 Nivel de conocimientos sobre contaminación según el sexo.

En la figura 9.6 se puede visualizar mejor el tanto por ciento de estudiantes en cada nivel, además se ve claramente la proporción de varones y mujeres en cada uno de estos niveles. Por ejemplo, se puede evidenciar que entre los estudiantes que tienen un buen nivel de conocimiento, las mujeres representan la mayor proporción en comparación con los varones.

Aquí concluye el análisis descriptivo de los resultados. Posteriormente, se debe utilizar la estimación de la proporción poblacional para contrastar la hipótesis de investigación.

9.4. Análisis descriptivo para variables cuantitativas

9.4.1. Análisis descriptivo para variable discreta

Ejemplo No. 1: Los administradores de la red de investigadores “S” desean conocer a cuántas organizaciones académicas están asociados sus miembros, teniendo en consideración la universidad de la que proceden dichos investigadores: universidad pública (Pu) o universidad privada (P). Por ello se realiza una encuesta y se obtienen los siguientes resultados:

Tabla 9.18 Datos resultantes de la encuesta.

No.	Sexo	Organizaciones	No.	Sexo	Organizaciones	No.	Sexo	Organizaciones	No.	Sexo	Organizaciones
1	Pu	8	20	Pu	12	39	P	14	58	P	9
2	Pu	12	21	Pu	12	40	P	12	59	P	10
3	Pu	15	22	Pu	13	41	P	8	60	P	11
4	Pu	8	23	Pu	12	42	P	10	61	P	16
5	Pu	7	24	Pu	10	43	P	8	62	P	11
6	Pu	12	25	Pu	13	44	P	14	63	P	11
7	Pu	10	26	Pu	10	45	P	13	64	P	14
8	Pu	10	27	Pu	11	46	P	13	65	P	12
9	Pu	11	28	Pu	11	47	P	14	66	P	14
10	Pu	10	29	Pu	12	48	P	7	67	P	15
11	Pu	15	30	Pu	7	49	P	8	68	P	12
12	Pu	12	31	P	7	50	P	9	69	P	12
13	Pu	12	32	P	12	51	P	9	70	P	16
14	Pu	10	33	P	7	52	P	10	71	P	13
15	Pu	12	34	P	10	53	P	10	72	P	15
16	Pu	10	35	P	9	54	P	9	73	P	15
17	Pu	10	36	P	10	55	P	8	74	P	11
18	Pu	14	37	P	11	56	P	9	75	P	11
19	Pu	16	38	P	14	57	P	9	76	P	12

Consideraciones metodológicas

Problema de investigación: ¿A cuántas organizaciones académicas están asociados los miembros de la red de investigadores “S”?

Problema específico 1: ¿A cuántas organizaciones académicas están asociados los miembros de la red de investigadores “S” que provienen de universidades públicas?

Problema específico 2: ¿A cuántas organizaciones académicas están asociados los miembros de la red de investigadores “S” que provienen de universidades privadas?

Objetivo de la investigación: Conocer a cuántas organizaciones académicas están asociados los miembros de la red de investigadores “S”.

Objetivo específico 1: Conocer a cuántas organizaciones académicas están asociados los miembros de la red de investigadores “S” que provienen de universidades públicas.

Problema específico 2: Conocer a cuántas organizaciones académicas están asociados los miembros de la red de investigadores “S” que provienen de universidades privadas.

Hipótesis de la investigación: No se ha planteado hipótesis.

Variable: Cantidad de organizaciones.

Escala de medición: Razón.

Tipo de investigación: Básica, cuantitativa, descriptiva, transversal y no experimental.

Resultados

Tablas de distribución de frecuencias

Tabla 9.19 Cantidad de organizaciones a las que están asociados los investigadores.

Organizaciones	Frecuencia	Porcentaje	Porcentaje acumulado
7,00	5	6,6	6,6
8,00	6	7,9	14,5
9,00	7	9,2	23,7
10,00	14	18,4	42,1
11,00	9	11,8	53,9
12,00	15	19,7	73,7
13,00	5	6,6	80,3
14,00	7	9,2	89,5
15,00	5	6,6	96,1
16,00	3	3,9	100,0
Total	76	100,0	

Interpretaciones:

En la tabla 9.19 se observa que las mayores frecuencias corresponden a los investigadores que están asociados a 10, 11 o 12 organizaciones académicas y estas frecuencias en conjunto representan el 50% de los investigadores que son parte del análisis.

El 23.7% de los investigadores está asociado, como máximo, a 9 organizaciones académicas y el 26.3% está asociado, como mínimo a 13 organizaciones académicas. Es decir, las menores frecuencias corresponden a los investigadores asociados a menos de 10 o más de 12 organizaciones académicas.

También se puede resaltar que solo el 6.6% de los investigadores está asociado a 7 organizaciones académicas y 3.9% está asociado a 16 organizaciones académicas. De otro lado, el 80.3% está asociado, como máximo a 13 organizaciones académicas.

Gráficos estadísticos

Los gráficos son una opción alterna para presentar los resultados y su interpretación es similar a la interpretación de la tabla. En la figura 9.7 se observa un diagrama de barras como grafico adecuado ya que no hay una gran variabilidad en los valores analizados y esto hace innecesario es uso de intervalos de clase.

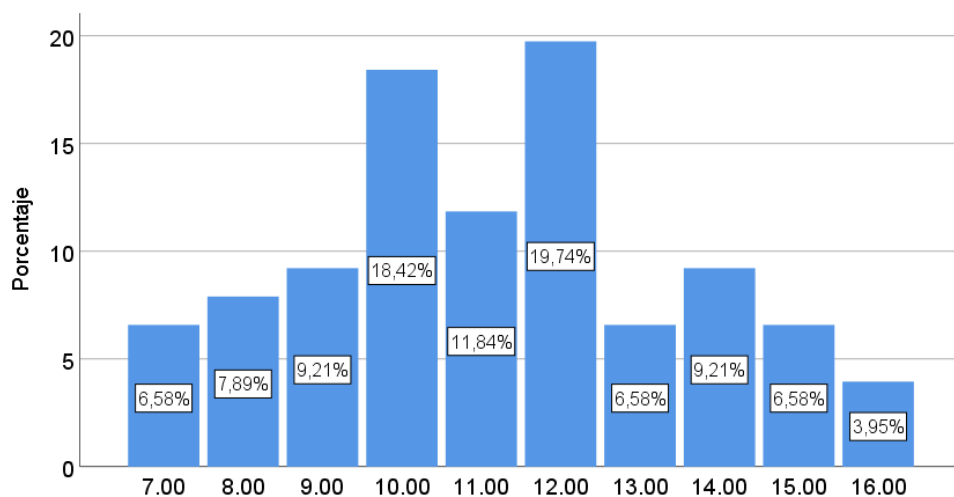


Figura 9.7 Cantidad de organizaciones a las que están asociados los investigadores.

Estadígrafos

Puesto que la variable es cuantitativa, se pueden calcular e interpretar las medidas de posición central, posición no central, de dispersión y de forma.

Tabla 9.20 Cantidad de organizaciones a las que están asociados los investigadores.

Estadístico	Valor
N	76
Media	11,1974
Mediana	11,0000

Moda		12,00
Desviación estándar		2,38339
Asimetría		,136
Curtosis		-,660
Mínimo		7,00
Máximo		16,00
Percentiles	25	10,0000
	50	11,0000
	75	13,0000

Interpretaciones:

En la tabla 9.20 se observa que la cantidad de organizaciones académicas a las que están asociados los 76 investigadores varía de 7 a 16 organizaciones.

En promedio, los investigadores están asociados a 11 organizaciones académicas aproximadamente. El 50% de investigadores está asociado como mínimo a 11 organizaciones académicas. Con mayor frecuencia se observa que los investigadores están asociados a 12 organizaciones académicas.

La desviación estándar se puede comparar con la media permitiendo calcular un coeficiente de variación de 21.28% lo que implica que la distribución es relativamente homogénea.

El signo del coeficiente de asimetría 0.136 indica que la gráfica distribución presenta un alargamiento hacia la derecha. El coeficiente de curtosis negativo indica que la distribución es platicúrtica. Es decir, la distribución presenta un achatamiento en la parte central en comparación con la distribución normal.

El 25% de los investigadores está asociado como máximo a 10 organizaciones académicas. El 50% de los investigadores está asociado como máximo a 11 organizaciones académicas. El 25% de los investigadores está asociado como mínimo a 13 organizaciones académicas

Comparación según la universidad de procedencia

Tabla 9.21 Cantidad de organizaciones a las que están asociados los investigadores según la universidad de procedencia.

Organizaciones	Pública	Privada	Total
7,00	2	3	5
8,00	2	4	6
9,00	0	7	7
10,00	8	6	14
11,00	3	6	9
12,00	9	6	15
13,00	2	3	5
14,00	1	6	7
15,00	2	3	5
16,00	1	2	3
Total	30	46	76

Interpretaciones:

En la tabla 9.21 se observa que la muestra se compone de una menor cantidad de investigadores que provienen de universidades públicas. Esta cantidad representa casi los dos tercios de la cantidad de investigadores que provienen de universidades privadas.

Observando las mayores frecuencias totales se puede ver que de los 14 investigadores que están asociados a 10 organizaciones académicas 8 investigadores provienen de universidades públicas y 6 provienen de universidades privadas.

Además, de los 15 investigadores que están asociados a 12 organizaciones académicas 9 investigadores provienen de universidades públicas y 6 provienen de universidades privadas.

En contraste, de los 9 investigadores que están asociados a 11 organizaciones académicas 3 investigadores provienen de universidades públicas y 6 provienen de universidades privadas.

Comparación gráfica según la universidad de procedencia

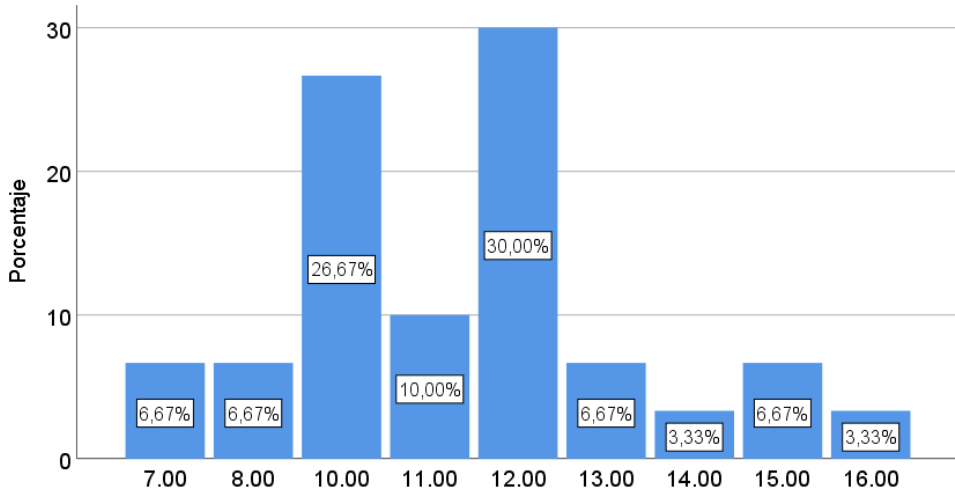


Figura 9.8 Cantidad de organizaciones a las que están asociados los investigadores que provienen de universidades públicas.

Interpretaciones

En el caso de los investigadores que provienen de universidades públicas, en la figura 9.8 se puede apreciar que las mayores frecuencias corresponden a los investigadores que están asociados a 10, 11 o 12 organizaciones académicas.

La mayor frecuencia corresponde a los investigadores es que están asociados a 12 organizaciones académicas y representa el 30%. La segunda mayor frecuencia corresponde a los investigadores que están asociados a 10 organizaciones académicas y representa el 26.67%.

El 20% de este grupo de investigadores está asociado como mínimo a 13 organizaciones académicas y el 13.34% está asociado a 7 u 8 de este tipo de organizaciones.

Solo 1 investigador de este grupo está asociado a 16 organizaciones académicas.

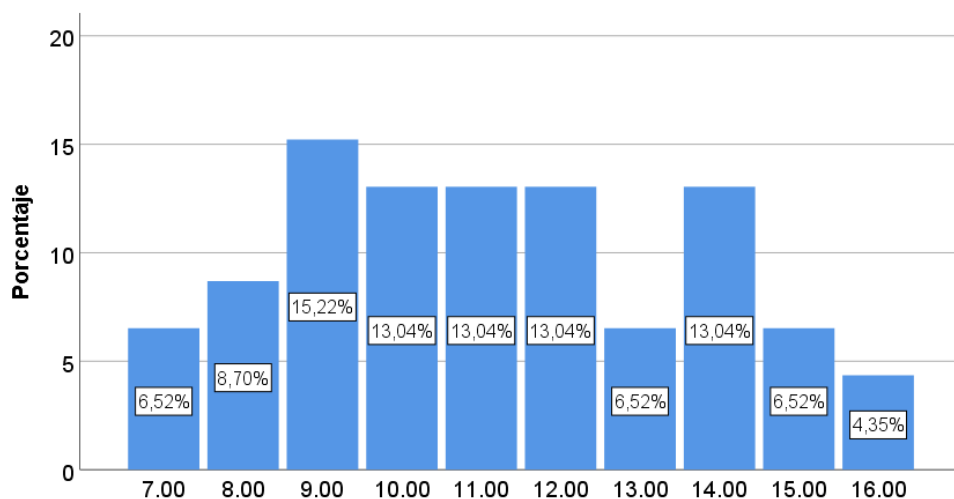


Figura 9.9 Cantidad de organizaciones a las que están asociados los investigadores que provienen de universidades privadas.

Interpretaciones

En la figura 9.9 se observa que en el grupo de investigadores que provienen de universidades privadas la distribución de frecuencias es relativamente más heterogénea con respecto a los investigadores que provienen de universidades públicas, dado que, en este último grupo solo dos frecuencias son las que presentan mayor diferencia con respecto a las demás.

La mayor frecuencia corresponde a los investigadores es que están asociados a 9 organizaciones académicas y representa el 15.22%. La segunda mayor frecuencia corresponde a los investigadores que están asociados a 10, 11, 12 y 14 organizaciones académicas y representa el 13.04 en cada caso.

El 30.43% de este grupo de investigadores está asociado como mínimo a 13 organizaciones académicas y el 15.22% está asociado a 7 u 8 de este tipo de organizaciones.

Solo 2 investigadores de este grupo están asociados a 16 organizaciones académicas.

Tabla 9.22 Cantidad de organizaciones según la universidad de procedencia.

Estadístico	Valor	
	Pública	privada
N	30	46
Media	11,2333	11,1739
Mediana	11,5000	11,0000
Moda	12,00	9,00
Desviación estándar	2,19220	2,52370
Asimetría	,060	,177
Curtosis	,115	-,938
Mínimo	7,00	7,00
Máximo	16,00	16,00
Percentiles	25	10,0000
	50	11,5000
	75	12,0000

Interpretaciones:

En la tabla 9.22 se observa que, en promedio, los investigadores que provienen de universidades públicas están asociados a 11.23 organizaciones académicas. Esta cantidad supera al 11.17 que representa el promedio de los investigadores que provienen de universidades privadas.

La mitad de los investigadores que provienen de universidades públicas está asociada por lo menos a 11.50 organizaciones en comparación con el grupo de los investigadores que provienen de universidades privadas donde la mitad de ellos está asociada como mínimo a 11.00 de este tipo organizaciones.

Los investigadores que provienen de universidades públicas están asociados con mayor frecuencia a 12 organizaciones académicas. En el grupo de investigadores que provienen de universidades privadas se observa que la cantidad de organizaciones más frecuente es 9.

El coeficiente de variación en el grupo de investigadores que provienen de universidades públicas es 19.52% y en el grupo de investigadores que provienen de universidades privadas es 22.56. Esto quiere decir que entre datos del grupo de investigadores que provienen de universidades privadas existe una mayor dispersión.

En la distribución de ambos grupos de investigadores se observa una asimetría positiva, esto indica que la distribución presenta un sesgo hacia la derecha y que hay una mayor concentración de datos en la primera mitad de la distribución.

En el grupo de investigadores que provienen de universidades públicas se observa que el coeficiente de curtosis es positivo, es decir, las mayores frecuencias se ubican en la parte central de la distribución. En el grupo de investigadores que provienen de universidades privadas se observa que el coeficiente de curtosis es negativo, es decir, la distribución es más plana que la distribución normal.

El 25% de los investigadores que provienen de universidades públicas está asociado como máximo a 10 organizaciones académicas y 25% del grupo de investigadores que provienen de universidades públicas está asociado como máximo a 9 organizaciones académicas.

El 75% de los investigadores que provienen de universidades públicas está asociado como máximo a 12 organizaciones académicas y 75% del grupo de investigadores que provienen de universidades públicas está asociado como máximo a 13.25 organizaciones académicas.

Diagrama de cajas

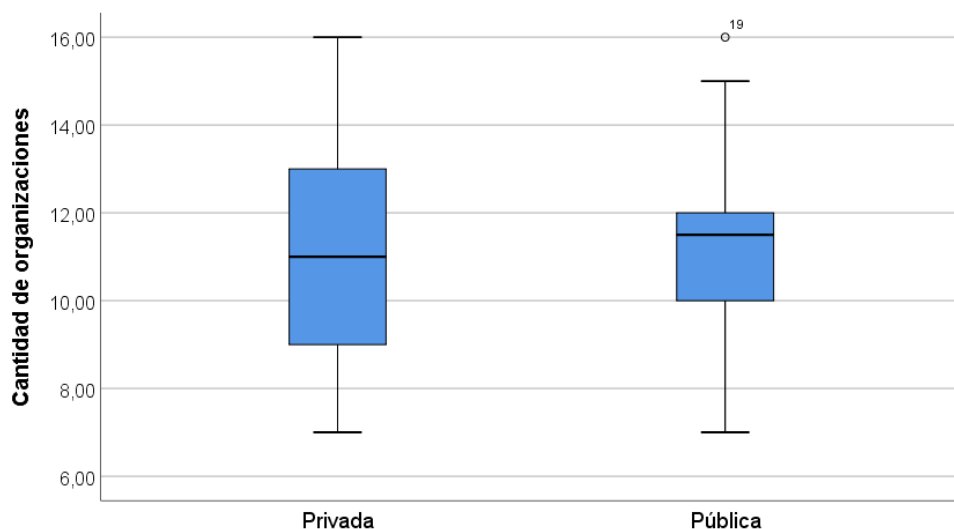


Figura 9.10 Comparación según la universidad de procedencia

El diagrama de cajas es un gráfico que se construye en base a los estadígrafos, por ello contiene más información que una tabla de distribución de frecuencias o un gráfico de barras. Al usar el diagrama de cajas se puede prescindir de la tabla 9.22 que contiene a los estadígrafos.

Interpretaciones:

Según la figura 9.10, los datos del grupo de investigadores que provienen de universidades públicas están menos dispersos en comparación con el grupo de investigadores que proviene de universidades privadas.

En el grupo de los investigadores que provienen de universidades privadas se observa que la mediana se ubica en la parte central de la caja lo que indica que la distribución es homogénea entre el primer y tercer cuartil. En cambio, en el grupo de los investigadores que provienen de universidades públicas la mediana está más cerca al tercer cuartil lo que indica que los datos se concentran en los valores altos.

También se observa que el grupo de los investigadores que provienen de universidades privadas no hay valores atípicos. En cambio, en el otro grupo sí se registra un valor atípico, este valor atípico es el 16, dato que ocupa el lugar 19 en la matriz de datos (tabla 9.18).

Aquí concluye el análisis descriptivo de los resultados. No es necesario el uso de la estadística inferencial porque no se ha planteado ninguna hipótesis de investigación.

9.4.2. Análisis descriptivo para variable continua

Ejemplo No. 1: En una investigación se desea conocer las características de la masa corporal de los empleados de la empresa “X”. Por ello, se utiliza una balanza digital y se realiza la medición en una muestra de 80 empleados de sexo masculino (M) y femenino (F).

Tabla 9.23 Datos referentes al sexo y a la masa corporal.

No	Sexo	Masa	No	Sexo	Masa	No	Sexo	Masa	No	Sexo	Masa
1	F	50.20	21	F	53.35	41	M	69.80	61	M	74.37
2	F	60.50	22	F	64.40	42	M	70.45	62	M	75.30
3	F	63.00	23	F	65.60	43	M	75.30	63	M	73.50
4	F	60.25	24	F	71.30	44	M	78.40	64	M	74.00
5	F	68.80	25	F	53.20	45	M	80.43	65	M	75.10
6	F	65.30	26	F	54.00	46	M	83.50	66	M	74.30
7	F	72.45	27	F	64.20	47	M	90.50	67	M	74.20
8	F	68.90	28	F	68.20	48	M	83.59	68	M	77.20
9	F	60.00	29	F	63.20	49	M	81.40	69	M	85.80
10	F	58.40	30	F	52.10	50	M	73.40	70	M	92.90
11	F	59.00	31	F	71.80	51	M	77.24	71	M	74.30
12	F	57.98	32	F	75.10	52	M	74.53	72	M	76.20
13	F	65.78	33	M	73.40	53	M	75.34	73	M	73.30
14	F	65.46	34	M	74.60	54	M	77.10	74	M	72.30
15	F	67.30	35	M	80.20	55	M	71.50	75	M	75.40
16	F	65.54	36	M	82.40	56	M	74.50	76	M	84.40
17	F	66.50	37	M	75.30	57	M	84.60	77	M	82.10
18	F	64.87	38	M	90.45	58	M	84.20	78	M	80.30
19	F	70.50	39	M	88.30	59	M	83.00	79	M	85.20
20	F	66.54	40	M	79.23	60	M	85.40	80	M	70.40

Consideraciones metodológicas

Problema de investigación: ¿Cuál son las características de la masa corporal de los empleados de la empresa “X”?

Problema específico 1: ¿Cuál son las características de la masa corporal de los varones empleados en la empresa “X”?

Problema específico 2: ¿Cuál son las características de la masa corporal de las mujeres empleadas en la empresa “X”?

Objetivo de la investigación: Conocer las características de la masa corporal de los empleados de la empresa “X”.

Objetivo específico 1: Conocer las características de la masa corporal de los varones empleados en la empresa “X”.

Objetivo específico 2: Conocer las características de la masa corporal de las mujeres empleadas en la empresa “X”.

Hipótesis de la investigación: No se ha planteado hipótesis.

Variable: Masa corporal.

Escala de medición: Razón.

Tipo de investigación: Básica, cuantitativa, descriptiva, transversal y no experimental.

Resultados

Tabla de frecuencias general

Ingresando los datos a algún programa de apoyo estadístico (SPSS, Minitab, R, etc.) se puede obtener la tabla de distribución de frecuencias. Puesto que la variable es continua y los datos presentan valores muy variados, se utiliza una tabla con intervalos de clase, frecuencia absoluta, frecuencia porcentual y frecuencia porcentual acumulada.

Tabla 9.24 Masa corporal de los empleados de la empresa “X”.

Masa corporal	Frecuencia	Porcentaje	Porcentaje acumulado
[50.20; 56.30>	5	6,3	6,3
[56.30; 62.40>	6	7,5	13,8
[62.40; 68.50>	14	17,5	31,3
[68.50; 74.60>	22	27,5	58,8
[74.60; 80.70>	17	21,3	80,0
[80.70; 86.80>	12	15,0	95,0
[86.80; 92.90]	4	5,0	100,0
Total	80	100,0	

Interpretaciones

De acuerdo con la tabla 9.24 solo el 5% de los empleados tiene una masa corporal de 86.80 kg a 92.90 kg.

Más de la cuarta parte de empleados (22) presenta una masa corporal mínima de 68.50 kg, pero menor a 74.60 kg.

El 58.8% de los empleados tiene una masa corporal menor a 74.46 kg.

El 80% de los empleados tiene una masa corporal menor a 80.70 kg.

Estadígrafos

Los programas de apoyo estadístico también permiten el cálculo automático de los estadígrafos o estadísticos de tendencia central, de dispersión, de posición o de forma. La tabla que se presenta a continuación generalmente se incluye en los anexos y en el cuerpo del análisis estadístico se redacta la interpretación de estos estadígrafos.

Tabla 9.25 Estadísticos sobre la masa corporal de los empleados.

Estadístico	Valor	
n	80	
Media	72,529	
Mediana	73,750	
Moda	72,25	
Desviación estándar	9,485	
Varianza	89,972	
Asimetría	-,239	
Curtosis	-,226	
Rango	42,70	
Mínimo	50,20	
Máximo	92,90	
Cuartiles	1	65,645
	3	79,023
	10	59,100
Percentiles	20	64,956
	80	81,206
	90	84,580

Interpretaciones

En la tabla 9.25 se observa que la muestra se compone de 80 observaciones en las que se ha registrado un mínimo y máximo valor de la masa corporal de 50.20 kg y 92.90 kg respectivamente estableciendo un rango muy significativo de 42.70 kg.

La masa corporal promedio entre los empleados es de 72.529 kg, la mitad de los empleados tiene una masa corporal mínima de 73.75 kg y la masa corporal más frecuente es de 72.25 kg.

La desviación estándar calculada es de 9,485 kg y comparando este valor con la media se obtiene un coeficiente de variación de 13.078 % el mismo que indica que la muestra es relativamente homogénea.

El coeficiente de asimetría calculado es de -0.239 lo que indica la gráfica de frecuencias tiene un sesgo hacia la izquierda, es decir, que entre los datos existe una tendencia a los valores más grandes. La curtosis es negativa, por lo tanto, la distribución tiene menor cantidad de valores atípicos extremos en comparación con la distribución normal.

El 25% de los empleados tiene una masa corporal máxima de 65,645 kg. El 75% de ellos tiene una masa corporal máxima de 79,023 kg. Si analizamos los valores de las observaciones más extremas tenemos que el 10% de los empleados presenta una masa corporal máxima de 59,100 kg y el 10 % de los empleados tiene una masa corporal mínima de 84,580 kg.

Gráficos estadísticos

El uso de los gráficos es otra opción para presentar la distribución de datos.

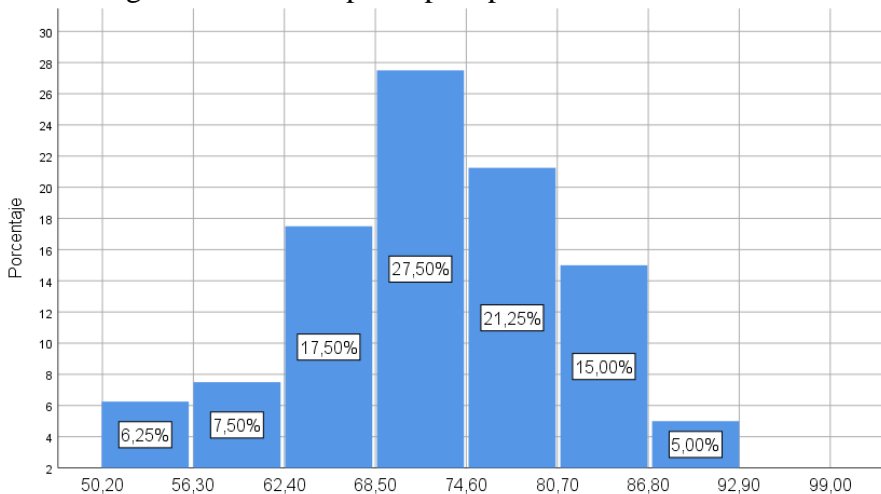


Figura 9.11 Masa corporal de los empleados de la empresa “X”.

Del histograma (figura 9.11) se pueden realizar interpretaciones similares a las que se elaboró a partir de las tablas.

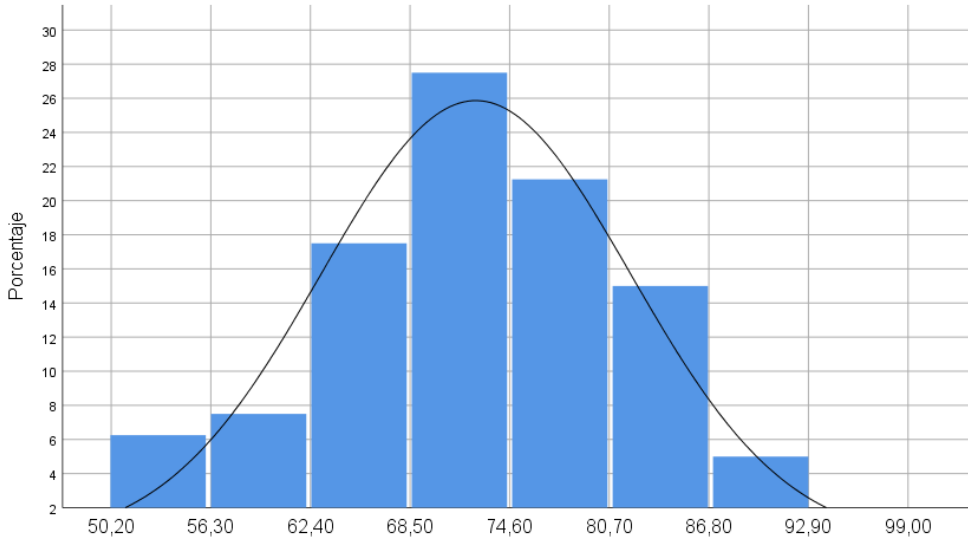


Figura 9.12 Masa corporal de los empleados de la empresa "X".

Si se coloca la distribución normal sobre el histograma (figura 9.12) se puede verificar la asimetría negativa observando el sesgo hacia la izquierda en el histograma. También se verifica la curtosis negativa, puesto que la mayor parte de frecuencias centrales están por debajo de la curva normal.

Masa corporal según el sexo

En la tabla de datos registrados al inicio, además de la masa corporal se registró el sexo de los empleados, esta información responde a los objetivos específicos del estudio. El análisis descriptivo se hace más específico cuando se realiza una comparación de las características de la masa corporal en las mujeres y en los varones por separado. Este análisis depende de los objetivos de la investigación y de los datos que se han recolectado, es decir, se además de haber registrado el sexo de los empleados se hubiese registrado también el estado civil correspondiente a cada empleado se podría

realizar un análisis de la masa corporal según el estado civil del empleado (soltero, casado, viudo, etc.).

Estadígrafos para la masa corporal en las mujeres

En la sección anterior ya se detalló una tabla de distribución de frecuencias, esa tabla también se puede elaborar para conocer la distribución de la masa corporal según el sexo. En esta parte solo interpretaremos los estadísticos de la variable masa corporal para el grupo de mujeres que integran la muestra.

Tabla 9.26 Estadísticos sobre la masa corporal.

Estadístico	Valor	
n	32	
Media	63,554	
Mediana	65,085	
Moda	64.725	
Desviación estándar	6,336	
Varianza	40,143	
Asimetría	-,445	
Curtosis	-,417	
Rango	24,90	
Mínimo	50,20	
Máximo	75,10	
Cuartiles	1	59,250
	3	67,975
Percentiles	10	53,245
	90	71,650

Interpretaciones

En la tabla 9.26 se observa que la muestra de mujeres se compone de 32 observaciones en las que se ha registrado un mínimo y máximo valor de la masa corporal de 50.20 kg y 75.10 kg respectivamente estableciendo un rango de 24.90 kg.

Entre las mujeres, la masa corporal promedio es de 63.554 kg, la mitad tiene una masa corporal mínima de 65.085 kg y la masa corporal más frecuente es de 64.725 kg.

La desviación estándar calculada es de 6,336 kg y comparando este valor con la media se obtiene un coeficiente de variación de 9.969 % que indica que la muestra es relativamente homogénea.

El coeficiente de asimetría calculado es -0.445 lo que indica la gráfica de frecuencias tiene un sesgo hacia la izquierda, es decir, que entre los datos existe una tendencia a los valores más grandes. La curtosis es negativa, por lo tanto, la distribución tiene menor cantidad de valores atípicos extremos en comparación con la distribución normal.

El 25% de las mujeres tiene una masa corporal máxima de 59.250 kg. El 75% de ellas tiene una masa corporal máxima de 67.975 kg. Si analizamos los valores de las observaciones más extremas tenemos que el 10% de las mujeres presenta una masa corporal máxima de 53.245 kg y otro 10 % tiene una masa corporal de por lo menos de 71.650 kg.

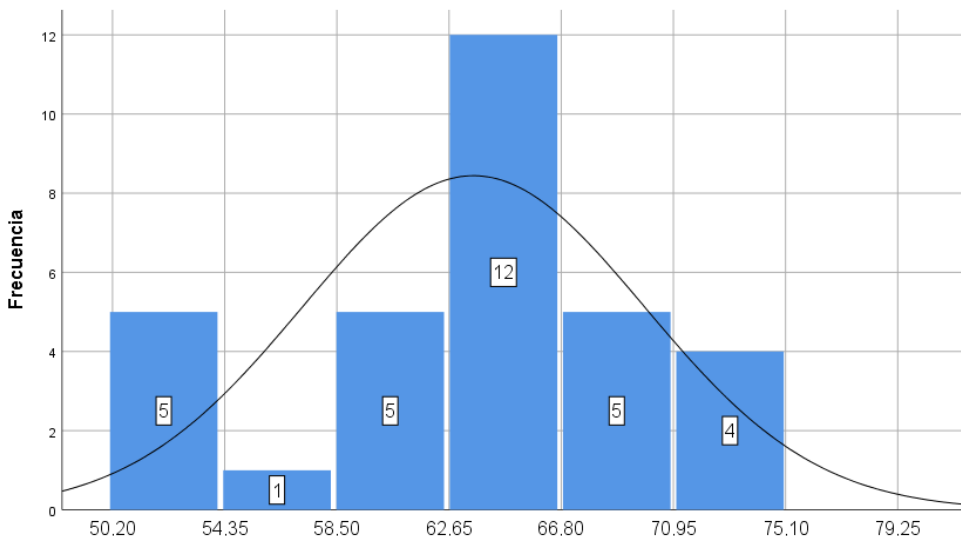


Figura 9.13 Masa corporal en las mujeres.

En la figura 9.13 se puede verificar la asimetría negativa. La curtosis es negativa, pues se observa que solo una columna central presenta una altura por encima de la curva normal.

Estadígrafos para la masa corporal en los varones

De la misma forma que en análisis descriptivo del grupo de mujeres, en esta parte solo interpretaremos los estadísticos del grupo de varones que integran la muestra inicial.

Tabla 9.27 Estadísticos sobre la masa corporal.

Estadístico	Valor	
N	48	
Media	78,513	
Mediana	76,650	
Moda	75,30	
Desviación estándar	5,774	
Varianza	33,334	
Asimetría	,671	
Curtosis	-,375	
Rango	23,10	
Mínimo	69,80	
Máximo	92,90	
Cuartiles	1	74,300
	3	83,375
Percentiles	10	72,220
	90	86,050

Interpretaciones

En la tabla 9.27 se puede observar que la muestra de varones se compone de 48 observaciones en las que se ha registrado un mínimo y máximo valor de la masa corporal de 69.80 kg y 92.90 kg respectivamente estableciendo un rango de 23.10 kg.

La masa corporal promedio entre los varones es de 78.513 kg, la mitad de los empleados tiene una masa corporal de por lo menos de 76.650 kg y la masa corporal más frecuente es de 75.30 kg.

La desviación estándar calculada es de 5.774 kg y comparando este valor con la media se obtiene un coeficiente de variación de 7.354% que indica que la muestra es relativamente homogénea.

El coeficiente de asimetría calculado es 0.671 lo que indica la gráfica de frecuencias tiene un sesgo hacia la derecha, es decir, que entre los datos existe una tendencia a los valores más pequeños. La curtosis es negativa, por lo tanto, las mayores frecuencias no se ubican en la parte central de la distribución.

El 75% de los varones tiene una masa corporal mínima de 74.30 kg. El 25% de ellos tiene una masa corporal mínima de 83.375 kg. Si analizamos los valores de las observaciones más extremas tenemos que el 10% de los varones presenta una masa corporal máxima de 72.22 kg y otro 10 % tiene una masa corporal de por lo menos de 86.050 kg.

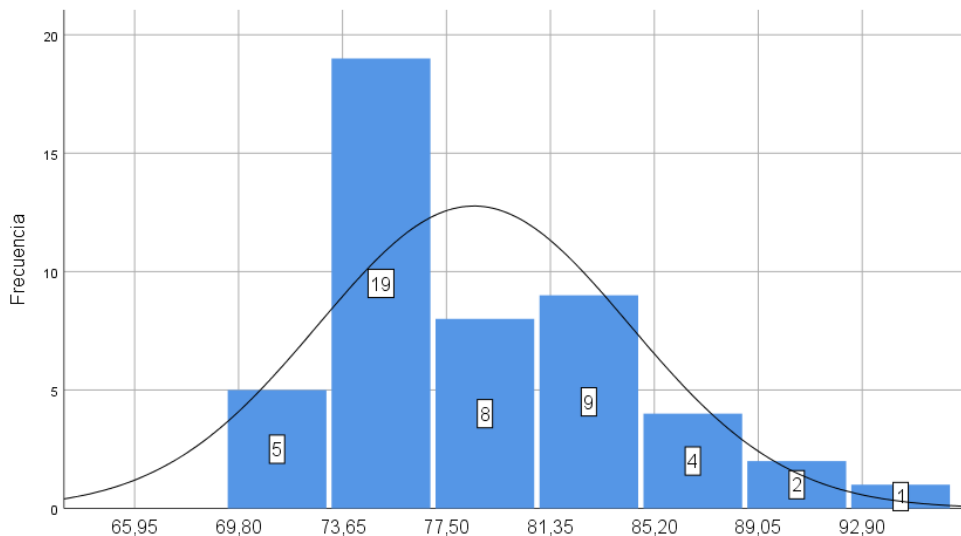


Figura 9.14 Masa corporal en los varones.

En la figura 9.14 se puede verificar la asimetría positiva, es decir, el alargamiento hacia el lado derecho y una mayor concentración de datos en los valores más bajos. El tipo de curtosis también se puede verificar observando que solo una columna central presenta una altura por encima de la curva normal.

Entre los datos no contamos con otro tipo de información como la estatura o la edad, sin embargo, en muchos estudios sobre la masa corporal se verifica que los varones presentan una mayor masa corporal en comparación con la masa corporal de las mujeres. Por ello, es válido orientar el análisis hacia una comparación simplificada sobre la media y la variabilidad de la masa corporal en ambos grupos.

Tabla 9.28 Estadísticos sobre la masa corporal por grupo.

Grupo	Media	Desviación estándar	Coefficiente de variación
Varones	78,513	5,774	7.354%
Mujeres	63. 554	6,336	9.969 %

Interpretaciones

En la tabla 9.28 se verifica que los varones presentan una mayor masa corporal promedio. También, se puede concluir que la masa corporal en el grupo de mujeres presenta una mayor dispersión en comparación con la masa corporal del grupo de varones.

También se puede utilizar el rango intercuartil para conocer la concentración del 50% de datos en la parte central de la distribución.

Tabla 9.29 Rango intercuartil de la masa corporal por grupo.

Grupo	Q_1	Q_3	R_{IQ}
Varones	74,300	83,375	9.075
Mujeres	59,250	67,975	8.725

Interpretaciones

En la tabla 9.29 se puede ver que el rango intercuartil para cada grupo es casi el mismo. El rango intercuartil del grupo de las mujeres es ligeramente menor con respecto al rango intercuartil de los varones lo que nos permite afirmar que en el grupo de las mujeres existe una mayor variabilidad de masas corporales.

Como se comentó anteriormente, es el investigador quien debe decidir qué tipo de información estadística es más adecuada presentar. Esta decisión va de acuerdo con los objetivos de su investigación. Existen algunas investigaciones donde solo es necesario mostrar la siguiente información como resultado del análisis descriptivo:

Tabla 9.30 Estadísticos sobre la masa corporal.

Grupo	Media	Desviación estándar
Varones	78,513	5,774
Mujeres	63. 554	6,336

En la tabla 9.30 se observa otra forma de presentación de los resultados que se debe a otro tipo de objetivo de estudio. Por ejemplo, cuando se desea determinar si existen diferencias significativas entre la masa corporal de un grupo de varones y mujeres que presentan algunas características similares como la estatura y la edad. En este tipo de investigaciones es suficiente contar con la media y la desviación estándar como resultados del análisis descriptivo. En base a estos datos, luego se aplicarán otros procedimientos propios de la estadística inferencial para la comparación de medias.

Aquí concluye el análisis descriptivo de los resultados. No es necesario el uso de la estadística inferencial porque no se ha planteado ninguna hipótesis de investigación.

Ejemplo No. 2: Un equipo de estudiantes desarrolla una investigación experimental para identificar el efecto que tiene la proporción de levadura en la altura (en mm) del pan tipo “X” bajo un mismo proceso de elaboración. Para ello se prueba con 3 cantidades de levadura fresca y se registra la altura obtenida en cada pan.

Tabla 9.31 datos referentes a la cantidad de levadura y las alturas.

9 gramos		12 gramos		15 gramos	
Caso	Altura	Caso	Altura	Caso	Altura
1	30	1	27	1	32
2	32	2	28	2	35
3	29	3	30	3	30
4	30	4	32	4	37
5	29	5	29	5	30
6	31	6	29	6	33
7	28	7	30	7	34
8	26	8	32	8	32
9	30	9	30	9	36
10	25	10	31	10	36
11	27	11	30	11	29
12	29	12	27	12	30
13	28	13	32	13	32
14	28	14	28	14	32
15	28	15	27	15	30

Consideraciones metodológicas

Problema de investigación: ¿Cómo influye la proporción de levadura en la altura final del pan tipo “X”?

Objetivo de la investigación: Conocer la influencia de la proporción de levadura en la altura final del pan tipo “X”.

Hipótesis de la investigación: La proporción de levadura influye en forma directa en la altura del pan tipo “X”.

Variable independiente: Proporción de levadura.

Variable dependiente: Altura final del pan tipo “X”.

Escala de medición: Razón.

Tipo de investigación: Aplicada, cuantitativa, explicativa y experimental.

Resultados

Estadígrafos

Para este tipo de investigación donde el objetivo es la comparación de las alturas luego de los diferentes tratamientos no se requiere mostrar todos los estadísticos de manera detallada. Los estadígrafos que sirven de base para la etapa de la estadística inferencial son la media y la desviación estándar.

Tabla 9.32 Estadísticos de la altura según la cantidad de levadura.

Gramos	n	Mínimo	Máximo	Media	Desviación estándar
9	15	25,00	32,00	28,6667	1,83874
12	15	27,00	32,00	29,4667	1,80739
15	15	29,00	37,00	32,5333	2,55976

En la tabla 9.32 se observa que la mayor altura promedio se obtiene con 15 gramos de levadura y la menor altura promedio se obtiene cuando se utiliza 9 gramos de levadura. Sin embargo, la desviación estándar para el tercer grupo es mayor con respecto a la desviación estándar de los demás grupos. Para conocer si la diferencia, entre los promedios de alturas entre los grupos, es significativa se debe utilizar la estadística inferencial.

También se puede utilizar un diagrama de cajas para comparar las distribuciones de datos entre los grupos.

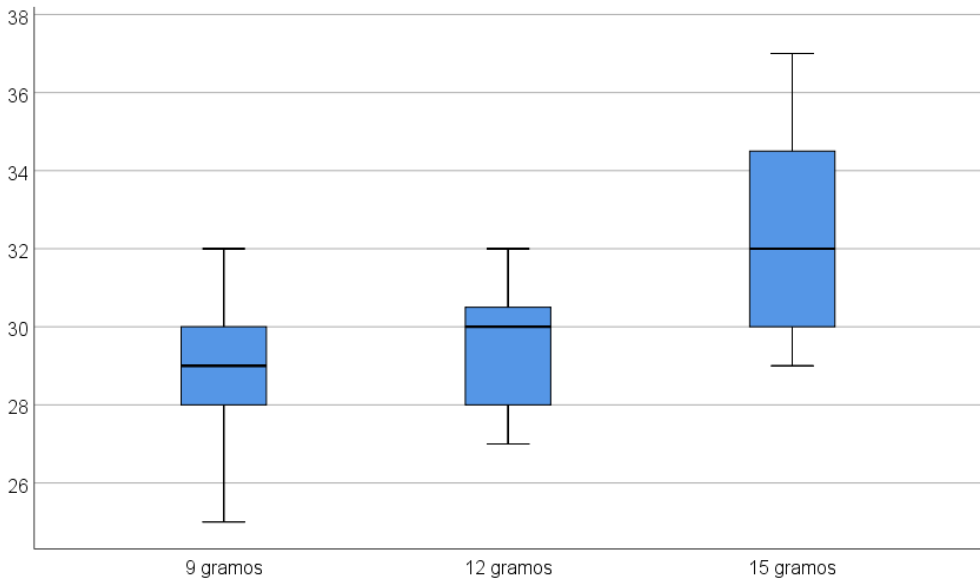


Figura 9.15 Distribución de las alturas según la cantidad de levadura.

En la figura 9.15 se observa que no existen valores atípicos en ninguno de los grupos lo que indica que las alturas de los panes dentro de cada grupo son relativamente homogéneas. En los grupos donde se utilizó 9 g y 15 g de levadura se observa que la mediana se ubica en la parte central de la caja lo que indica que presentan una distribución más homogénea en comparación con el grupo donde se utilizó 12 g de levadura. En este último grupo la mediana está más cerca al tercer cuartil lo que indica que una mayor concentración de los datos en los valores altos de esa distribución.

Como se indicó anteriormente, la etapa siguiente es la comparación de medias por medio de un procedimiento inferencial. Esta comparación de medias solo se podrá realizar si las distribuciones de los grupos son semejantes a la distribución normal. Para verificar esta condición podemos explorar superponiendo la curva normal sobre la gráfica de la distribución de cada grupo.

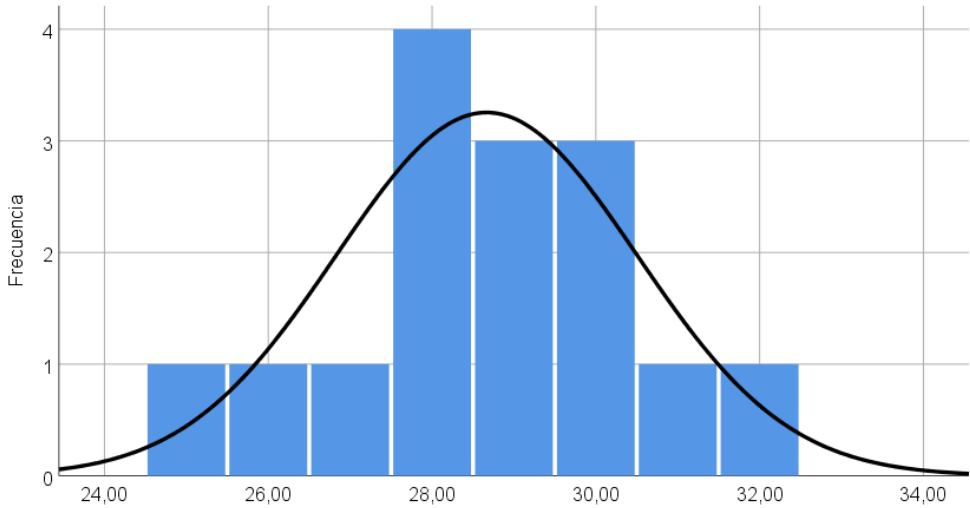


Figura 9.16 Alturas registradas cuando se usa 9 gramos de levadura.

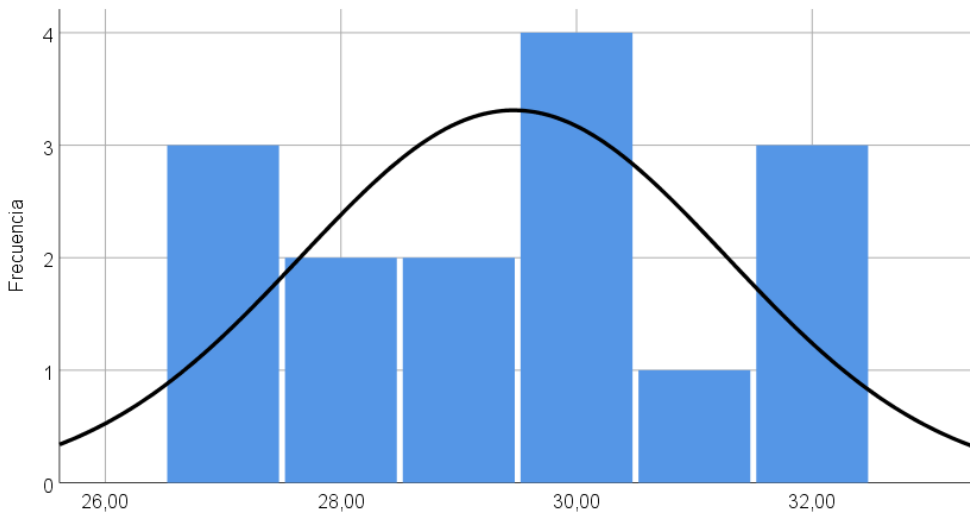


Figura 9.17 Alturas registradas cuando se usa 12 gramos de levadura.

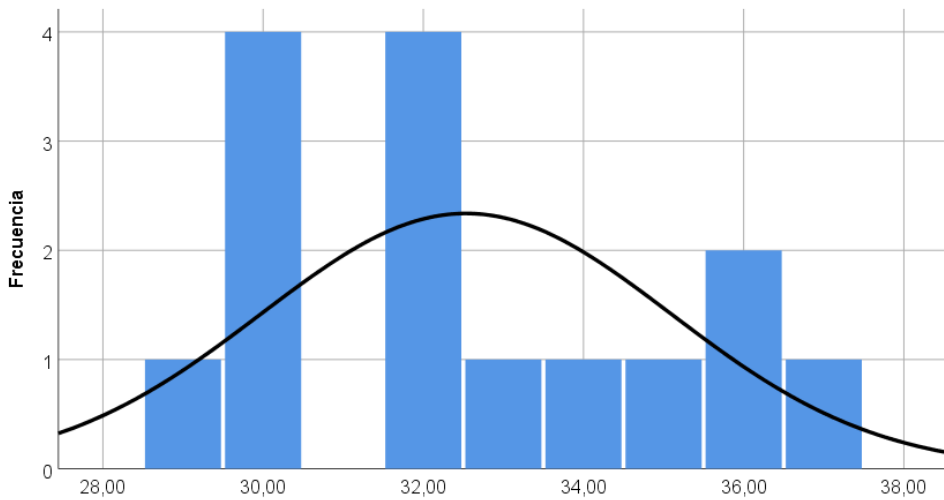


Figura 9.18 Alturas registradas cuando se usa 15 gramos de levadura.

Como se puede apreciar en las figuras 9.16, 9.17 y 9.18, solo en el grupo de alturas donde se utilizó 9 gramos de levadura se observa una semejanza de la distribución de estos datos con la distribución normal los demás grupos difieren notoriamente de la curva normal. Estas gráficas solo nos dan una noción del comportamiento de la distribución de datos. Sin embargo, existen pruebas estadísticas que pueden determinar si un conjunto de datos proviene de una población con distribución aproximadamente normal. Estas pruebas son la prueba de normalidad de Shapiro-Wilk y la prueba de Kolmogorov-Smirnov.

Aquí concluye el análisis descriptivo de los resultados. El paso siguiente consiste en seleccionar la prueba adecuada para establecer si la diferencia entre las medias obtenidas es significativa.

Referencias bibliográficas

- Anderson, D., Sweeney, D. y Williams, T. (2012). Estadística para negocios y economía. 11^o ed. México D.F. Cengage Learning Editores.
- Batanero, C. (2002). Los retos de la cultura estadística. Jornadas Interamericanas de Enseñanza de la Estadística. Buenos Aires, 2002. Conferencia inaugural. Recuperado de: <http://www.ugr.es/~batanero/pages/ARTICULOS/CULTURA.pdf>
- Batanero, C., Arteaga, P. y Ruiz, B. (2010). *Análisis de la complejidad semiótica de los gráficos producidos por futuros profesores de educación primaria en una tarea de comparación de dos variables estadísticas*. Enseñanza de las Ciencias, 28(1), 141-154. Recuperado de: https://www.researchgate.net/publication/44165784_Analisis_de_la_complejidad_semiotica_de_los_graficos_producidos_por_futuros_profesores_de_educacion_primaria_en_una_tarea_de_comparacion_de_dos_variables_estadisticas
- Bologna, E. (2011). Estadística para Psicología y Educación. Córdoba, Argentina. Editorial Brujas.
- Datum. (2016). Ideele Radio. Fernando Tuesta: Encuesta de Datum no indica una caída irreversible de la aprobación de PPK. Recuperado de: <https://ideeleradio.pe/lo-ultimo/fernando-tuesta-encuesta-de-datum-no-indica-una-caida-irreversible-de-la-aprobacion-de-ppk/>
- Gorgas, J., Cardiel, N. y Zamorano, J. (2011). *Estadística básica para estudiantes de ciencias*. Facultad de Ciencias Físicas de la Universidad Complutense de Madrid, España.
- Hernández, R., Fernández, C. y Baptista, M. (2010). *Metodología de la Investigación*. 5^o ed. México D.F. Editorial McGraw Hill.
- Hernández, A.A., Ramos, M.P., Placencia, B.M., Indacochea, B., Quimis, A.J. y Moreno, L.A. (2018). *Metodología de la Investigación Científica*. Alicante. Editorial Área de Innovación y Desarrollo.

Hopkins, K., Hopkins, B. y Glass, G. (1997). *Estadística básica para las ciencias sociales y del comportamiento*. México. Prentice-Hall Hispanoamericana.

IBM (s.f.) Knowledge Center: *Resumir Estadísticos*. Recuperado de: https://www.ibm.com/support/knowledgecenter/es/SSLVMB_sub/statistics_mainhelp_ddita/spss/base/idh_summ_stat.html

INEI, (2020). *Perú: Producto Bruto Interno por Departamentos 2007-2019*. Lima, Perú. Recuperado de: https://www.inei.gov.pe/media/MenuRecursivo/publicaciones_digitales/Est/Lib1759/libro.pdf

INEI, (2009). *Guía para la presentación de gráficos estadísticos*. Lima, Perú. Centro de Investigación y Desarrollo (CIDE).

INEI, (2020). *La Nupcialidad en los distritos de Lima Metropolitana y Región Lima, 2018*. Lima, Perú.

INEI (2020). *Indicadores de precios de la economía. Boletín mensual*. Lima, Perú.

INEI (2020). *Condiciones de vida de la población en riesgo frente a la pandemia del COVID-19. Encuesta Nacional de Hogares*. Lima, Perú.

INEI (2020). *Estado de la población peruana 2020*. Lima, Perú.

IPSOS. (2020). *Ficha técnica encuesta nacional urbana*. Recuperado de: https://www.ipsos.com/sites/default/files/ct/news/documents/2020-07/anexo_7_-_fichatecnica_julio_2020_nacional.pdf

Juárez, F., Villatoro, J., López, E. (2002), *Apuntes de Estadística Inferencial*. México, D.F.: Instituto Nacional de Psiquiatría Ramón de la Fuente.

Kazmier, L. y Díaz, A. (1991). *Estadística aplicada a administración y economía*. 2° ed. México D.F. Editorial McGraw Hill.

- Lerma, H. D. (2011). *Presentación de informes: el documento final de investigación*. 3° ed. Bogotá. Ecoe Ediciones.
- Leyton, O. (2018). *La investigación en ciencias sociales. Establecimiento de hipótesis, métodos y técnicas de investigación, Información cualitativa y cuantitativa*. 2° ed. México D.F. Editorial Trillas.
- Lind, D., Wathen, S.A. y Marchal, W. G. (2016). *Estadística aplicada a los negocios y la economía*. 16° ed. Madrid, España: Editorial McGraw Hill.
- Martinez, C. (2012). *Estadística y muestreo*. 13° ed. Bogotá, Colombia. ECOE ediciones.
- Martínez, H. (2012). *Metodología de la investigación*. México D.F. Cengage Learning Editores
- Medenhall, W., Beaver, R. y Beaver, B. (2010). *Introducción a la probabilidad y estadística*. 13° ed. México D.F. Cengage Learning Editores.
- Mendoza, J. y Garza, J. (2009) *La medición en el proceso de investigación científica: Evaluación de validez de contenido y confiabilidad*. México. UANL. Recuperado de: <http://eprints.uanl.mx/12508/1/A2.pdf>
- Moya, R. (2007). *Estadística descriptiva. Conceptos y aplicaciones*. Lima, Perú. Editorial San Marcos.
- Naciones Unidas (2009). *Cómo hacer comprensibles los datos Parte 2. Una guía para presentar estadísticas*. Comisión Económica Para Europa. Ginebra, Suiza. Recuperado de: https://unece.org/DAM/stats/documents/writing/MDM_Part2_Spanish.pdf
- Ñaupas, H., Mejía, E., Novoa, E. y Villagomez, A. (2014). 4° ed. Bogotá, Colombia. Ediciones de la U.

- Ruiz, F. (2005). *El concepto de medición. Reflexiones en torno a la medición de la innovación en América Latina*. Recuperado de: <https://www.redalyc.org/pdf/3442/344234270002.pdf>
- SEDAPAL. (2016). Estudio de impacto ambiental detallado del proyecto “Optimización de sistemas de agua potable y alcantarillado, sectorización, rehabilitación de redes y actualización de catastro – área de influencia planta Huachipa – área de drenaje Comas – chillón - lima”. Recuperado de: <http://www.sedapal.com.pe/Contenido/ambiental/ambiental/disco2/ANEXO%2016%20RESULTADO%20TALLERES%20INFORMATIVOS/Resultado%20de%20la%20encuesta.pdf>
- Siegel, S. y Castellan, J. (1995). *Estadística no paramétrica; Aplicada a las ciencias de la conducta*. 4° ed. México D.F. Editorial Trillas.
- Soto Abanto, S. E. (2018). Variables, dimensiones e indicadores en una tesis. Recuperado de: <https://tesisciencia.com/2018/08/20/tesis-variables-dimensiones-indicadores>.
- Spiegel, M. y Stephens, L. (2009). *Estadística*. 4° ed. México D.F. Editorial McGraw Hill.
- Tamayo, M. (2003). *El proceso de la investigación científica. Incluye evaluación y administración de proyectos de investigación*. 4° ed. México D.F. Limusa Editores.
- Triola, M. (2004). *Estadística*. 9° ed. México. Pearson Educación.
- Vargas, V. (2007). *Estadística descriptiva para ingeniería ambiental con SPSS*. Cali, Colombia. Universidad Nacional de Colombia.
- Wackerly, D., Mendenhall, W. y Scheaffer, R. (2010). *Estadística matemática con aplicaciones*. 7° ed. México D.F. Cengage Learning Editores.

ISBN: 978-612-00-6118-3



9 786120 061183