UNIVERSIDAD NACIONAL DE JAÉN

INSCID
Instituto de Investigación de
Ciencia de Datos

PRINCIPIOS DE LA ESTADÍSTICA DESCRIPTIVA

Autores
Dra. Marcela Yvone Saldaña Miranda
Mg. Rosario Yaqueliny Llauce Santamaria
Dra. Cinthya Yanina Santa Cruz López
Dr. Lenin Quiñones Huatangari

PRESENTACIÓN

Hacer la presentación del libro "Principios de la Estadística Descriptiva", de autoría de cinco grandes y comprometidos docentes de la Universidad Nacional de Jaén, es muy honroso para mí, a la vez que trae a mi memoria, algunos mitos sobre la "Estadística", mitos que los autores, han derrumbado de una forma muy sutil a través de las páginas de este libro.

Según presentan los autores, la estadística en sus inicios fue relacionada con censos y registros de bienes, posteriormente a la aparición de escuelas y su consideración como disciplina independiente, hasta la actualidad, que aparece el concepto de métodos para la obtención de conclusiones generales, con un nuevo enfoque que denomina la Estadística inductiva o Estadística inferencial.

Mientras que, para algunos, la estadística es sencilla y de gran importancia en la formación profesional, para otros, sobre todo, para los estudiantes, la estadística es sinónimo de dificultad, angustia, ansiedad, sin embargo, con la publicación del libro "Principios de la Estadística Descriptiva", se plantea una estadística fácil de aprender y aplicar.

Quien tenga en sus manos este libro, tenga por seguro, que tendrá la oportunidad de aprender y comprender la estadística descriptiva, desde las definiciones básicas hasta los conceptos fundamentales, todo ello en base a la serie de ejercicios resueltos y propuestos de manera didáctica.

Dra. Bertha Talledo Torres

Dra. Marcela Yvone Saldaña Miranda
Mg. Rosario Yaqueliny Llauce Santamaria
Dra. Cinthya Yanina Santa Cruz López
Dr. Lenin Quiñones Huatangari

PRINCIPIOS DE LA ESTADÍSTICA DESCRIPTIVA

INSCID - INSICIENSA

PRINCIPIOS DE LA ESTADISTICA DESCRIPTIVA

Autores:
Dra.Marcela Yvone Saldaña Miranda
Mg.Rosario Yaqueliny Llauce Santamaria
Dra.Cinthya Yanina Santa Cruz López
Dr.Lenin Quiñones Huatangari

Edición de:
Universidad Nacional de Jaén. Fondo Editorial
Dirección: Km.243 de la carretera Jaén - San Ignacio, Cajamarca -Perú.
https://unj.edu.pe/

1era Edición digital – septiembre 2024
Libro electrónico disponible en:

Hecho el depósito legal en la Biblioteca Nacional del Perú N°
ISBN:

PRESENTACIÓN

Este libro es fruto de la experiencia como docentes universitarios y aborda de manera clara las técnicas y métodos esenciales de la estadística descriptiva. Fue diseñado con el propósito de cubrir necesidades formativas y como una herramienta didáctica para superar la carencia bibliográfica; mediante la explicación sencilla de los métodos proporcionará una base sólida y comprensible, facilitando el aprendizaje y aplicación de la estadística descriptiva.

Representa una contribución por parte de los docentes hacia el sistema educativo y fortalece las competencias de la comunidad estudiantil. Es un testimonio de esfuerzo y fe en la educación peruana; se entrega en manos de aquellos que participan activamente en la labor educativa, reflejando la convicción de que día a día se fortalece en las perspectivas de hombres y mujeres que desempeñan un papel crucial en el ámbito educativo peruano.

El libro "Principios de la Estadística Descriptiva" está organizado con las temáticas en sus primeros seis capítulos: Definiciones básicas de la estadística, organización y clasificación de los datos, análisis descriptivo de datos cuantitativos, medidas de forma de la distribución, distribuciones bidimensionales y coeficiente de correlación lineal. Además, en el capítulo siete se cubre de manera integral los conceptos fundamentales de la estadística con ejercicios resueltos y finalmente el capítulo ocho se plantean ejercicios propuestos, proporcionando una comprensión práctica de los conceptos presentados en cada capítulo.

Los editores

ÍNDICE

1 ESTADÍSTICA GENERAL 11

1.1 Etimología del término estadística y definición 11

1.2 Reseña histórica 11

1.3 División de la estadística 13

1.3.1 Estadística descriptiva 13

1.3.2 Estadística Inferencial 13

1.3.3 Utilidad e importancia de la estadística 14

1.3.4 Fases de la estadística 15

1.4 La estadística en el desarrollo de investigación 16

1.5 Etapas de la investigación estadística 16

1.5.1 Formulación del problema 16

1.5.2 Diseño de la investigación 17

1.5.3 Recolección de datos 17

1.5.4 Organización y presentación de datos 18

1.5.4.1 Análisis de datos 18

1.5.4.2 Interpretación de resultados 18

1.5.4.3 Conclusión y recomendaciones 18

1.5.4.4 Comunicación de resultados 18

1.6 Instrumentos de recolección de datos 19

1.6.1 Cuestionarios 19

1.6.2 Entrevistas 20

1.6.3 Observación 20

1.6.4 Registros y documentos 20

1.6.5 Escalas de medición 21

1.6.6 Grupos focales 21

1.6.7 Experimentos 21

1.7 Fuentes de información para la recolección de los datos 21

1.7.1 Fuentes primarias 21

1.7.2 Fuentes secundarias 22

1.7.3 Fuentes digitales 22

1.8 Censo 23

1.9 Registros 23

1.10 Población 23

Parámetro 24

1.11 Muestra 24

Estadístico 25

1.12 Variable 25

Clasificación de las variables 26

1.12.1 Según la naturaleza de la variable 26

1.12.2 Según su escala de medición 27

1.13 Unidad de análisis 28

1.14 Unidad de muestreo 28

1.15 Marco muestral 28

1.16 Diseño muestral 29

1.17 Datos 29

2 ORGANIZACIÓN Y CLASIFICACIÓN DE DATOS 30

2.1 Introducción 30

2.2 Revisión y corrección de datos 30

2.3 Distribución de frecuencias 30

2.3.1 Frecuencia absoluta 31

2.3.2 Frecuencia relativa 31

2.4 Distribución de frecuencias para variables cualitativas 33

2.4.1 Gráficas para variables cualitativas 34

2.4.1.1 Grafica de barras 34
2.4.1.2 Grafica de sectores 35
2.5 Distribución de frecuencias para variables cuantitativas discretas 37
2.5.1 Gráficas para variables cuantitativas discretas 37
2.5.1.1 Grafica de bastón 37
2.5.1.2 Grafica de escalera 38
2.6 Distribución de frecuencias para variables cuantitativas Continuas 41
2.6.1 Construcción de la distribución de frecuencias 41
2.6.2 Gráfica de la distribución por intervalos 42
2.6.2.1 Histograma 42
2.6.2.2 Polígono. 42
2.6.2.3 Ojivas 43
2.6.3 Límites reales de clases 46
2.6.4 Simetría de una tabla de distribución de frecuencias 47
2.6.5 Diagrama de hojas y tallos 50
3 ANÁLISIS DESCRIPTIVO DE DATOS CUANTITATIVOS 51
3.1 Introducción 51
3.2 Estadígrafo de posición 52
3.2.1 Promedio o media aritmética 52
3.2.1.1 Cálculo de la media aritmética 52
3.2.1.2 Propiedades de la media o promedio 55
3.2.1.3 Ventajas y desventajas: 57
3.2.1.4 Aplicaciones 57
3.2.2 Media geométrica 58
3.2.2.1 Cálculo de la media geométrica 58
3.2.2.2 Ventajas y Desventajas 60
3.2.2.3 Aplicaciones de la media geométrica 60

3.2.3 Media armónica 63

3.2.3.1 Cálculo de la media armónica 63

3.2.3.2 Ventajas y desventajas 64

3.2.3.3 Aplicaciones 64

3.2.4 Mediana 65

3.2.4.1 Cálculo de la mediana: 65

3.2.4.2 Propiedades de la mediana. 73

3.2.4.3 Ventajas y desventajas 75

3.2.5 Moda. 77

3.2.5.1 Cálculo de la moda 78

3.2.5.2 Ventajas y desventajas 81

3.2.5.3 Relación entre media, mediana y moda 82

3.3 Medidas de dispersión 84

3.3.1 Introducción 84

3.3.2 Rango de la variable 84

3.3.3 Recorrido intercuartílico 85

3.3.4 Desviación del cuartil 85

3.3.5 Desviación media absoluta 86

3.3.6 Varianza 89

3.3.6.1 Varianza poblacional 89

3.3.6.2 Varianza muestral 91

3.3.7 Desviación típica o desviación estándar 93

3.3.7.1 Ventajas y Desventajas 95

3.3.8 Coeficiente de variación 96

4 MEDIDAS DE FORMA DE LA DISTRIBUCION 97

4.1 Introducción 97

4.2 Medidas de asimetría 97

4.2.1 Coeficiente de Pearson. 97
4.2.1.1 Representación gráfica del coeficiente de asimetría 99
4.3 Medidas de apuntamiento o curtosis 101
4.3.1 Tipos de curtosis 101
4.3.1.1 Representación gráfica de la curtosis 101
4.3.2 Cálculo de la *curtosis*: 101
5 DISTRIBUCIONES BIDIMENSIONALES 103
5.1 Introducción 103
5.2 Tablas estadísticas bidimensionales 104
5.2.1 Frecuencias absolutas 104
5.2.2 Frecuencias Relativas 105
5.2.3 Distribuciones marginales. 107
5.2.4 Distribución condicional. 109
6 CORRELACIÓN Y REGRESIÓN LINEAL 117
6.1 Introducción 117
6.2 Diagrama de dispersión 117
6.3 Correlación Lineal 118
6.4 Coeficiente de determinación 122
6.5 Regresión lineal 123
7 EJERCICIOS RESUELTOS 125
Aplicaciones estadísticas utilizando el software Jamovi 125
7.1 Capítulo 1 125
7.2 Capítulo 2 132
7.3 Capítulo 3 144
7.4 Capítulo 4 160
7.5 Capítulo 5 170
7.6 Capítulo 6 177

7.7 Capítulo 7 184
8 EJERCICIOS PROPUESTOS 194
8.1 Capítulo 1 194
8.2 Capítulo 2 205
8.3 Capítulo 3 213
8.4 Capítulo 4 218
8.5 Capítulo 5 220
8.6 Capítulo 6 223
9 REFERENCIAS BIBLIOGRÁFICAS 225

INTRODUCCIÓN

Este libro de estadística descriptiva está diseñado para satisfacer las necesidades de estudiantes de ciencias, ingenierías y especialidades afines, brindándoles una herramienta práctica y comprensible para desarrollar conceptos básicos de estadística y investigación científica.

Su objetivo es llenar el vacío existente en el aprendizaje del estudiante, proporcionando una guía integral para clases teóricas y prácticas. Los autores, con experiencia docente en la asignatura, han creado este material para optimizar el aprendizaje estudiantil.

El libro incluye:

- Conceptos básicos de estadística y investigación científica
- Ejercicios prácticos con el software estadístico Jamovi
- Ejemplos detallados para guiar el uso del software
- Aplicaciones en diversas situaciones de investigación científica

Este enfoque busca fortalecer las habilidades de análisis de datos en los estudiantes, permitiéndoles interpretar resultados de forma precisa y eficiente. El libro es una herramienta indispensable para cualquier estudiante que busque desarrollar habilidades en estadística y investigación científica."

Los autores

1 ESTADÍSTICA GENERAL

1.1 ETIMOLOGÍA DEL TÉRMINO ESTADÍSTICA Y DEFINICIÓN

El origen del término “estadística” no está bien definido. Existen varias posiciones al respecto, de ellas destaca las que sostiene que el término se deriva de las siguientes palabras:

Staat que en alemán significa “estado” o

Status que en latín significa “situación” o “estado”

En sus comienzos, la estadística estuvo asociada con las actividades del gobierno, como la recopilación de datos sobre nacimientos, fallecimientos, producción, impuestos, empleo, entre otros. Además, del hecho de que en Alemania se creó por primera vez la cátedra de Estadística como disciplina independiente.

La estadística estudia los métodos científicos para recolectar, organizar, resumir y analizar datos, así como para sacar conclusiones efectivas y tomar decisiones razonables.

1.2 RESEÑA HISTÓRICA

Durante su proceso de formación y desarrollo la estadística ha pasado por tres etapas que se describen:

Etapa inicial

Se remonta a la antigüedad y comprende hasta mediados del siglo XVIII. Está relacionada con censos de población, registro de bienes y datos económicos de los pueblos o Estados. Estos registros incluyen actas de nacimientos, defunciones, impuestos, recaudación, etc.

Cabe destacar los hechos importantes:

- En Egipto se realizaban periódicas recolecciones de datos de interés para la administración estatal. Divinizaron a SAPHKIT, diosa de los libros y cuentas.
- Los hebreos llevaban registros de datos estadísticos. La Biblia menciona el censo que Moisés realizó por encargo de Jehová.
- En China, el emperador Yao ordenó un censo en el año 1158 A.C.
- Los romanos llevaban registros estadísticos con fines tributarios.

- Los árabes, en el año 727, realizaron un censo en la península ibérica.
- Durante la Edad Media el clero se dedicó a llevar estadísticos demográficos, en forma sistemática.
- Destacan las estadísticas económicas contenidas en el archivo de la Indias de Sevilla.
- A partir del concilio de Trento (1 545 – 1 563) se establece la obligatoriedad de inscribir los nacimientos, matrimonios y defunciones.
- En América pre - colombiana, en el imperio Tahuantinsuyo, se llevaron registros de datos demográficos y socioeconómicos que estaban a disposición del Inca mediante el empleo del quipu, que era utilizado por funcionario del estado llamado Quipu Camayoc.

Etapa de sistematización

Aparecen escuelas que sistematizan como disciplina independiente, a mediados del siglo XVII, en la etapa inicial se encuentra antecedentes vinculados al registro de datos demográficos y socio – económicos.

Escuela Alemana (Escuela administrativa): Nace en Alemania, creando la primera cátedra de estadística, considerándola por primera vez como disciplina independiente cuya función es la descripción de los fenómenos concernientes a la administración del estado. En 1 660, Hermann Conring (1 606 – 1 681), comenzó a enseñar un curso de Estadística en la Universidad de Helmstadt. Uno de sus seguidores fue Godofredo de Achenwall (1 719 – 1 772) fue gran teórico de esta ciencia y le dio el nombre de Estadística, basado en el término "Status".

Escuela inglesa también conocida como la (Escuela demográfica): surgió en Inglaterra alrededor de la misma época. Su objetivo principal era establecer de manera cuantitativa las leyes que gobiernan el comportamiento de los fenómenos políticos y sociales; los mismos que no pueden ser independientes del tamaño, estructura y distribución de la población. Destacan William Petty (1 623 – 1 687), Edmund Halley (1 662 – 1 855) y John Graunt (1 620 – 1 674).

Escuela francesa (Escuela Probabilística): Introduce la teoría de probabilidades como fundamento matemático de la estadística. Destacan en el desarrollo e impulso de la teoría de probabilidades

Etapa actual

Comprende desde principios del siglo XIX hasta la actualidad. Al comienzo el interés fundamental era contar con la colección completa de datos relacionados con el estudio. A partir de los treinta del siglo XX crece un nuevo concepto que consistió en desarrollar métodos que hagan posible obtener conclusiones generales a partir de un estudio de muestras; es decir, de estudios parciales. Este nuevo enfoque de la estadística se denomina Estadística Inductiva o Estadística Inferencial.

1.3 DIVISIÓN DE LA ESTADÍSTICA

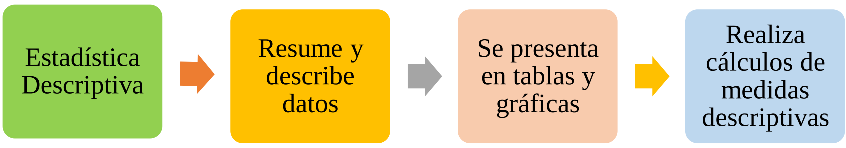
Se clasifica en dos principales categorías:

1.3.1 ESTADÍSTICA DESCRIPTIVA

Implica la presentación de datos mediante tablas y gráficos, incluyendo todas las actividades asociadas con estos datos. Su propósito es resumir un conjunto de datos sin hacer inferencias que vayan más allá de los datos en sí.

Figura 1

Pasos de la estadística descriptiva



1.3.2 ESTADÍSTICA INFERENCIAL

La estadística inferencial se dedica a investigar o analizar una población en base a una muestra seleccionada. Ante ello, los problemas frecuentes son la estimación de parámetros y pruebas de hipótesis.

Figura 2

Pasos de la estadística inferencial

Ejemplo 1.

- El gerente del "Hospital General de Jaén" quiere conocer el rendimiento laboral promedio de sus trabajadores, pero no dispone de tiempo ni recursos suficientes para evaluar a todo el personal. Por ello, opta por evaluar a 20 empleados del hospital. A partir de esta muestra, busca hacer una estimación del rendimiento laboral promedio de todos los trabajadores. Este proceso corresponde a una inferencia estadística, ya que se utiliza la información obtenida de un grupo reducido de empleados para hacer suposiciones sobre el comportamiento general del total de trabajadores en el hospital.

- Un gerente asegura que el 95% de los pacientes aceptará el nuevo medicamento que se va a probar. Se selecciona una muestra de 50 pacientes para evaluar esta afirmación, obteniendo los siguientes resultados: 20 pacientes aceptan el nuevo medicamento, mientras que los demás prefieren el anterior. Ante estos resultados, surge la pregunta de si la afirmación del gerente debería ser descartada. El análisis de esta situación se enmarca en la inferencia estadística, pues se busca evaluar la proporción de pacientes que aceptan el medicamento en función de los datos obtenidos de la muestra seleccionada.

$$p = \frac{20}{50} = 0.40$$

La proporción, expresada en porcentajes, muestra solo un 40% de aceptación, lo que es significativamente menor que el 95% indicado por el gerente. Por lo tanto, basándonos en estos resultados, se rechaza la afirmación inicial del gerente sobre el nivel de aceptación del nuevo medicamento entre los pacientes.

1.3.3 UTILIDAD E IMPORTANCIA DE LA ESTADÍSTICA

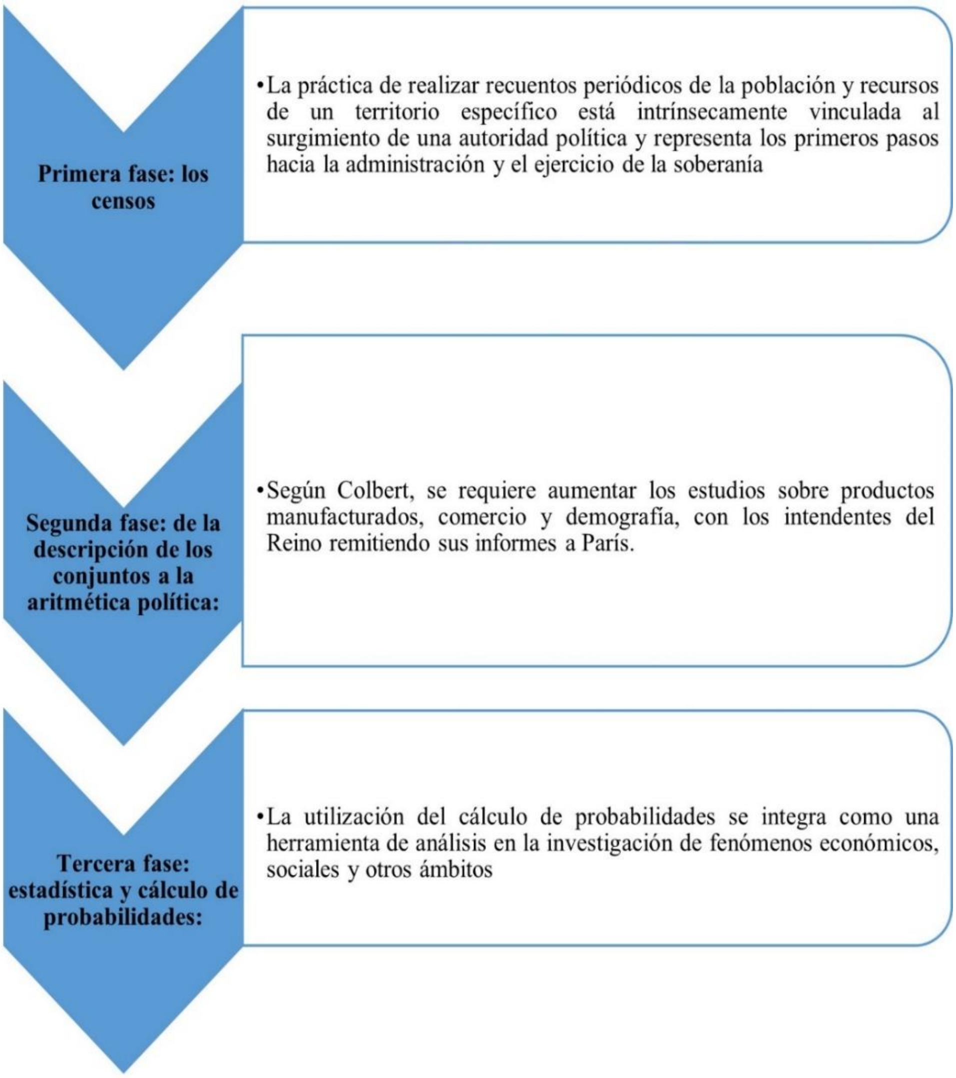
La estadística es esencial para analizar datos y apoyar la toma de decisiones informadas en diversas áreas como la investigación científica, la salud pública, la economía y la industria. Facilita la descripción de fenómenos complejos, permite identificar patrones, y es clave en la predicción de escenarios futuros. Además, es fundamental en el control de calidad, la mejora de procesos y el análisis de grandes volúmenes de datos. Su aplicación ayuda a transformar la información en conocimiento valioso para resolver problemas y generar avances significativos en múltiples disciplinas

1.3.4 FASES DE LA ESTADÍSTICA

El desarrollo histórico de la estadística se puede dividir en tres fases distintas.

Figura 3

Fases de la estadística



1.4 LA ESTADÍSTICA EN EL DESARROLLO DE INVESTIGACIÓN

La estadística desempeña un papel crucial en el desarrollo de la investigación, ya que permite diseñar estudios rigurosos, recopilar y analizar datos de manera eficiente, y extraer conclusiones válidas. Ayuda a formular hipótesis, interpretar resultados y evaluar la fiabilidad de los hallazgos. Además, mediante técnicas estadísticas, se puede medir la variabilidad y controlar los posibles errores, lo que asegura que los resultados obtenidos sean representativos y precisos. Así, la estadística es fundamental para la creación de conocimiento científico basado en evidencia sólida.

1.5 ETAPAS DE LA INVESTIGACIÓN ESTADÍSTICA

Las etapas de la investigación estadística son un conjunto de pasos sistemáticos que permiten recopilar, analizar e interpretar datos para resolver un problema o responder a una pregunta de investigación. Estas etapas aseguran que el proceso de investigación se realice de manera estructurada y científica, permitiendo obtener resultados fiables y válidos.

Incluyen la definición del problema de investigación, el diseño de la metodología**,** la recolección de datos mediante técnicas apropiadas, el procesamiento y análisis de esos datos utilizando herramientas estadísticas, la interpretación de los resultados obtenidos y, finalmente, la presentación de conclusiones y posibles recomendaciones.

A continuación, se describen las principales etapas:

1.5.1 FORMULACIÓN DEL PROBLEMA

En esta fase se define claramente el objetivo de la investigación, se identifican las preguntas que se quieren responder y las hipótesis que se van a probar. También se determina el tipo de información que se necesita recolectar.

El planteamiento o preparación de una investigación comprende:

- Formulación del problema.
- Justificación del estudio.
- Planteamiento de objetivos.
- Formulación de hipótesis.
- Identificación de variables.
- Selección de fuentes de información.

- Antecedentes de la investigación.
- Determinación ámbito geográfico y periodo de datos.
- Selección de la población y muestra.
- Determinación de métodos, técnicas e instrumentos para recopilar y analizar información.
- Elaboración y aplicación de instrumentos para la recolección de datos.
- Elaboración del cronograma de actividades.
- Realización del presupuesto y fuentes de financiación.
- Capacitación del equipo de trabajo.

1.5.2 DISEÑO DE LA INVESTIGACIÓN

Se planifica cómo se recopilarán los datos, definiendo el método de muestreo (aleatorio, estratificado, sistemático, etc.), el tamaño de la muestra y el tipo de datos (cuantitativos o cualitativos). También se eligen las herramientas para la recolección de datos, como encuestas, experimentos u observaciones.

1.5.3 RECOLECCIÓN DE DATOS

Es el proceso mediante el cual se obtienen y reúnen datos o información relevante para responder a una pregunta de investigación o para analizar un fenómeno en particular. Este proceso es crucial en cualquier investigación, ya que la calidad de los datos recolectados influye directamente en la validez y confiabilidad de los resultados.

La recolección de datos se puede realizar mediante dos modalidades:

a. La Investigación bibliográfica o Investigación Documental

Consiste en tomar información estadística (datos procesados de acuerdo con los objetivos y criterios específicos) de las fuentes documentales disponibles tales como: Oficinas de estadística, Archivos o registros administrativos, Informes, boletines estadísticos y publicaciones de organismos especializados en temas relacionados con el problema que se estudia. A estas fuentes de datos se le denomina fuentes secundarias.

b. El trabajo de Campo

Consiste en tomar datos directamente de las unidades de análisis que conforman la muestra o población en estudio. El investigador entra en contacto directo con las personas, instituciones u objetos que constituyen las unidades de análisis, obteniendo datos originales. Por lo tanto, en este caso las unidades de análisis son las fuentes primarias de datos.

1.5.4 ORGANIZACIÓN Y PRESENTACIÓN DE DATOS

Una vez recopilados, los datos deben organizarse y presentarse en tablas, gráficos o diagramas para facilitar su análisis y comprensión. Esta etapa permite visualizar de manera clara la información recogida.

1.5.4.1 Análisis de datos

Se aplican métodos estadísticos para examinar los datos, detectar patrones, tendencias o relaciones y evaluar las hipótesis planteadas. Este análisis puede incluir tanto medidas descriptivas (media, mediana, varianza) como inferenciales (pruebas de hipótesis, regresión, etc.).

1.5.4.2 Interpretación de resultados

En esta fase, se interpretan los resultados del análisis estadístico en el contexto del problema inicial. Se discuten las implicaciones de los hallazgos y se decide si confirman o refutan las hipótesis planteadas.

1.5.4.3 Conclusión y recomendaciones

Se presentan las conclusiones derivadas del análisis y se formulan recomendaciones basadas en los resultados obtenidos. Esta etapa también puede incluir sugerencias para futuras investigaciones o para la aplicación práctica de los hallazgos.

1.5.4.4 Comunicación de resultados

Finalmente, los resultados se comunican a través de informes, publicaciones o presentaciones. Es crucial que los hallazgos sean explicados de manera clara y comprensible para el público objetivo, ya sea académico, empresarial o gubernamental.

Figura 4

Etapas del Método Científico

1.6 INSTRUMENTOS DE RECOLECCIÓN DE DATOS

Los instrumentos de recolección de datos son las herramientas que utilizamos para obtener la información necesaria para responder a nuestras preguntas de investigación. Estos instrumentos varían en complejidad y se adaptan a diferentes tipos de estudios y objetivos.

Los instrumentos de recolección de datos más comunes son:

1.6.1 CUESTIONARIOS

Son una serie de preguntas, generalmente estructuradas, que pueden ser respondidas de manera escrita o digital. Pueden contener ***preguntas Cerradas*****:** Ofrecen opciones de respuesta predefinidas (opción múltiple, Likert, dicotómica), ***Abiertos*****:** Permiten al encuestado expresar sus opiniones libremente y ***Mixtos*****:** Combinan preguntas cerradas y abiertas. Se utilizan cuando se necesita obtener información de un gran número de personas. Son comunes en encuestas, estudios de mercado, investigaciones sociales, entre otros.

- **Ventajas**: Bajo costo, facilidad de administración a gran escala.
- **Desventajas**: Riesgo de respuestas inexactas si las preguntas no están bien diseñadas.

1.6.2 ENTREVISTAS

Interacción cara a cara (o virtual) entre un entrevistador y un entrevistado, donde se formulan preguntas directas. Las entrevistas pueden ser estructuradas (preguntas preestablecidas), semiestructuradas (Se tiene una guía de preguntas, pero se permite la flexibilidad) o no estructuradas (Se basa en temas generales y la conversación fluye libremente). Se emplea para recolectar información cualitativa en profundidad sobre opiniones, experiencias y actitudes.

- **Ventajas**: Permiten obtener información detallada y profunda.
- **Desventajas**: Requieren más tiempo y recursos; pueden introducir sesgos si no se administran correctamente.

1.6.3 OBSERVACIÓN

Consiste en la recolección de datos a través de la observación directa de comportamientos, eventos o fenómenos en su entorno natural o en un entorno controlado. Puede ser participante (El investigador se integra al grupo estudiado) o no participante (El investigador observa desde afuera). Se aplica en estudio de comportamientos humanos, procesos sociales o interacciones en el campo de la sociología, antropología, psicología, entre otros.

- **Ventajas**: Proporciona datos en tiempo real y puede capturar detalles que otras técnicas podrían no registrar.
- **Desventajas**: Puede ser subjetiva, requiere un buen entrenamiento del observador y puede ser costosa y demorada.

1.6.4 REGISTROS Y DOCUMENTOS

Uso de fuentes documentales o registros existentes, como informes, actas, historias clínicas (registros médicos de pacientes), archivos (documentos históricos, legales, etc.) y bases de datos (conjuntos estructurados de datos), publicaciones oficiales, entre otros. Se usa en investigación histórica, análisis de tendencias, estudios de archivo.

- **Ventajas**: Acceso a grandes cantidades de datos históricos o ya recolectados.
- **Desventajas**: Dependencia de la calidad y confiabilidad de los registros existentes.

1.6.5 ESCALAS DE MEDICIÓN

Herramientas que permiten cuantificar actitudes, opiniones o comportamientos en términos numéricos. Ejemplos comunes son la escala de Likert, la escala diferencial semántica, entre otras. Se emplea evaluar percepciones, actitudes o satisfacción en diversas áreas, como la psicología, sociología o el marketing.

- **Ventajas**: Facilitan la comparación y el análisis estadístico.
- **Desventajas**: Pueden limitar las respuestas de los participantes y no siempre captan la complejidad de las opiniones.

1.6.6 GRUPOS FOCALES

Reunión de un pequeño grupo de personas donde se discute sobre un tema específico bajo la moderación de un facilitador. Se aplica en investigaciones cualitativas, especialmente para explorar percepciones, actitudes y opiniones sobre productos, servicios, o fenómenos sociales.

- **Ventajas**: Permite una interacción dinámica que puede generar ideas profundas y variadas.
- **Desventajas**: Sesgos del grupo, posible influencia entre los participantes.

1.6.7 EXPERIMENTOS

Método en el que los investigadores manipulan una o más variables y observan los efectos en otra(s) variable(s). Puede ser en un entorno controlado (laboratorio) o en el campo. Se utiliza en investigaciones científicas en las que se desea establecer relaciones de causa y efecto.

- **Ventajas**: Alto nivel de control, permite establecer causalidad.
- **Desventajas**: Puede ser artificial y no reflejar condiciones del mundo real.

1.7 FUENTES DE INFORMACIÓN PARA LA RECOLECCIÓN DE LOS DATOS

La recolección de datos se puede realizar a través de diversas fuentes de información, dependiendo del tipo de investigación o estudio que estés realizando. Las principales fuentes de información para la recolección de datos incluyen:

1.7.1 FUENTES PRIMARIAS

Estas son fuentes donde se recogen datos originales, directamente del objeto de estudio. Algunos ejemplos son:

- **Encuestas y cuestionarios**: Se recopilan respuestas directas de los encuestados.
- **Entrevistas**: Conversaciones directas con personas clave, donde se puede obtener información en profundidad.
- **Observación directa**: Recolección de datos mediante la observación del comportamiento o fenómenos sin intervención.
- **Experimentos**: Datos obtenidos a través de pruebas controladas y manipulaciones de variables en un entorno experimental.
- **Grupos focales**: Conversaciones grupales donde los participantes discuten un tema específico bajo la moderación de un facilitador.

1.7.2 FUENTES SECUNDARIAS

Son datos ya existentes que fueron recolectados por otras personas u organizaciones, que luego pueden ser utilizados para el análisis. Algunos ejemplos son:

- **Estudios y reportes previos**: Informes, tesis, publicaciones científicas, libros y artículos periodísticos.
- **Bases de datos**: Repositorios de información pública o privada, como bases de datos de instituciones gubernamentales, universidades, empresas, etc.
- **Censos y estadísticas oficiales**: Datos demográficos, económicos, de salud, entre otros, publicados por instituciones gubernamentales.
- **Documentos históricos**: Archivos, registros antiguos, actas de reuniones, cartas, etc.
- **Publicaciones comerciales**: Informes de mercado, estudios de tendencias industriales, entre otros.

1.7.3 FUENTES DIGITALES

Hoy en día, las tecnologías de la información permiten obtener datos de varias maneras, tales como:

- **Redes sociales**: Análisis de comportamientos, preferencias y tendencias a partir de datos de redes sociales como Twitter, Facebook, Instagram, entre otras.
- **Big Data**: Grandes volúmenes de datos que se generan de forma continua a través de sensores, dispositivos, transacciones electrónicas, etc.

- **Páginas web y blogs**: Información publicada en sitios web especializados, blogs personales o corporativos.

1.8 CENSO

Proceso de recolectar y presentar datos demográficos, económicos y sociales de un momento específico, que incluye a todas las personas de un país o territorio determinado.

El objetivo primordial, es cumplir con ciertas necesidades de información estadística relacionadas con todos los habitantes de un país, con el fin de planificar diversos aspectos como la provisión de alimentos, la cantidad de escuelas y hospitales necesarios, así como su ubicación, entre otro.

1.9 REGISTROS

Son métodos empleados para estudiar los cambios y la comparación de:

- Variaciones en el tamaño de la población.
- Aumento de la población debido a nacimientos.
- Disminución de la población por emigración.
- Reducción de la población por fallecimiento.

A través de los sistemas de registros, se implementa un proceso continuo que monitorea de forma ininterrumpida los movimientos y cambios en la población.

1.10 POBLACIÓN

Es el conjunto de elementos, individuos, objetos o eventos que comparten una característica común y son objeto de estudio en una investigación o análisis estadístico

Ejemplo 2.

- Número de pacientes que presentan enfermedad del Coronavirus en el Perú en el 2022.
- Edades de los alumnos Universitarios del Perú en el semestre 2022-I
- Tipo de virus que producen enfermedades respiratorias agudas en los pacientes del centro de salud de Morro Solar - Jaén en el año 2020.
- Usuarios que han adquirido la instalación de equipos en el distrito de Jaén en el año 2022.

La población se clasificación en:

a. Población finita. Consiste en un número limitado de elementos, se denota por “**N**”.

Ejemplo 3.

Nivel de comprensión sobre enfermedades de transmisión sexual en todos los estudiantes actualmente matriculados en las universidades del Perú.

b. Población infinita. Es aquella que carece de límites o restricciones

Ejemplo 4.

La calidad de todas las unidades producida mediante un proceso de fabricación.

PARÁMETRO

Es una métrica descriptiva que se aplica a las observaciones de interés para el investigador en una población. Estas características son cuantificables dentro de la población y requieren el uso de datos poblacionales para su determinación. Los indicadores más comunes son:

- Media poblacional (μ).
- Proporción poblacional (P).
- Desviación típica poblacional (σ).

Ejemplo 5.

- El ingreso promedio de todos los asalariados de salud en el Perú (μ).
- La proporción de la proporción total de las plantas manufactureras (P).
- La variabilidad que hay en la producción total de ingreso familiar de la ciudad de Jaén (σ).

1.11 MUESTRA

Es una parte representativa o un subconjunto seleccionado de una población, se denota por “**n”**

Ejemplo 6:

- Pacientes que llegan al servicio de Urgencias al Hospital General de Jaén en el primer trimestre del 2023.
- Alumnos universitarios del Perú cuyas edades se encuentran entre 16 a 18 años.
- Producción de oro en la empresa Antamina. entre los meses de agosto a noviembre del año 2022.

- Estudio de una muestra aleatoria de 150 estudiantes de la UNJ según nivel socioeconómico.

ESTADÍSTICO

Es un valor resumido que proporciona información sobre una característica particular de una muestra, y se utiliza como una estimación del parámetro correspondiente de la población. Los más usados son:

- Media muestral ($\bar{x}$).
- Proporción muestral (p).
- Desviación típica muestral ($s_{\bar{x}}$).

Ejemplo 7.

- Número promedio de pacientes con TBC en el departamento de Cajamarca ($\bar{x}$).
- La proporción de los valores del colesterol HDL de los habitantes del Perú (p).
- La variabilidad de los niveles séricos de plomo encontrados en los niños del departamento de Cajamarca ($s_{\bar{x}}$).

1.12 VARIABLE

Es una característica o atributo que puede tomar diferentes valores y puede ser medida u observada en un estudio. En el contexto de la estadística y la investigación, las variables representan las características que se están estudiando y pueden ser de diferentes tipos, como variables cualitativas (categóricas) o variables cuantitativas (numéricas)

Ejemplo 8.

- Edades
- Raza
- Diámetro de los árboles
- Número de inspecciones de calidad
- Nivel de glucosa en sangre
- Calificaciones
- Millas recorridas, etc.
- Sexo
- Saturación de oxígeno
- Volumen de las Precipitación
- Número de órdenes de producción
- Tamaño de una lesión
- Ingreso Mensuales
- Peso de los expedientes

Clasificación de las variables

1.12.1 SEGÚN LA NATURALEZA DE LA VARIABLE

- **Variable cualitativa. -** Es una característica que describe cualidades o atributos de una población.

  Ejemplo 9.

  - Lesión intraepitelial de cuello uterino
  - Campo laboral de los Biólogos
  - Grado de instrucción
  - Lugar de nacimiento
  - Nacionalidad
  - Estado civil
  - Factor RH
  - Presión arterial

- **Variable cuantitativa.** Cuando el valor de la variable se expresa en una cantidad.

  Ejemplo 10.

  - Números de hijos
  - Temperatura corporal
  - Horas extras de los trabajadores
  - Número de ventas del centro de producción
  - Tiempo de respuesta a un tratamiento
  - Número de empleados
  - Tasa de mortalidad
  - Presión arterial de los pacientes (mm/Hg)
  - Calidad de los productos alimenticios
  - Nivel de glucosa en sangre (mg/dL)

La variable cuantitativa se clasifica en:

a. **Variable cuantitativa discreta.** Se refiere a aquellas variables que resultan de un proceso de conteo y generalmente toman valores enteros.

   Ejemplo 11.

   - Número de compras
   - Número de motores arreglados.
   - Número de cheques.
   - Número de vacunas rechazadas.
   - Número de productos alimenticios caducados
   - Número de horas extras

b. **Variables cuantitativas continuas.** Se obtienen al medir una característica numérica y puede asumir cualquier valor dentro de un intervalo.

Ejemplo 12.

- Producción de una empresa
- La temperatura de los productos alimenticios
- Egresos de una empresa
- Colesterolemia
- Índice de masa corporal (IMC)
- Ingresos mensuales

1.12.2 SEGÚN SU ESCALA DE MEDICIÓN

❖ **Variables nominales.** Son aquellas que clasifican los elementos en distintas categorías sin establecer un orden específico entre ellas, y estas categorías son mutuamente excluyentes

Ejemplo 13.

- Estilos de vida
- Formas de pagos
- Tipos de minerales
- Clases de parásitos
- Estado de salud del paciente)
- Género del paciente

❖ **Variables ordinales.** Aquellas que clasifican objetos individuales, en categorías organizadas con el fin de establecer comparaciones. Es decir, son susceptibles de ordenación, pero no de medición cuantitativas. La distancia entre cada una de los valores es muy difícil de determinar.

Ejemplo 14.

- Riesgo cardiovascular
- Nivel de ventas
- Grado de instrucción
- Estadios clínicos de los tumores
- Medidas de glucosa
- Nivel de dolor percibido

❖ **Variables de intervalo.** Son aquellas que representan mediciones numéricas en las que las diferencias entre los valores tienen un significado específico y constante. En estas variables, el cero no indica la ausencia de la cantidad medida, sino simplemente un punto de referencia

Ejemplo 15.

- Cociente de inteligencia
- Puntuación en una escala de calificación
- Temperatura (Agua congelada (°C) (32°F))

❖ **Variable de razón.** abarcan todas las características de los tipos anteriores, incluyendo la categorización, orden, distancia y un punto de origen común. Los valores de estas variables se expresan como números reales.

Ejemplo 16.

- Edad
- Pesos
- Ingreso
- Producción anual
- Estatura
- Cantidad de accidentes

-

1.13 UNIDAD DE ANÁLISIS

Elemento u objeto indivisible de la población que será examinado, proporcionando datos concretos.

Ejemplo 17.

- Se quiere estudiar la demanda de un champú X en la ciudad de Jaén. La unidad de análisis serán las familias.
- En una encuesta de presupuestos familiares la unidad de análisis puede ser la familia o las familias, de cierto nivel de ingresos.
- Un usuario que ha adquirido la instalación de equipos en el distrito de Jaén en el año 2022.

1.14 UNIDAD DE MUESTREO

Es una unidad seleccionada del marco muestral, se define como un elemento o una entidad individual dentro de la población que es seleccionada para formar parte de la muestra en un estudio de investigación

Ejemplo 18.

- Un estudio sobre la efectividad de una vacuna en una comunidad. La unidad de muestreo son los pacientes vacunados.
- Una corporación realiza estudios sobre la calidad de infraestructura (como puentes, carreteras o edificios), por lo tanto, el proyecto completo es la unidad de muestreo.

1.15 MARCO MUESTRAL

Es un registro físico (como una lista o un croquis) que enumera individualmente las unidades de muestreo, permitiendo la selección aleatoria de las unidades que formarán la muestra. Este documento es esencial en cualquier diseño de muestreo

1.16 DISEÑO MUESTRAL

El diseño muestral tiene 2 aspectos:

- **El proceso de selección**: Consiste en normas y procedimientos a través de los cuales se eligen ciertos elementos de la población para formar parte de la muestra.
- **Un proceso de estimación**: se encarga de calcular las estimaciones de la muestra, asimismo, abarca las tareas de selección y estimación para realizar inferencias.

1.17 DATOS

Notación o son símbolos que describen condiciones, hechos, situaciones o valores.

Ejemplo 19:

x_i: 18 años, 1.55 cm, católica, Alto.

2

ORGANIZACIÓN Y CLASIFICACIÓN DE DATOS

2.1 INTRODUCCIÓN

Frente a un conjunto de datos, el primer paso es organizarlos y clasificarlos según criterios relevantes, lo que nos permitirá visualizar sus características y obtener conclusiones, ya sea directamente o mediante cálculos adicionales. Este proceso comprende los siguientes pasos:

- Revisión y corrección de los datos.
- Construcción de tablas de frecuencias.
- Presentación de los datos en forma de tablas estadísticas y gráficos para una mejor comprensión visual

2.2 REVISIÓN Y CORRECCIÓN DE DATOS

Es un paso imprescindible antes de proceder a la clasificación y procesamiento de la información.

Una regla empírica ampliamente contrastada (Huber 1 984) es esperar entre un 2 y un 5 % de observaciones con errores de medición, trascripción, etc. Entonces antes de utilizar los datos muestrales conviene aplicar técnicas simples para probarlos, como dar respuestas a las preguntas:

a. ¿La conclusión es coherente? ¿Se han extraído conclusiones que no estén sustentado por los datos?
b. ¿Cuántas observaciones hay? ¿Representan adecuadamente a todos los grupos que desean estudiar?

2.3 DISTRIBUCIÓN DE FRECUENCIAS

Es una tabla que muestra cómo se distribuyen los datos en diferentes categorías, modalidades o intervalos. Su objetivo es resumir un conjunto de datos para visualizar cuántas veces ocurren ciertos valores o rangos de valores en un conjunto de observaciones.

Existen dos tipos comunes de distribuciones de frecuencias:

2.3.1 FRECUENCIA ABSOLUTA

Indica el número de veces que ocurre cada valor o intervalo de valores en el conjunto de datos.

Se clasifica en:

- **Frecuencia absoluta simple. -** cantidad de veces que se repite un valor específico de una variable. Se representa por f_i y cumple con lo siguiente:

$$f_i \in Z^+ \Rightarrow 0 \leq f_i \leq n \Rightarrow \sum_{i=1}^{n} f_i$$

- **Frecuencia absoluta acumulada. -** suma de las frecuencias absolutas de todos los valores de la variable. La última frecuencia es igual al número de observaciones. Denotamos por F_i

  Donde:

$$F_k = f_1 + f_2 + f_3 + ... + f_k$$

$$F_k = \sum_{j=1}^{i} f_i$$

Observación

$$i.\ F_1 = f_1$$
$$F_2 = f_1 + f_2 \Rightarrow f_1 \neq f_2 \neq 0$$

$$ii.\ F_1 < F_2 \Rightarrow f_1\, v\, f_2 \neq 0$$
$$\therefore F_1 \leq F_2 \leq F_3 \leq ... \leq F_K$$

También debe cumplir:

$$f_1 = F_1 \leq F_2 \leq F_3 \leq \cdots \leq F_k = n$$

2.3.2 FRECUENCIA RELATIVA

Muestra la proporción o porcentaje de ocurrencias de cada valor o intervalo respecto al total de datos.

Se clasifica en:

- **Frecuencia relativa simple. -** Es el cociente de la frecuencia absoluta y el número total de observaciones. Se denota por h_i :

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

Las frecuencias relativas deben cumplir:

i. $h_i \geq 0$

ii. $0 \leq h_i \leq 1$

iii. $\sum_{i=1}^{n} h_i = 1$

- **Frecuencia relativa acumulada. -** Es la suma de los diferentes valores de la frecuencia relativa. La última frecuencia relativa acumulada es igual a uno. Se denota por H_i

Cumple:

$h_1 = H_i \leq H_2 \leq H_3 \leq \ldots \leq H_k = 1.00$

2.4 DISTRIBUCIÓN DE FRECUENCIAS PARA VARIABLES CUALITATIVAS

Es una tabla que organiza y presenta los datos categóricos o cualitativos, mostrando la frecuencia con la que ocurre cada categoría en el conjunto de datos.

Las partes de una tabla son:

- **Número de la tabla:** Es el código que facilita la localización de la tabla dentro de un documento.
- **Título**: Describe el contenido o propósito de la tabla de forma clara y concisa. Se ubica generalmente en la parte superior.
- **Encabezado.** Es la descripción de filas y columnas de la tabla estadística. Se ubica en la parte superior del cuerpo. Señala las variables con sus categorías o intervalos.
- **Contenido de la tabla:** Se encuentran los datos relacionados con las variables mencionadas en el encabezado. Muestra cómo están distribuidos los datos según la clasificación en las categorías de las variables.
- **Notas al pie (opcional)**: Información adicional que se encuentra generalmente en la parte inferior de la tabla. Explican detalles importantes o aclaraciones sobre los datos presentados.
- **Fuente de los datos** (opcional): Se incluye cuando es necesario mencionar el origen de los datos utilizados en la tabla.

Tabla 1

Distribución de la variable cualitativa

Categoría de variable (X_i)	f_i	F_i	h_i	H_i	$\%h_i$	$\%H_i$
C_1	f_1	F_1	h_1	H_1	$\%h_1$	$\%H_1$
C_2	f_2	F_2	h_2	H_2	$\%h_2$	$\%H_2$
C_3	f_3	F_3	h_3	H_3	$\%h_3$	$\%H_3$
...	...	...	...	...	...	...
C_k	f_k	F_k	h_k	H_k	$\%h_k$	$\%H_k$
Total	**N**	**-**	**1.00**	**-**	**100%**	**-**

2.4.1 GRÁFICAS PARA VARIABLES CUALITATIVAS

Las partes de un gráfico son:

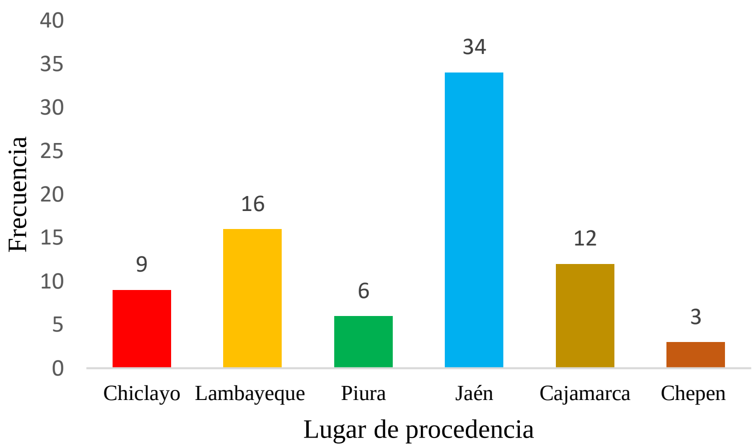
- **Número de la gráfica:** Es el código que facilita la localización de la figura dentro de un documento.
- **Título**: Proporciona una descripción breve y clara del contenido del gráfico. Ayuda a los observadores a entender de qué trata la representación visual.
- **Leyenda**: Explica los colores, símbolos o patrones utilizados en el gráfico para representar diferentes conjuntos de datos o variables.
- **Etiquetas de los ejes**: Describen lo que representan las medidas o categorías en los ejes X e Y.
- **Fuente de los datos**: A **veces** se incluye una nota en la parte inferior indicando de dónde provienen los datos, especialmente en gráficos académicos o informativos.
- **Notas o anotaciones** (opcional): Comentarios que aclaran información específica o destacan ciertos aspectos importantes dentro del gráfico.

2.4.1.1 Grafica de barras

Las categorías de las variables se representan por una barra rectangular vertical (horizontal), donde la altura (largo) es proporcional a su frecuencia.

Figura 5

Trabajadores según lugar de procedencia

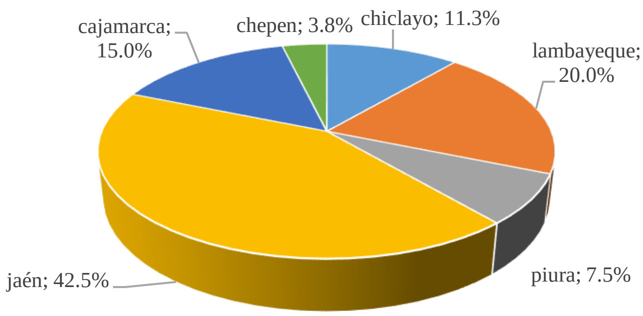


2.4.1.2 Grafica de sectores

Los datos de cada categoría se representan por un sector circular cuyo ángulo es: $s_i = h_i \times 360^0$

Figura 6

Porcentaje de trabajadores según lugar de procedencia



Ejemplo 20:

Estudio socio económico de 20 familias del distrito de La coipa – San Ignacio, 2024. Se asignarán valores: bajo (B), medio (M), alto (A). Los resultados fueron:

M	B	A	B	M	M	B	M	M	M
B	M	B	M	B	B	A	B	A	B

a. ¿Identifica la variable, según su naturaleza?
b. ¿Cuál es su escala de medición?
c. Construir la distribución de frecuencias.
d. Interpretar: $f_1; \%h_1$
e. Realizar el grafico de barras e interpretarlo

Solución

Variable en estudio: **Nivel socioeconómico**

Tipo de variable según su naturaleza: **Cualitativa**

Su escala de medición es: **Ordinal**

Tabla 2

Distribución de las familias según su nivel socio económico, 2024

Nivel socioeconómico	f_i	F_i	h_i	H_i	$\%h_i$	$\%H_i$
Bajo	9	9	0.45	0.45	45%	45%
Medio	8	17	0.40	0.85	40%	85%
Alto	3	20	0.15	1.00	15%	100%
Total	**20**	**-**	**1.00**	**-**	**100%**	**-**

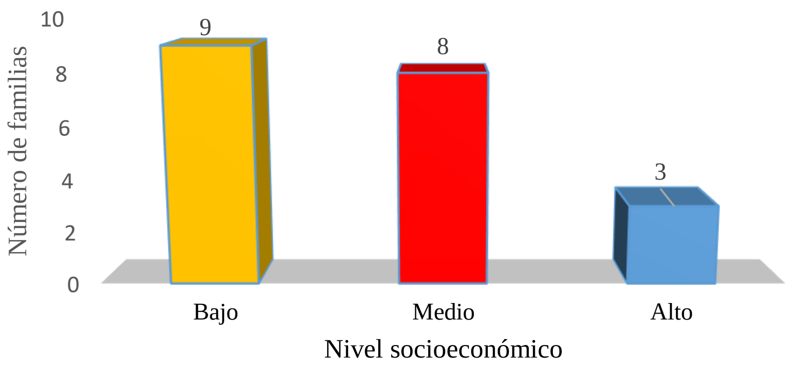
Interpretación

$f_1 = 9$ **:** Significa que hay 9 familias cuyo nivel socio económico es bajo.

$\%h_1 = 45\%$ El 45% de las familias que viven en el distrito de la Coipa tienen el nivel socio económico es bajo

Figura 7

Nivel socioeconómico de 20 familias del distrito de la coipa en el año 2023



.

Comentario: La mayoría de las familias pertenecen a los niveles socioeconómicos bajo y medio, con un número significativamente menor en el nivel alto. Esto indica una predominancia de condiciones socioeconómicas más bajas en la población estudiada.

2.5 DISTRIBUCIÓN DE FRECUENCIAS PARA VARIABLES CUANTITATIVAS DISCRETAS

Organiza los datos de una variable numérica que solo puede tomar un número limitado de valores enteros. Este tipo de distribución se utiliza para variables que se pueden contar, como el número de pacientes, la cantidad de autos en un estacionamiento, número de productos lácteos, etc.

Tabla 3

Distribución de frecuencias de variable discreta

Variable Y_i	f_i	h_i	F_i	H_i	$\%h_i$	$\%H_i$
Y_1	f_1	h_1	F_1	H_1	$\%h_1$	$\%H_1$
Y_2	f_2	h_2	F_2	H_2	$\%h_2$	$\%H_2$
Y_3	f_3	h_3	F_3	H_3	$\%h_3$	$\%H_3$
...	...	...	...	...	...	...
Y_k	f_k	h_k	F_k	H_k	$\%h_k$	$\%H_k$
Total	**N**	**1.00**	**-**	**-**	**100.00**	**-**

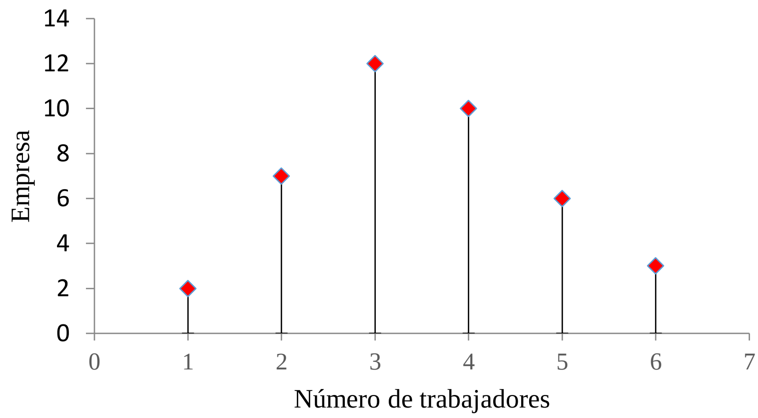
2.5.1 GRÁFICAS PARA VARIABLES CUANTITATIVAS DISCRETAS

2.5.1.1 GRAFICA DE BASTÓN

Se traza en cada valor de la variable un segmento de recta proporcional a su frecuencia.

Figura 8

Número de trabajadores según la empresa

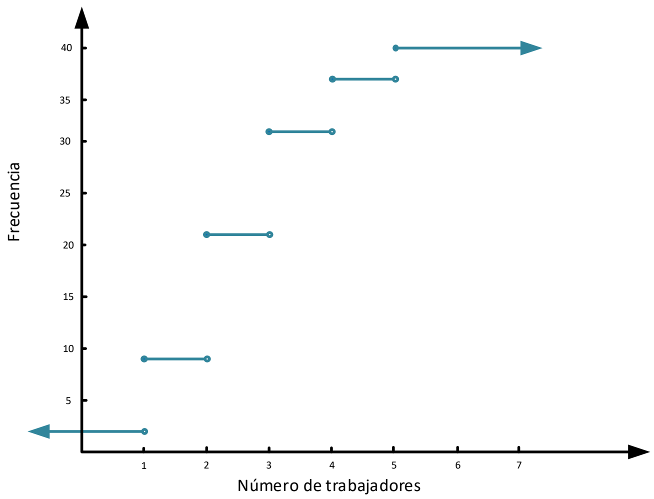


2.5.1.2 GRAFICA DE ESCALERA

Se caracteriza por mostrar cómo una variable aumenta de manera escalonada a lo largo de un eje. Cada cambio en el valor de la variable genera un "escalón" en la gráfica.

Figura 9

Empresas según número de trabajadores



Ejemplo 21.

Una empresa de confección, realiza su pedido de 20 lotes de cuero y cada lote contiene 48 artículos. Se encontró los siguientes números de artículos por lote.

3,	5	1	2	1	3	2	4	4	2
2	6	3	6	1	4	4	3	3	3

a. Indicar el tipo de variable en estudio
b. Construir la tabla de distribución de frecuencias
c. Interpretar f_2, $\%F_4$, $\%h_5$
d. Realizar el gráfico con frecuencias simples
e. Graficar las frecuencias acumuladas
f. ¿Qué porcentaje de lotes tienen entre dos y cuatro artículos defectuosos?

Solución

Variable en estudio: **Número de artículos defectuosos**

Tipo de variable: **Cuantitativa discreta / Ordinal**

Tabla 4

Distribución de 20 lotes de cuero según el número de artículos defectuosos

Artículos Defectuosos (Y_i)	f_i	h_i	F_i	H_i	$\%h_i$	$\%H_i$
1	3	0.15	3	0.15	15	15
2	4	0.20	7	0.35	20	35
3	6	0.30	13	0.65	30	65
4	4	0.20	17	0.85	20	85
5	1	0.05	18	0.90	5	90
6	2	0.10	2	1.00	10	100
Total	**20**	**1.00**	**-**	**-**	**100.0**	**-**

Interpretación:

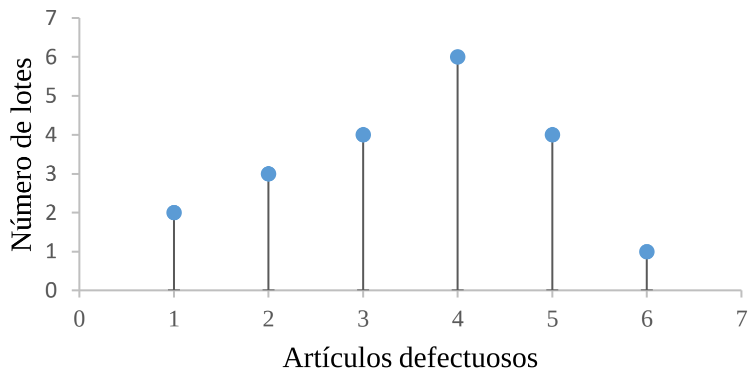
$f_2 = 3$ Hay 3 lotes que tienen 1 artículo defectuoso

$F_4 = 15$ Hay 15 lotes que tienen a lo más 3 artículos defectuosos

$\%h_5 = 20\%$ el 20% de los lotes tienen 4 artículos defectuosos

Figura 10

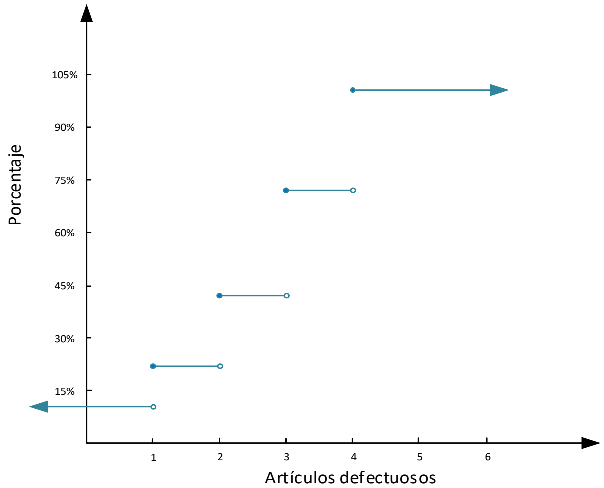
Lotes de cuero según el número de artículos defectuosos



Comentario: La mayor parte de los lotes de cuero presenta entre 2 y 4 artículos defectuosos, siendo 4 el número de artículos defectuosos más habitual, lo cual corresponde al 30% de lotes. De ello podemos inferir que es frecuente encontrar entre 2 y 4 defectos por lote, mientras que es menos común encontrar lotes con una cantidad muy elevada o sin defectos.

Figura 11

Lotes según el número de artículos defectuosos



Comentario: En el grafico podemos observar que la mayoría de los lotes tiende a acumular varios artículos defectuosos, alcanzando hasta 6 artículos defectuosos, representando el 100% del total estudiado.

El porcentaje de lotes que tienen dos o más pero menos de 4 artículos defectuosos es: $20\,\% + 30\,\% = 50\,\%$

2.6 DISTRIBUCIÓN DE FRECUENCIAS PARA VARIABLES CUANTITATIVAS CONTINUAS

Se utiliza cuando la variable estadística es continua o cuando la variable discreta tiene un número grande de valores distintos, organiza y resume un conjunto de datos numéricos en intervalos o clases, mostrando cuántas observaciones caen dentro de cada intervalo.

2.6.1 CONSTRUCCIÓN DE LA DISTRIBUCIÓN DE FRECUENCIAS

Aquí están los pasos básicos para crear esta distribución:

- Establecer el rango (R):

  $R = X_{max} - X_{min}$

- Determinar el número de intervalos, "m", empleando la regla de Sturges

  $m = 1 + 3.33\, log(n)$

  Redondear el resultado al número entero superior.

- Calcular la amplitud interválica "C":

  $C = \frac{R}{m}$

- Determinar los límites de los intervalos.

  I_1 = [Xmin , Xmin + C)

  I_2 = [Xmin + C , Xmin + 2C)

  $\vdots$

  I_k = [Xmin +(K-1) C , Xmin + KC)

- La marca de clase: Se simboliza por "Y_i". Se define como el punto medio del intervalo

$$Y_i = \frac{L_i + L_s}{2}$$

Tabla 5

Distribución de frecuencias de variable continua

$[l_i - l_s>$	Conteo	Y_i	f_i	h_i	F_i	H_i	$\%h_i$	$\%H_i$	$^*F_i$	$^*H_i$
$l_1 - l_2$	$\vert\vert\vert$.. $\vert\vert\vert$..	Y_1	f_1	h_1	F_1	H_1	$\%h_1$	$\%H_1$	$^*F_1$	$^*H_1$
$l_2 - l_3$	$\vert\vert\vert$..	Y_2	f_2	h_2	F_2	H_2	$\%h_2$	$\%H_2$	$^*F_2$	$^*H_2$
$l_3 - l_4$	$\vert\vert\vert$..	Y_3	f_3	h_3	F_3	H_3	$\%h_3$	$\%H_3$	$^*F_3$	$^*H_3$
…	…	…	…	…	…	…	…	…	…	…
$l_k - l_k$	$\vert\vert\vert$..	Y_k	f_k	h_k	F_k	H_k	$\%h_k$	$\%H_k$	$^*F_k$	$^*H_k$
Total			**n**	**1.00**			**100%**			

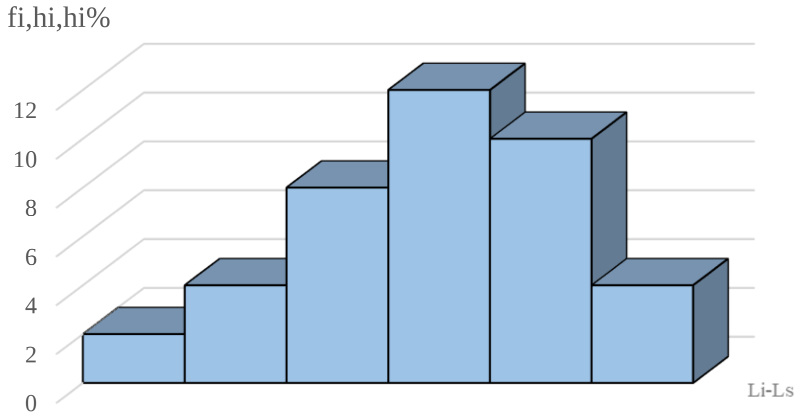
2.6.2 GRÁFICA DE LA DISTRIBUCIÓN POR INTERVALOS

2.6.2.1 HISTOGRAMA

Gráfico compuesto por barras rectangulares verticales adyacentes, en el cual la base de cada barra es proporcional a la amplitud del intervalo y la altura representa su frecuencia.

Figura 12

Gráfico de histograma

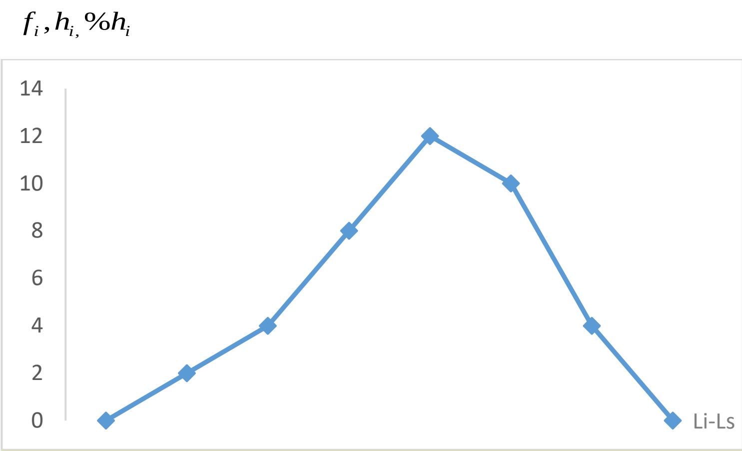


2.6.2.2 POLÍGONO

Se crea conectando con los segmentos de línea los puntos que tienen, como abscisas las marcas de clase y como ordenadas, las frecuencias respectivas. Las marcas de clase adyacentes de frecuencia cero se cierra en ambos extremos.

Figura 13

Gráfico de polígonos

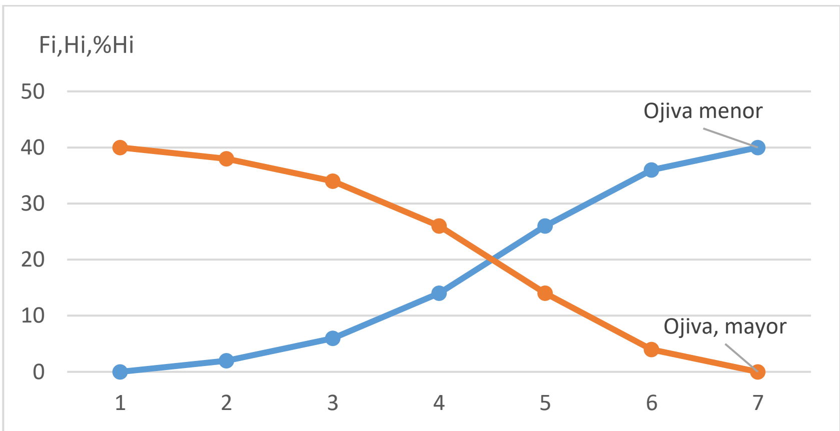


2.6.2.3 OJIVAS

Se obtiene al unir con segmentos de recta los puntos donde la abscisa es proporcional al límite superior de cada intervalo y la ordenada es proporcional a la frecuencia acumulada correspondiente, ya sea absoluta, relativa o en porcentaje

Figura 14

Gráfico de ojivas



Ejemplo 22.

Se tiene información de las presiones sanguíneas (mmHg) de 20 adolescentes moderadamente obesos. El dato se obtuvo de los pacientes del Hospital General de Jaén.

143	128	113	98	142	127	112	97	141	126
111	96	140	125	110	95	133	118	103	89
132	117	102	89	131	116	101	89	130	115

a. Indicar la variable y su tipo

b. Construir la tabla

c. Interpretar $f_{1;}$ % h_5

d. ¿Qué porcentaje de adolescentes tienen una presión sanguínea de 107 a más, pero menos de 134?

e. ¿Cuántos adolescentes tienen una presión sanguínea de 120 a más?

Solución

V.E.: Presión sanguínea

T.V.: Cuantitativa continua

Rango

$R = 143 - 89 = 54$

Número de intervalos

m = 1 + 3.33 log (30) = 5.92

El número de intervalos necesariamente debe ser un número entero, por ello, se debe redondear al entero superior

m = 6

Amplitud interválica

$$C = \frac{R}{m} = \frac{54}{6} = 9$$

Tabla 6

Distribución de las presiones sanguíneas de los adolescentes moderadamente obesos

$[L_I - L_S)$	Y_i	f_i	F_i	h_i	H_i	$\%h_i$	$\%H_i$
89 - 98	93.5	6	6	0,200	0.200	20%	20.0%
98 - 107	102.5	4	10	0.133	0.333	13.3%	33.3%
107 - 116	111.5	5	15	0.167	0.500	16.7%	50.0%
116 - 125	120.5	3	18	0.100	0.600	10.0%	60.0%
125 - 134	129.5	8	26	0.267	0.867	26.7%	86.7%
134 - 143	138.5	4	30	0.133	1.00	13.3%	100.0%
Total	**n**	**30**	**-**	**1.000**	**-**	**100.0**	**-**

Interpretación:

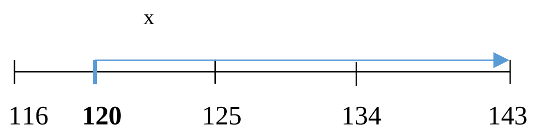
$f_i = 6$: Podemos indicar que hay un adolescente moderadamente obeso y su presión sanguínea es de 89 a más, pero menos de 98 mmHg

$\%h_5 = 26.7\%$: el 26.7% de los adolescentes que muestran una obesidad moderada presentan una presión sanguínea de 125 a más, pero menos de 134 mmHg

- El porcentaje de adolescentes tienen una presión sanguínea de 107 a más, pero menos de 134 mmHg es de 53.4%

- Como se observa en la tabla no tiene ningún intervalo cuyo límite superior sea 120 mmHg,

$\therefore$ Se ubica en el cuarto intervalo $(I_4 = 116; 125)$

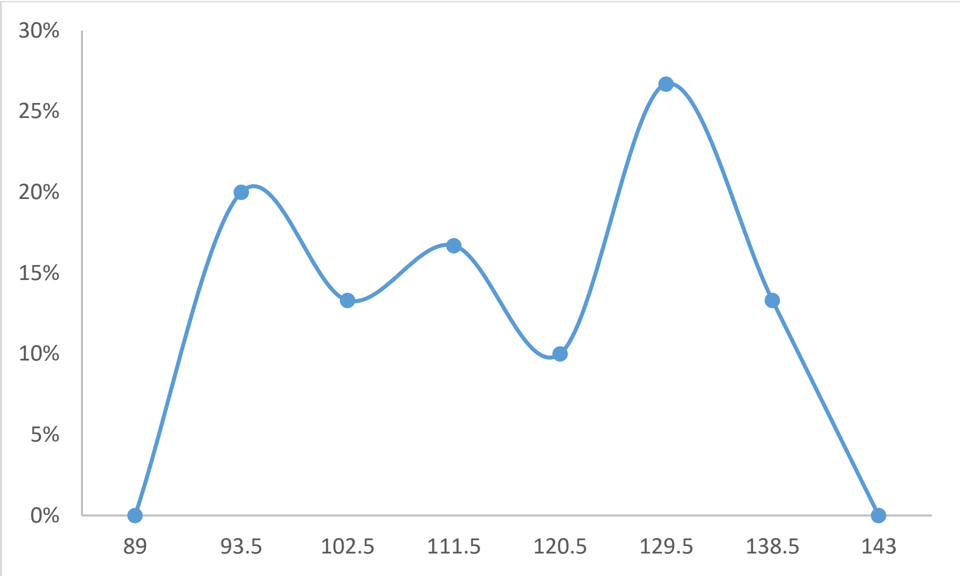


$\frac{125-120}{125-116} \times (3) = 1.667 \cong 2$ adolescentes

$\Rightarrow$ X = 2+8+4=14 adolescentes

Figura 15

Distribución de las presiones sanguíneas de los adolescentes moderadamente obesos



Comentario: En la figura se observa que las presiones sanguíneas de los adolescentes moderadamente obesos se concentran principalmente en torno a los 93.5 mmHg **y** 129.5 mmHg**,** con un mayor pico alrededor del segundo valor. Asimismo, se afirma que la cantidad de adolescentes con presiones fuera de este rango es baja, lo que indica que la mayoría se sitúa entre estos valores.

2.6.3 LÍMITES REALES DE CLASES

El límite real de un intervalo de clase es igual al valor aparente más o menos la mitad de la unidad de medida utilizada.

Ejemplo 23.

Las cuotas anuales de 30 trabajadores que aportaron al banco de Crédito para su seguro de vida (en dólares) teniendo los siguientes intervalos:

60 – 70 70 – 80 80 – 90 90 – 100 100 – 110 110 – 120

Determine los límites de clase.

Solución

Los límites reales son:

- 59.5 - 70.5
- 90.5 - 100.5

- 70.5 - 80.5
- 80.5 - 90.5
- 100.5 - 110.5
- 110.5 - 120.5

2.6.4 SIMETRÍA DE UNA TABLA DE DISTRIBUCIÓN DE FRECUENCIAS

- Si la tabla de distribución de frecuencias tiene un número impar (m = 2 k-1) de clases, se considera simétrica, si las dos clases centrales tienen frecuencias iguales.

$$n_{k-j} = n_{k+j} \quad , \ 1 \leq j \leq k-1$$

- Si la tabla de distribución de frecuencias tiene un número par (m = 2K) de clases, se considera simétrica, si las dos clases centrales tienen frecuencias iguales y las clases equidistantes de estas centrales tienen también frecuencias iguales.

Ejemplo 24.

La cantidad diaria de arroz registrada durante ciento noventa días en un supermercado, se tabuló en una distribución de frecuencias simétrica de cinco intervalos de amplitud iguales a 4. Si la marca de Clase del intervalo central es 12 y la curva de frecuencias absolutas cumple con la relación:

$$f(x) = -(x - 12)^2 + 70$$

Reconstruir la distribución y graficar su histograma.

Solución:

Recopilando los datos se tiene:

- ❖ $n = 190 \qquad C = 4$
- ❖ La distribución es Simétrica:

$f_1 = f_5 \qquad f_2 = f_4 \qquad f_3 = ?$

La marca de clase central es igual a doce, entonces: $Y_3 = 12$

Pero: $Y_3 = Y_2 + C$

De ello se tiene:

$Y_2 = Y_3 + C$

$Y_2 = 12 - 4 = 8$

$Y_1 = 8 - 4 = 4$

$Y_4 = 12 + 4 = 16$

$Y_5 = 16 + 4 = 20$

Se sabe que: $Y_1 = \frac{Y_o + Y_1^{'}}{2}$

Pero:

$Y_1^{'} = Y_0 + C$

$Y_1^{'} = Y_0 + C + Y_0$

$2Y_0 + 4 = 8$

$Y_0 = 4$

- Para trabajar con la ecuación se debe tomar la marca de clase como "X" y de esto se tiene:

$f(x) = -(x - 12)^2 + 70$

$f(12) = -(12 - 12)^2 + 70$

$f(12) = 70$

Esto está representando el $f_3 = 70$

$f(x) = -(x - 12)^2 + 70$

$f(4) = -(4 - 12)^2 + 70$

$f(4) = 6$

Esto está representando el $f_1 = 6$

Por simetría el $f_6 = 6$

$f(x) = -(x - 12)^2 + 70$

$f(8) = -(8 - 12)^2 + 70 = 54$

Esto está representando el $f_2 = 54$

Por simetría el $f_2 = 54$

- Trabajando el cuadro se tiene lo siguiente:

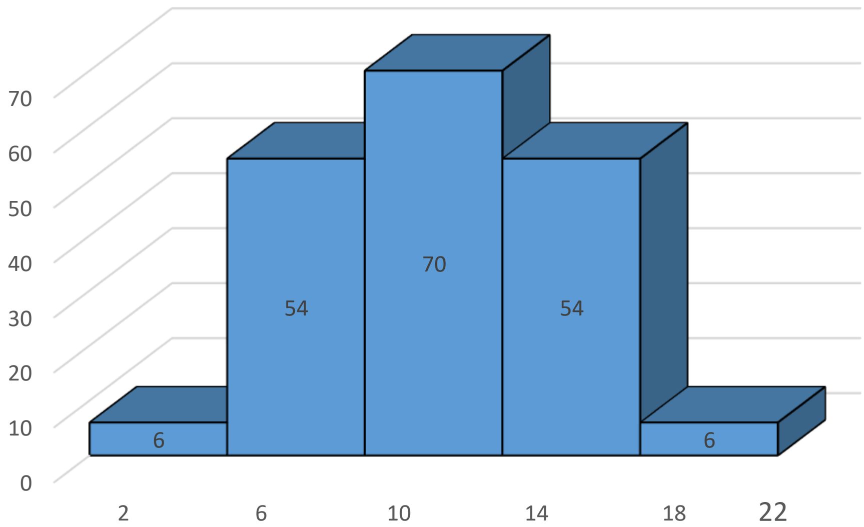
Tabla 7

La demanda diaria de azúcar (en decenas de kilos)

$[L_i - L_s)$	Y_i	f_i
2 – 6	4	6
6 – 10	8	54
10 – 14	12	70
14 – 18	16	54
18 – 22	20	6
Total	-	**190**

Figura 16

Distribución de la demanda diaria de azúcar (en decenas de kilos)



Comentario: En la figura se observa que la demanda diaria de azúcar sigue una distribución aproximadamente simétrica, con un pico en la demanda de entre 100 y 140 kilos diarios. Las frecuencias disminuyen tanto en los rangos más bajos como en los más altos.

2.6.5 DIAGRAMA DE HOJAS Y TALLOS

Un procedimiento semi gráfico (tabular y gráfico) de presentar la información para datos cuantitativos, que es especialmente útil cuando el número total de observaciones es pequeño (menor que 50), es el diagrama de hojas y tallos de TUKEY. Los principios básicos para construirlo son:

a. Redondear los datos de dos o tres cifras significativas y expresar en unidades convenientes.

b. Organizar los datos en una tabla con dos columnas separadas:

- Para los datos numéricos, ubicar a la izquierda los dígitos de las decenas, que constituyen el tallo y a la derecha los dígitos de las unidades, que forman las hojas. Por ejemplo 62 se escribe:

Tallo	Hoja
6	2

- Para datos estén constituidos por tres dígitos, en el tallo se escribirán únicamente centenas y decenas hacia la izquierda. Por ejemplo 259 se escribirá:

Tallo	Hoja
25	9

c. Cada tallo representa una clase y se escribe por única vez, mientras que el número de “hojas” denota la frecuencia de dicha clase y se ubica en una tercera columna del diagrama.

3 ANÁLISIS DESCRIPTIVO DE DATOS CUANTITATIVOS

3.1 INTRODUCCIÓN

Los datos presentados en una distribución de frecuencias resaltan aspectos importantes como la marca de clase, el punto medio y la forma de la distribución (si es asimétrica o simétrica). Sin embargo, para obtener una descripción más precisa de los datos, se deben calcular indicadores de resumen, que son medidas descriptivas que se centran en aspectos como la centralización, la posición, la dispersión o variabilidad, así como la simetría y la curtosis de los datos.

Los valores descriptivos derivados de una muestra, representados por ($x_{1,} x_2, x_3 \ldots x_n$); se conocen como estadísticos o estadígrafos. Estos se dividen en cuatro categorías: de posición, dispersión, concentración y forma. Los estadísticos de posición describen la ubicación de la distribución de frecuencias en relación con un valor particular de la variable. Se pueden distinguir dos tipos: estadísticos de tendencia central y de localización.

Los estadísticos de tendencia central se denominan así porque tienden a ocupar posiciones intermedias entre el valor más bajo y el más alto del conjunto de datos. Los más utilizados son la mediana y la media aritmética, geométrica, armónica, cuadrática. Los de localización señalan valores frecuentes o extremos. Los usados son: moda y cuartiles.

Los estadísticos de dispersión muestran el grado en que los datos están dispersos; un valor más alto indica mayor dispersión de las observaciones. Estos estadísticos reflejan cómo se agrupan los valores del conjunto de datos alrededor del promedio. El más relevante de estos es la varianza, y otros relacionados incluyen la desviación estándar y el coeficiente de variación

Las características de forma, la forma de la curva (o polígono) de distribución de frecuencias y en especial su simetría o asimetría (no tiene simetría) y forma más o menos aplastada o en punta.

3.2 ESTADÍGRAFO DE POSICIÓN

3.2.1 PROMEDIO O MEDIA ARITMÉTICA

Se calcula sumando los valores observados de la variable, dividida por el número total de observaciones, se denota por $\bar{x}$

3.2.1.1 CÁLCULO DE LA MEDIA ARITMÉTICA

❖ **Para datos no tabulados**

El promedio para datos no tabulados se calcula de la siguiente manera:

$$\overline{x} = \sum_{i=1}^{n} \frac{x_i}{n}$$

Ejemplo 25.

Los datos proporcionados sobre el número de pacientes en 10 centros hospitalarios de Cajamarca, son: 3, 2, 1, 3, 1, 4, 2, 0, 4, 3. Calcular e interpretar la media aritmética.

Solución

$$\overline{x} = \sum_{i=1}^{10} \frac{x_i}{n} = \frac{3+2+1+3+1+4+2+0+4+3}{10} = 2.3$$

$$\overline{x} \approx 2\,pac.$$

Interpretación: El número promedio de pacientes en 10 centros hospitalarios de Cajamarca es dos.

❖ **Para datos tabulados**

a. Variable cuantitativa discreta.

Si “n” de una variable estadística discreta, x se clasifican en n valores de distintos o frecuencias absolutas $f_1 + f_2 + f_3, ..., f_k$. Entonces el promedio aritmético se calcula con la fórmula:

$$\overline{x} = \sum_{i=1}^{n} \frac{y_i f_i}{n}$$

Ejemplo 26.

Calcular e interpretar la media aritmética del número de exámenes de laboratorio programados por paciente:

2	1	2	4	1	3	2	3	2	4
3	2	1	3	2	3	3	1	2	0

Solución:

Variable: Número de exámenes de laboratorio

Tipo de variable: Cuantitativa discreta/Ordinal

Tabla 8

Distribución del número de compras por familia

y_i	f_i	$y_i f_i$
0	1	0
1	4	4
2	7	14
3	6	18
4	2	8
Total	20	44

.

Calculando su promedio:

$$\overline{x} = \sum_{i=1}^{n} \frac{y_i f_i}{n} = \frac{44}{20} = 2.22 \cong 2 exámenes$$

Interpretación: El número promedio de exámenes clínicos de laboratorio programados es dos.

b. Variable cuantitativa continua.

Si "n" valores de una variable z están tabulados en una distribución de frecuencias de "m" intervalos donde $m_1, m_2, m_3 ... m_k$, son la marca de clase $f_1 + f_2 + f_3, ..., f_k$ son las frecuencias absolutas, entonces la media aritmética es.

$$\overline{x} = \sum_{i=1}^{n} \frac{y_i f_i}{n}$$

Ejemplo 27.

Se busca establecer el consumo promedio diario de insumos en 45 fábrica alimentos procesados.

63	64	70	62	51	55	67	53	56	56
89	59	57	26	62	59	61	64	62	43
36	35	62	64	60	60	67	76	63	68
49	78	43	72	71	67	51	44	60	52
61	57	81	73	61					

Solución

Procesando los datos se obtuvo la siguiente tabla:

Tabla 9

Distribución del consumo diario de insumos en 45 fábrica de alimentos procesados

$[L_I - L_S)$	y_i	f_i	$y_i f_i$
26 – 35	30.5	1	30.5
35 – 44	39.5	4	15.8
44 – 53	48.5	5	242.5
53 – 62	57.5	14	805
62 – 71	66.5	14	931
71 – 80	75.5	5	577.5
80 – 89	84.5	2	169
Total		45	2713.5

El promedio del consumo diario de insumos es:

$$\overline{x} = \sum_{i=1}^{n} \frac{y_i f_i}{n} = \frac{2713.5}{45} = 60.3$$

Interpretación: El promedio del consumo diario de los insumos de 45 fábricas de alimentos procesados es 60.

Nota: La media aritmética se calcula utilizando frecuencias relativas, reemplazando $h_i = \frac{f_i}{n}$ .Para datos agrupados, por el intervalo se tiene $\overline{x} = \sum_{i=1}^{n} y_i h_i$

Ejemplo 28.

Calcular la media aritmética de los datos agrupados en la tabla 9 Distribución del consumo diario de insumos en 45 fábrica de alimentos procesados.

Tabla 10

Distribución del consumo diario de insumos en 45 fábrica de alimentos procesados

$[L_I - L_S)$	y_i	f_i	h_i	$y_i h_i$
26 – 35	30.5	1	0.022	0.671
35 – 44	39.5	4	0.090	3.555
44 – 53	48.5	5	0.111	5.383
53 – 62	57.5	14	0.311	17.883
62 – 71	66.5	14	0.311	20.681
71 – 80	75.5	5	0.111	8.580
80 – 89	84.5	2	0.044	3.718
Total		45	1.000	60.274

Promedio del consumo diario de insumos

$\overline{x} = \sum_{i=1}^{n} y_i\, h_i = 60.274 \approx 60$ plántulas

Interpretación: El promedio del consumo diario de los insumos de 45 fábricas de alimentos procesados es 60.

3.2.1.2 PROPIEDADES DE LA MEDIA O PROMEDIO

a. La suma de sus desviaciones respecto a la media es siempre igual a cero:

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right) = 0$$

b. La media de una muestra es igual al promedio ponderado de las sub muestras:

$$\overline{x} = \frac{\overline{x}_1 n_1 + \overline{x}_2 n_2}{n}$$

Ejemplo 29.

Utilizando los datos muestrales, calcular el promedio del incremento anual en el diámetro de los árboles (IDA) con el objetivo de planificar las cosechas futuras y evaluar la productividad del bosque.

$n = 20$ $\quad$ $n_1 = 12$ $\quad$ $n_8 = 8$

$\overline{x}_1 = 1.333\ cm$ $\quad$ $\overline{x}_2 = 2.15\ cm$

Solución

La media aritmética ponderada es:

$$\overline{x} = \frac{(1.333)(12) + 2.15(8)}{20} = 1.66\ cm$$

Interpretación: El promedio del incremento anual en el diámetro de los árboles es de 1,66 cm.

c. La media aritmética de una constante por una variable es igual al producto de la constante por la media aritmética de la variable.

$$M[kx] = kM[x]$$

d. La media aritmética constante más una variable es igual a la constante más la media aritmética de la variable.

$$M[k \pm x] = k \pm M[x]$$

e. La suma de los cuadrados de las desviaciones es mínimo cuando dichas desviaciones se toman con respecto a la media aritmética.

$$\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2 \leq \sum_{i=1}^{n}\left(x_i - a\right)^a$$

f. La media de la suma de dos variables es igual a la suma de las medias de dichas variables.

$$M_{(x_i \pm y_i)} = M_{(xi)} \pm M_{(y_i)} = \overline{x} \pm \overline{y}$$

3.2.1.3 VENTAJAS Y DESVENTAJAS:

- **Ventajas:**
  - Sencilla de entender y ampliamente usada.
  - Se puede obtener cuando no es posible calcular otros tipos de medidas de tendencia central.
  - Se establece en el muestreo.
- **Desventajas:**
  - Puede verse afectado por los valores extremos que no son representativos del resto de las observaciones. De modo que es altamente sensible a cualquier cambio de una distribución de datos.

3.2.1.4 APLICACIONES

- Se utiliza ampliamente en diversas áreas. En educación, se emplea para calcular promedios de calificaciones. En finanzas, ayuda a obtener promedios de precios o rendimientos de inversiones. En salud, puede calcular promedios de indicadores clínicos como presión arterial o niveles de glucosa.
- Se usa en encuestas para promediar opiniones y en análisis estadísticos para describir conjuntos de datos simples.

3.2.2 MEDIA GEOMÉTRICA

La media geométrica de "n" valores positivos $x_1, x_2, x_3, \ldots, x_n$, es el número M_g que se define como la raíz enésima:

$$M_g = \sqrt[n]{x_1 . x_2 . \ldots\, x_n} = \sqrt[n]{\underset{i=1}{\overset{n}{\pi}} x_i}$$

Ejemplo 30:

Supongamos que durante una semana se registra el crecimiento diario de casos de una enfermedad infecciosa en una población. Los incrementos diarios de casos son los siguientes:

x_i: 1.50 1.33 1.25 1.20 1.33 1.25

Solución

La media geométrica es:

$M_g = \sqrt[6]{1.50 \times 1.33 \times 1.25 \times 1.20 \times 1.33 \times 1.25} = 1.307$

Interpretación: La tasa de crecimiento diaria promedio de la enfermedad infecciosa durante esa semana es de 1.31.

3.2.2.1 CÁLCULO DE LA MEDIA GEOMÉTRICA

❖ **Para datos no tabulados:**

La fórmula para calcular la media geométrica cuando las observaciones se presentan en forma:

$$Log[M_g] = \frac{\sum \log x_i}{n}$$

Ejemplo 31.

Dada una producción de algodón se obtuvo los datos (en quintales)

x_i: 22 28 58 32 24 42

Solución

Sacar el logaritmo de cada dato.

log22 = 1.342 log28 = 1.447 log58 = 1.763

log32 = 1.505 log24 = 1.380 log42 = 1.623

Calculando se tiene:

$$Log[M_g] = \frac{1.342 + 1.447 + 1.763 + 1.505 + 1.380 + 1.623}{6} = 1.51$$

Utilizando antilogaritmo se tiene:

$$M_g = 32.359$$

Interpretación

La producción promedio de algodón es de 32.359 quintales.

- **Para datos tabulados:**

Cuando los datos se representan agrupados se presenta la fórmula:

$$Log[M_g] = \frac{\sum f_i \log x_i}{n}$$

Ejemplo 32:

De una población de 120 fábricas se selecciona una muestra de 55 fábricas, analizar su promedio de producción semanal en toneladas.

Tabla 11

Distribución de la población de 120 fábricas

Y_i	f_i	$\log Y_i$	$f_i \log Y_i$
1	5	0	0
2	10	0.301	301
3	15	0.477	7.156
4	12	0.602	7.224
5	13	0.698	9.086
Total	**55**		**26.476**

.

Solución

Calcula los logaritmos de Y_i y multiplicarlo con la frecuencia.

$$Log[M_g] = \frac{26.476}{55} = 0.481$$

Aplicando antilogaritmo se tiene:

$$M_g = 3.0296$$

Interpretación: La producción promedio semanal de las fábricas es de 3.0296 toneladas.

3.2.2.2 VENTAJAS Y DESVENTAJAS

- **Ventajas:**
  - Apropiada para tasas de crecimiento
  - Menos afectada por valores extremos
  - Captura mejor la realidad en datos multiplicativos
  - Aplicable a proporciones y tasas
  - Mejor para variaciones en el tiempo

- **Desventajas:**
  - Esta limitado por valores positivos.
  - Si algún valor de la variable es cero, la media geométrica no puede determinarse.

3.2.2.3 APLICACIONES DE LA MEDIA GEOMÉTRICA

- A pesar de las limitaciones previamente señaladas, la media geométrica encuentra aplicaciones valiosas en ciertas variables, especialmente en aquellas que siguen una tendencia exponencial, como las variables cronológicas. En estos casos, su uso se vuelve fundamental cuando se necesita calcular valores intermedios mediante interpolación lineal.
- También se usa cuando se desea promediar tasas de cambio, proporciones, índices.

Ejemplo 33:

Calcular el promedio del cambio (factor de crecimiento) de la población (en millones de habitantes).

Tabla 12

Promedio del cambio

Años	**Población**	**Cambio** (X_i)
1980	2	-
1990	8	4
2000	128	16

Solución

El cambio promedio de crecimiento, usando la media aritmética simple será:

$$\bar{x} = \frac{4+16}{2} = \frac{20}{2} = 10$$

Según el promedio, el crecimiento de la población será:

Tabla 13

Crecimiento de la población

Años	**Población** $\bar{x}$
1 980	2
1 990	2 x 10 = 20
2 000	20 x 10 =200

Pero el cuadro inicial dado muestra un crecimiento real de sólo 2, 8, 128. Por tanto, el cambio promedio correcto de crecimiento, deberá ser menor que 10. Es decir, en este caso la media aritmética para la tasa de crecimiento es incorrecta. Si se suma la media geométrica para obtener este promedio de cambio, se tiene:

$$M_g = \sqrt[2]{4 \times 16} = 8 \text{ veces cada 10 años}$$

Según el promedio, de crecimiento de la población será:

Tabla 14

Crecimiento de la población

Años	Población $\overline{x}_G$
1 980	6
1 990	6 x 8 = 48
2 000	48 x 8 = 384

Como se observa, la media geométrica da un crecimiento más próximo al crecimiento real. Por lo tanto, la mejor medida para promediar el cambio de crecimiento es la media geométrica.

Ejemplo 34.

Suponga que se depositaron 100 nuevos soles inicialmente y que se acumulan los intereses a tasas variables de 5, 10, 15, -2, 18% anual, durante cinco años. Hallar el factor de crecimiento promedio anual.

Tabla 15

Depósito en cinco años

Años	Tasa de interés (%)	Factor de crecimiento (X_i)	Ahorros al final del año
1 998	5	1.05	100 x1.05 =105.00
1 999	10	1.10	105 x1.10 = 115.50
2 000	15	1.15	115.5 x1.15 =132.825
2 001	-2	0.98	132.825 x 98=130.1685
2 002	18	1.18	130.1685 x1.18 =153.60

Donde la columna llamada "factor de crecimiento" se obtiene con la fórmula:

$$1 + \frac{tasa\ de\ interés}{100}$$

3.2.3 MEDIA ARMÓNICA

La media armónica de n valores no nulos $x_i, x_2, x_3, \ldots x_n$, es el número M_a que se define como el recíproco de la media aritmética de los recíprocos de esos valores.

3.2.3.1 CÁLCULO DE LA MEDIA ARMÓNICA

- **Para datos no tabulados:**

La fórmula para los datos en forma simple se presenta:

$$M_a = \frac{1}{\frac{1}{n}\sum\left(\frac{1}{x_i}\right)}$$

Ejemplo 35.

Se ha comprado 3000 mil antibióticos que cuesta S/. 5.00 el ciento y 3000 mil antibióticos que cuestan S/. 3.00 el ciento ¿Cuál es el promedio porcentual de los antibióticos?

Solución

300 antibióticos $\rightarrow s/.5.00\ el\ \%$

300 antibióticos $\rightarrow s/.3.00\ el\ \%$

$$M_a = \frac{1}{\frac{1}{2}\left(\frac{1}{5}+\frac{1}{3}\right)} = 4$$

Interpretación: El promedio porcentual de antibióticos es de 4.

- **Para datos tabulados**

La media armónica para datos tabulados (media armónica ponderada) se define por:

$$M_a = \frac{1}{\frac{1}{n}\sum\frac{n_i}{x_i}}$$

Ejemplo 36.

Un atleta se dirige de Jaén a Trujillo, ciudades que distan 700Km. Los primeros 300 Km los recorre a 150 Km/h. y el resto a1 100 Km/h. Calcule e interprete la velocidad media del atleta.

Solución

Puesto que la velocidad y el tiempo están en relación inversa, vamos a calcular el promedio de la velocidad con la media armónica.

$$M_a = \frac{1}{\frac{1}{n}\sum\frac{n_i}{x_i}} = \frac{1}{\frac{1}{700}\left(\frac{300}{150}+\frac{500}{100}\right)} = 100Km$$

Interpretación: El promedio del recorrido de la velocidad del atleta es de 100km/h.

3.2.3.2 VENTAJAS *Y* DESVENTAJAS

- **Ventajas:**
  - Adecuada para promediar tasas o ratios
  - Considera la proporcionalidad inversa
  - Útil en ciertos contextos financieros

- **Desventajas:**
  - Sensibilidad a valores muy pequeño
  - Es afectado por los valores extremos, de modo que otorga a los valores pequeños, le da un peso mayor al de las medias aritmética y geométrica.
  - No está definida, si alguno de los valores es cero.

3.2.3.3 APLICACIONES

Se aplica en los casos:

- Cuando se tiene términos, cuyo recíprocos se quiere calcular su media.
- Cuando se presentan en relación inversa entre las variables implícitas, por ejemplo, entre la productividad y el tiempo.

3.2.4 MEDIANA

Valor mediano de una serie de valores observados en el número que supera a la serie de datos ordenados en forma creciente o decreciente en dos partes iguales. Depende de datos ordenados y no de valores que dependen de estos datos.

3.2.4.1 CÁLCULO DE LA MEDIANA

❖ **Para datos no tabulados:**

Para calcular la mediana en datos no tabulados, hay que distinguir dos situaciones:

a. Número impar de observaciones:

Si se toman los datos originados para calcular la mediana lo primero que se debe hacer es ordenar los datos de menor a mayor o de mayor a menor y tomar como valor la posición central.

$$Me = X_{\frac{n+1}{2}}$$

Ejemplo 37.

En un laboratorio se busca eliminar el ruido o valores anómalos en redes de sensores que miden la vibración. Los datos se dan a continuación:

x_i: 100, 30, 3, 10, 25.

Calcular e interpretar la mediana.

Solución

Se ha ordenado a los datos en forma ascendente:

3	10	25	30	100	$n = 5$
x_1	x_2	x_3	x_4	x_5	impar

Realizando el cálculo se tiene:

$$M_e = X_{\frac{5+1}{2}} = X_3 = 25$$

Interpretación: El 50% del ruido o valores anómalos en redes de sensores que miden la vibración es menor o igual a 25, mientras que el otro 50% supera a dicho peso.

Nota: La mediana puede determinarse de manera directa seleccionando el valor central del conjunto de datos.

b. Número par de observaciones:

Si el número de observaciones es par, es la semisuma de los valores centrales.

$$Me = \frac{X_{\frac{n}{2}} + X_{\frac{n+1}{2}}}{2}$$

Ejemplo 38.

Un programa registra los tiempos de falla (min) de los equipos que se encuentran en mantenimiento en el laboratorio de la universidad nacional de Jaén. Los datos se dan a continuación:

x_i: 100, 30, 3, 10, 26, 50.

Calcular e interpretar el valor mediano.

Solución

Ordenando los datos en forma ascendente:

3	10	26	30	50	1000	$n = 6$
x_1	x_2	x_3	x_4	x_5	x_6	par

$$M_e = \frac{x_3 + x_4}{2} = \frac{26 + 30}{2} = 28$$

Interpretación

El 50% de los equipos que se encuentran en mantenimiento en el laboratorio de la universidad nacional de Jaén es menor o igual a 28 min. mientras que el otro 50% supera a dicho tiempo.

❖ **Para datos tabulados**

a. Para variables discretas

- Se construye la tabla de distribución de frecuencias absolutas acumuladas menor que.
- Se determina la menor frecuencia absoluta acumulada F_i que supere a $\frac{n}{2}$; es decir: $\frac{n}{2} < F_i$

  En esta situación puede ocurrir que $\frac{n}{2} \geq F_{i-1}$ o sea que se puede $F_{i-1} \leq \frac{n}{2} < F_i$

i. Cuando $\frac{n}{2} \geq F_{i-1}$ entonces la mediana es: $M_e = Y_i$

ii. Cuando $\frac{n}{2} = F_{i-1}$, entonces y_{i-1}, y_i cualquiera entre estos dos se puede considerar como la mediana. La mediana se calcula:

$$M_e = \frac{y_{i-1} + y_i}{2}$$

Ejemplo 39.

Se revisaron 20 lotes de materia prima, que llegaron a Jaén para abastecer a la empresa "Dennis", cada lote contiene 48 artículos. Para ello se realiza un control de calidad de los artículos defectuosos, encontrando los resultados por lote:

3	4	3	2	1	3	2	4	5	3
2	0	4	0	1	4	3	3	1	2

Solución:

Variable: Artículos defectuosos

Tipo de variable: Cuantitativa discreta/Ordinal

Tabla 16

Artículos defectuosos por lotes de materia prima

	Y_i	f_i	F_i
	0	2	2
	1	3	5
	2	4	9
M_e	3	6	15
	4	4	19
	5	1	20
	Total	**20**	**-**

Para ubicar la mediana se selecciona el total de los datos y se divide entre dos.

$$\frac{n}{2} = \frac{20}{2} = 10$$

Luego ver la condición:

$$F_i \geq \frac{n}{2} \begin{cases} 1)\ \mathrm{F_i} > \dfrac{n}{2} \Rightarrow \mathrm{Me} = \mathrm{Y_i} \\ \quad \text{o} \\ 2)\ F_i = \dfrac{n}{2} \Rightarrow Me = \dfrac{Y_{i-1} + Y_i}{2} \end{cases}$$

Cumple la primera condición:

$M_e = 3$

Interpretación

El 50% de los lotes tienen tres o menos artículos defectuosos mientras que el otro 50% supera a dicho número.

Ejemplo 40.

La Tabla 17 presenta el número de empleados que tienen las empresas de la avenida Mesones Muro de la ciudad de Jaén.

Tabla 17

Número de empleados que tienen las empresas

	Y_i	f_i	F_i
	1	8	8
	2	6	14
Y_i	3	9	23
Y_{i-1}	**4**	8	**31**
	5	5	36
	6	2	38
	7	7	45
	8	5	50
	9	6	56
	10	6	62
	Total	**62**	**-**

Hallar la mediana de la siguiente tabla.

Solución:

Variable: Número de empleados

Tipo de variable: Cuantitativa discreta/Ordinal

Para ubicar la posición de la mediana se selecciona el total de los datos y se divide entre dos.

$$\frac{n}{2}=\frac{62}{2}=31$$

Luego ver la condición:

$$F_i \geq \frac{n}{2}\begin{cases} 1)\ \mathrm{F_i} > \dfrac{n}{2} \Rightarrow \mathrm{Me} = \mathrm{Y_i} \\ \quad \text{o} \\ 2)\ F_i = \dfrac{n}{2} \Rightarrow Me = \dfrac{Y_{i-1}+Y_i}{2} \end{cases}$$

Cumple la segunda condición:

$$M_e = \frac{Y_i + Y_{i-1}}{2} = \frac{4+5}{2} = 4.5$$

Interpretación

El 50% de las empresas de la avenida Mesones Muro de la ciudad de Jaén que pertenecen al régimen general tienen 5 o menos trabajadores mientras que el otro 50% supera a dicho número.

b. Variables continuas:

- Construir la tabla de frecuencias acumuladas "menor que"
- Determinar la menor de las frecuencias absolutas acumuladas

  F_i ; tal que $F_i > \frac{n}{2}$

  En esta situación puede ocurrir que $\frac{n}{2} \geq F_{i-1}$, es decir se puede tener: $F_{i-1} \leq \frac{n}{2} < F_i$ (o $H_{i-1} \leq \frac{1}{2} < H_i$) La clase que contiene a M_e es $LI - LS$

  i. Si $F_{i-1} = \frac{n}{2}$ o $H_{i-1} = \frac{1}{2}$ la mediana es: $M_e = Y'_{i-1}$

  Dónde Y'_{i-1}: Límite superior de la clase mediana

  ii. Si ocurre $\frac{n}{2} \geq F_{i-1}$ o $(H_{i-1} < \frac{1}{2})$ la mediana se debe encontrar dentro de la clase *j*, es decir, en el intervalo $[L_i - L_s)$, llamada clase mediana, por lo tanto, se calcula mediante la ecuación:

$$M_e = Li + C \times \left(\frac{\frac{n}{2} - F_{i-1}}{f_i} \right)$$

Donde:

L_i = Límite inferior de la clase mediana

C = Amplitud interválica

F_{i-1}	=	Frecuencia absoluta acumulada anterior de la clase de la mediana
f_i	=	Frecuencia absoluta de la clase de la mediana.
n	=	Total de observaciones

Ejemplo 41.

Calcular e interpretar la mediana de los datos de la tabla N° 6 (distribución de las presiones sanguíneas de 20 adolescentes moderadamente obesos)

Tabla 18

Distribución de las presiones sanguíneas de 20 adolescentes moderadamente obesos

$[L_I - L_S)$	Y_i	f_i	F_i
89 - 98	93.5	6	6
98 - 107	102.5	4	10
107 - 116	111.5	3	13 $= F_{i-1}$
116 - 125	120.5	5	18 $= F_i$
125 - 134	129.5	8	26
134 - 143	138.5	4	30
Total	**n**	**30**	**-**

Solución

- Para ubicar la mediana se selecciona el total de los datos y se divide entre dos. $\frac{n}{2} = \frac{30}{2} = 15$
- Observar la primera frecuencia acumulada que sea igual o que exceda a 15. En este caso es 18

Luego ver la condición:

$$F_i \geq \frac{n}{2} \begin{cases} 1)\ F_i = \frac{n}{2} \Rightarrow M_e = Y_i' \\ \quad \text{o} \\ 2)\ F_i > \frac{n}{2} \Rightarrow M_e = Y_{i-1}' + C\left[\frac{n/2 - F_{i-1}}{f_i}\right] \end{cases}$$

Cumple la segunda condición:

$$M_e = LI_i + C_i \times \left(\frac{\frac{n}{2} - F_{i-1}}{f_i} \right) = 116 + 9 \times \left(\frac{15-13}{5} \right) = 119.6$$

Interpretación: El 50% de los adolescentes moderadamente obesos presentan como máximo una presión sanguínea de 119.6 mmHg mientras que el otro 50% supera a dicha presión.

Ejemplo 42.

Hallar la mediana de los sueldos (miles de soles) de 50 trabajadores de una empresa de textil del mes de agosto de 2023.

Tabla 19

Sueldos de los trabajadores de una empresa de textil del mes de agosto de 2023

$[Y'_{i-1} \quad Y'_i\rangle$	f_i	F_i	
91.5 – 92.5	7	7	
92.5 – 93.5	10	17	
93.5 – **94.5**	8	**25**	$= F_i$
94.5 – 95.5	10	35	
95.5 – 96.5	15	50	
Total	**50**	**-**	

.

Solución

Para ubicar la mediana se selecciona el total de los datos y se divide entre dos.

$$\frac{n}{2} = \frac{50}{2} = 25$$

Luego ver la condición:

$$F_i \geq \frac{n}{2} \begin{cases} 1)\ F_i = \dfrac{n}{2} \Rightarrow M_e = Y'_i \\ \quad \text{o} \\ 2)\ F_i > \dfrac{n}{2} \Rightarrow M_e = Y'_{i-1} + C\left[\dfrac{n/2 - F_{i-1}}{f_i}\right] \end{cases}$$

Cumple la primera condición:

$M_e = Y_i' = 94.5$

Interpretación: El 50% de los trabajadores perciben el sueldo de 94500 nuevos soles mientras que el otro 50% de los trabajadores superan dicho sueldo.

Nota: Cuando se trabaja las frecuencias relativas acumuladas, la fórmula para obtener la mediana está dada por:

$$\mathrm{Me} = \mathrm{Y}_{\mathrm{i}}' + C\left(\frac{\frac{1}{2} - H_{i-1}}{h_i}\right)$$

3.2.4.2 PROPIEDADES DE LA MEDIANA

- La suma de las desviaciones absolutas de las observaciones con relación a la mediana es mínima con relación a cualquier otro valor de la distribución. En símbolos:

  $$\sum_{i=1}^{n} |Y_i - M_e| = \text{mínimo}, \qquad \text{para datos no tabulados}$$

  $$\sum_{i=1}^{n} f_i |Y_i - M_e| = \text{mínimo}, \qquad \text{para datos tabulados}$$

  Si *a* es cualquier valor, entonces la propiedad se escribe:

  $$\sum_{i=1}^{n} |Y_i - M_e| \le \sum_{i=1}^{n} |Y_i - a| \qquad \text{para datos no tabulados}$$

  $$\sum_{i=1}^{n} f_i |Y_i - M_e| \le \sum_{i=1}^{n} f_i |Y_i - a| \qquad \text{para datos tabulados}$$

- En una distribución simétrica se cumple:

  $M_e = Y'_{\frac{m+1}{2}}$, si m es impar

$$M_e = Y'_{\frac{m}{2}} \qquad \text{, si m es par}$$

Donde:

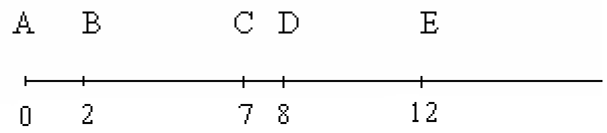
Y_i = Marca de clase

m = número de clase

Y'_i= límite de clase

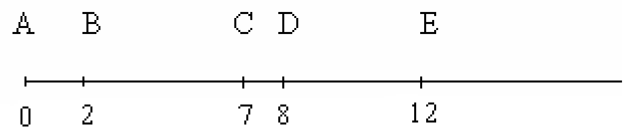
Ejemplo 43.

Cinco amigas viven en Bellavista, como muestra la figura:



¿Cuál de estos puntos deben ser elegidos para un encuentro de modo que el gasto para recogerlas sea mínimo?

Solución



Si m es par $M_e = Y'_{\frac{m}{2}}$ entonces:

$M_e = Y'_{\frac{6}{2}} = Y'_3$ Ocupa el tercer lugar

Interpretación: Entonces el punto de encuentro debe ser "C"

Ejemplo 44.

Hallar la mediana de los sueldos (miles de soles) de 50 trabajadores de una empresa de textil del mes de agosto de 2023. Los datos se presentan en la tabla siguiente

Tabla 20

Sueldos de los trabajadores de una empresa de textil

$[L_I - L_S)$	Y_i	f_i	F_i
91.5 – 92.5	92	8	9
92.5 – 93.5	93	13	21
93.5 – 94.5	94	8	29
94.5 – 95.5	45	13	32
95.5 – 96.5	96	8	40
Total	**-**	**50**	**-**

Solución

- Calcular la frecuencia absoluta acumulada "F_i"
- Ubicar la posición donde se encontrará la mediana $\left(\frac{50}{2}\right) = 25$; en la frecuencia absoluta acumulada y se debe cumplir la siguiente condición.

$F_{i-1} < \frac{n}{2} < F_i$ Donde i determina el intervalo donde se encuentra la M_e

$\Rightarrow$ $21 < 25 < 29$ Donde i=3 determina el intervalo donde se encuentra la M_e

- Calculamos la mediana.

$$M_e = 93.5 + 1 \times \left[\frac{25 - 21}{8}\right] = 94$$

Interpretación: El 50% de los sueldos de trabajadores es menor o igual a 94 mil soles y el otro 50% supera dicho sueldo.

3.2.4.3 VENTAJAS Y DESVENTAJAS

- **Ventajas:**
  - Es fácil de entender y se puede calcular a partir de cualquier clase de datos.
  - Se puede calcular para cualquier tipo de datos cuantitativos.

- Se puede encontrar inclusive en datos cualitativos ordinales.

Ejemplo 45.

La siguiente tabla muestra la distribución de la calidad de la fotocopia de 37 máquinas.

Encontrar la mediana e interpretarlo.

Tabla 21

Distribución de la calidad de la fotocopia de las máquinas

Fotocopia	f_i	F_i
Nítida	10	10
Normal	12	20
Opaca	17	37
Total	**37**	**-**

Solución

Se tiene:

$$\frac{n}{2} = \frac{37}{2} = 18.5$$

Ubicando la mediana en la primera frecuencia absoluta acumulada que supere a 18.5.

$$\Rightarrow \quad M_e = Normal$$

- **Desventajas**
  - Organizar los datos antes de realizar cualquier tipo de cálculo.
  - No es adecuada a manipulaciones algebraicas.

3.2.5 MODA

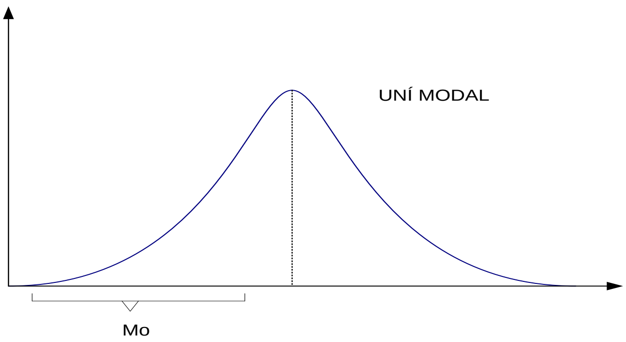
Es una medida de posición, es el valor más frecuente de la variable, es decir, es el valor de la variable que le corresponde la mayor frecuencia simple.

Existen varios tipos de moda según las características del conjunto de datos. Aquí te presento los principales:

- **Unimodal**: Cuando un conjunto de datos tiene una sola moda, es decir, un único valor que se repite con mayor frecuencia.

  Figura 17

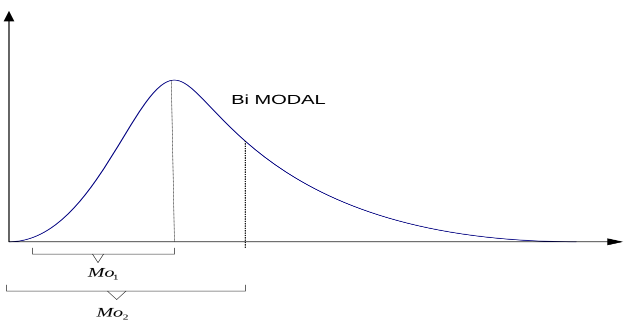
  Uní modal



- **Bimodal**: Se refiere a un conjunto de datos que tiene dos modas, o dos valores que se repiten con la misma mayor frecuencia. Esto ocurre cuando hay dos picos en la distribución de los datos.

  Figura 18

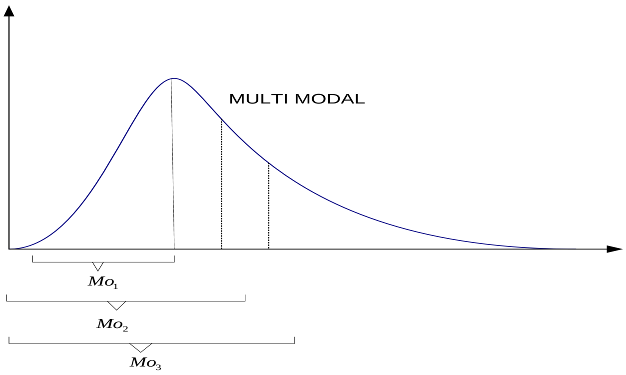
  Bimodal



- **Multimodal**: Ocurre cuando hay tres o más valores que tienen la misma frecuencia máxima. Un conjunto de datos multimodal tiene varias modas que pueden representar diferentes subgrupos dentro de los datos.

Figura 19

Multimodal



- **Amodal**: Cuando no hay ningún valor que se repita, es decir, todos los datos tienen la misma frecuencia, por lo que no existe una moda definida.

3.2.5.1 CÁLCULO DE LA MODA

❖ **Para datos no tabulados:**

La forma como se trabaja para datos no agrupados se presenta a continuación:

Ejemplo 46:

Estudio sobre el número de medicamentos que se administran diariamente por vía oral a 9 pacientes con hipertensión. Determine e interprete la moda.

x_i	2	3	5	2	3	7	3	1	3

Solución

El que se repite con mayor frecuencia es el valor 3

$\Rightarrow \qquad M_o = 3$

Interpretación: El mayor número de medicamentos que se administran por vía oral a los pacientes con hipertensión es 3.

Ejemplo 47.

En un estudio sobre el crecimiento de los árboles en una zona forestal, se toma una muestra para medir el crecimiento de 5 árboles en un año y estos son: 30 cm, 35 cm, 32 cm, 33 cm y 34 cm respectivamente. Hallar e interpretar la moda del crecimiento anual.

Solución

- Observamos que no se repite ninguna medición.

$\therefore$ **La distribución es AMODAL**

Ejemplo 48.

El área de control de calidad analiza diariamente las muestras seleccionadas en el día para asegurar la calidad de los nuevos productos lanzado al mercado en el primer trimestre del 2024. Los datos fueron: 32; 30, 35, 32, 33 y 34; 35. Hallar e interpretar la moda.

Solución

- Observamos que el valor que se repite frecuentemente es 32 Y 35.

$\therefore \mathrm{M}_{o1} = 32$

$\mathrm{M}_{o2} = 35$

Interpretación:

M_{o1} : El mayor tamaño de muestras seleccionadas por día para asegurar la calidad de los nuevos productos es 32

M_{o2} : El mayor tamaño de muestras seleccionadas por día para asegurar la calidad de los nuevos productos es 32.

- **Para datos tabulados**

a. Para variable cuantitativa discreta

Una vez agrupados los datos, es posible determinar el valor modal; bastará con fijar el valor que más se repite en la frecuencia absoluta simple.

Ejemplo 49.

La tabla 22 muestra la distribución de 20 familias del distrito de Jaén según el número de compras por familia.

Tabla 22

Distribución de familias del distrito de Jaén según el número de compras por familia

	(Y_i)	f_i
	0	1
	1	4
M_o	**2**	**7**
	3	6
	4	2
	Total	**20**

Solución

La frecuencia absoluta simple más alta es 7

$\therefore \quad M_o = 2$

Interpretación:

La mayoría de las familias del distrito de Jaén realizan dos compras.

b. Para variable cuantitativa continua:

Se calcula usando la fórmula:

$$M_o = LI_i + C_i \times \left(\frac{(f_i - f_{i-1})}{(f_i - f_{i-1}) + (f_i - f_{i+1})} \right)$$

Donde:

LI_i	:	Límite inferior de la clase modal
C_i	:	Amplitud interválica de la clase modal
f_i	:	Frecuencia absoluta de la clase modal
f_{i-1}	:	Frecuencia absoluta anterior de la clase modal
f_{i+1}	:	Frecuencia absoluta posterior de la clase modal

Ejemplo 50.

Calcular la moda del tiempo de tardanza de los administrativos en la Universidad Nacional de Jaén en el año 2024. Los datos se muestran en la tabla 23.

Tabla 23

Tiempo de tardanza de los administrativos en la Universidad Nacional de Jaén en el año 2024

$[L_I - L_S)$	Y_i	f_i
3 – 5	4	7
5 – 7	6	3
7 – 9	**8**	**15**
9 – 11	10	7
11 – 13	12	5
Total		**37**

Solución

Ubicar la clase modal (es el que tiene mayor frecuencia) en este caso es 15

Usando la fórmula para realizar los cálculos se tiene:

$$M_o = 7 + 2 \times \left[\frac{(15-3)}{(15-3)+(15-7)}\right] = 8.2$$

Interpretación:

La mayoría de los administrativos que trabajan en la "Universidad Nacional de Jaén" tienen 8.2 minutos de tardanza.

3.2.5.2 VENTAJAS Y DESVENTAJAS

- **Ventajas:**
  - Es fácil No es afectada por los valores extremos.
  - No es necesario ordenar los datos.
  - Es el más descriptivo y puede ser calculado a partir de datos
  - cuantitativos con intervalos abiertos, también como datos cualitativos.

- **Desventajas**
  - Su significado es restringido cuando el número de observaciones es pequeño.
  - Una distribución puede tener más de un valor modal, como también no puede tener, y esto hace difícil interpretar o continuar.
  - Cuando los datos se encuentran en intervalos pueden tomar valores que no serán sus valores originales.

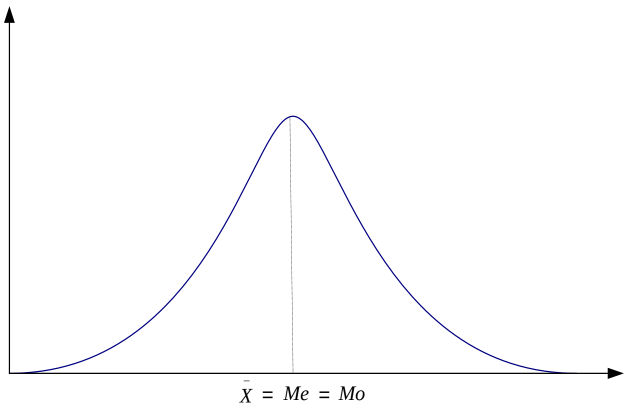
3.2.5.3 RELACIÓN ENTRE MEDIA, MEDIANA Y MODA

- **En distribuciones simétricas**

En una distribución de frecuencias simétricas (Figura 22) y es uní modal, coinciden el mismo valor la mediana, media y moda. Es decir: $\bar{x} = M_e = M_o$

Tabla 24

Distribución simétrica



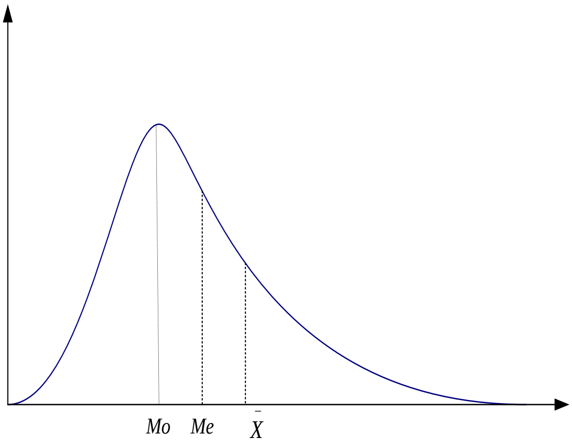
- **Distribuciones asimétricas.**

Si la distribución es unimodal (Figura 23), pero tiene simetría, las tres medianas forman valores diferentes y la mediana queda comprendida entre la moda y la media aritmética.

i) Si la distribución es alargada para valores grandes de la variable (asimétrica al derecho o positiva), entonces la situación general es: $\bar{x} > M_e > M_o$

Tabla 25

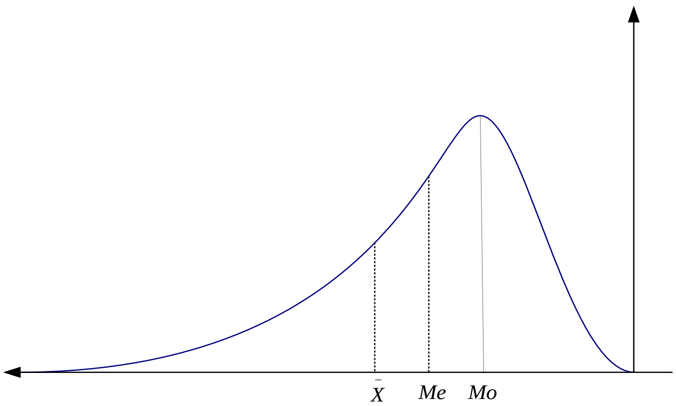
Distribución asimétrica a la derecha



ii) Si la distribución es alargada para valores pequeños de la variable (asimétrica a la izquierda o negativa), la situación general es: $\bar{x} < M_e < M_o$

Tabla 26

Distribución asimétrica a la izquierda



.

3.3 MEDIDAS DE DISPERSIÓN

3.3.1 INTRODUCCIÓN

La noción de dispersión o variabilidad está relacionada con cuánto se alejan los datos de una serie con respecto a un valor central, que suele ser la media aritmética. Es decir, se examina cómo se distribuyen los valores de los datos, analizando si están más concentrados o más dispersos.

No es suficiente calcular e interpretar las medidas de tendencia central o de posición, ya que a menudo hay datos extremos de la media aritmética, los cuales no están siendo representados por este parámetro.

Dos conjuntos de datos con la misma media aritmética no garantizan que sus distribuciones sean idénticas, es necesario examinar el grado de homogeneidad. Por ejemplo, los conjuntos de valores 5, 50, 95 y 49, 50, 51 tienen la misma media aritmética y mediana, pero en el primer conjunto, Rango la media aritmética se encuentra muy alejada de los valores extremos 5 y 95, a diferencia del segundo conjunto.

Para cuantificar el grado de dispersión, se utilizan los indicadores:

- Desviación media.
- Varianza.
- Desviación estándar.
- Coeficiente de variación.

3.3.2 RANGO DE LA VARIABLE

Se considera los dos valores extremos de una colección de datos.

$R = X_{máx} - X_{min}$

Donde:

$X_{máx}$: valor máximo de la variable x.

X_{min}: valor mínimo de la variable x.

Ejemplo 51.

La lista de datos representa las notas de 15 alumnos que llevan el curso de Estadística y Probabilidades:

12	05	13	10	11	17	19	07	13
12	15	10	12	08	09			

Solución

$R = X_{máx} - X_{min}$= 19 – 05 = 14

3.3.3 RECORRIDO INTERCUARTÍLICO

Se define como la diferencia entre el tercer y primer cuartil

$RI = Q_3 - Q_1$

Ejemplo 52.

Los datos representan las edades de un grupo de 20 turistas que están alojados en el hotel El León, durante el mes de febrero del 2023.

54	23	23	27	22	20	29	43	33	28
23	33	25	58	32	41	41	50	39	50

Solución

Ordenar los datos en forma ascendente:

20	22	23	23	23	25	27	28	29	32
33	33	39	41	41	43	50	50	54	58

Calculando el cuartil uno y cuartil tres se tiene:

$$\frac{nk}{4} = \frac{20 \times 1}{4} = 5$$

$$\frac{nk}{4} = \frac{20 \times 3}{4} = 15$$

$Q_1 = 23$ $\quad$ $Q_3 = 41$

$\Rightarrow$ $RI = Q_3 - Q_1 = 41 - 23 = 18$

3.3.4 DESVIACIÓN DEL CUARTIL

Es la mitad del recorrido intercuartil, es decir:

$$Q_D = \frac{Q_3 - Q_1}{2}$$

Donde:

Q_1 : Primer cuartil de los datos.

Q_3 : Tercer cuartil de los datos

La desviación cuartílica mide la distancia promedio entre los valores de un cuarto de los datos. Es representativo de la dispersión de los datos, se obtiene al promediar la mitad de los elementos intermedios en lugar de seleccionar uno de los cuartiles.

Nota: Si la distribución es muy asimétrica, es preferible calcular la desviación estándar.

Ejemplo 53.

Los datos son las edades de 20 turistas que se alojan en el hotel El León, en el mes de febrero del 2023.

54	23	23	27	22	20	29	43	33	28
23	33	25	58	32	41	41	50	39	50

Solución

Ordenar los datos en forma ascendente:

20	22	23	23	23	25	27	28	29	32
33	33	39	41	41	43	50	50	54	58

Calculando el cuartil uno y cuartil tres, se tiene:

$$\frac{nk}{4} = \frac{20 \times 1}{4} = 5$$

$$\frac{nk}{4} = \frac{20 \times 3}{4} = 15$$

$Q_1 = 23$ $\qquad Q_3 = 41$

La desviación cuartílica es:

$$Q_D = \frac{Q_3 - Q_1}{2} = \frac{41 - 23}{2} = 9$$

3.3.5 DESVIACIÓN MEDIA ABSOLUTA

$$DM = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$ Datos no tabulados

Datos tabulados

$$DM = \frac{\sum_{i=1}^{m} f_i \left| y_i - \bar{y} \right|}{n}$$

Donde:

m: Número de clase

f_i : Frecuencia absoluta de la clase

Y_i : Marca de clase

Esta última formula se puede escribir también así:

$$DM = \sum_{i=1}^{n} \frac{f_i}{n} \left| y_i - \bar{y} \right| = \sum_{i=1}^{m} h_i \left| y_i - \bar{y} \right|$$

La desviación de medias se utiliza como medida de dispersión en aquellas distribuciones en que la medida de tendencia central más significativa haya sido la media.

Ejemplo 54.

Se presentan los datos que corresponden a las edades de un grupo de 20 turistas que están hospedados en el Hotel El León durante el mes de febrero del 2023.

54	23	23	27	22	20	29	43	33	28
23	33	25	58	32	41	41	50	39	50

Determine la desviación de media absoluta.

Solución

Hallando el promedio:

$$\bar{x} = \frac{54 + 23 + 23 + \ldots + 39 + 50}{20} = 34.7$$

La desviación de medias es:

$$DM = \frac{\sum_{i=1}^{20} |x_i - \bar{x}|}{n} = \frac{|(54 - 34.7) + \ldots + (39 - 34.7) + (50 - 34.7)|}{20}$$

$DM = 9.84$

Ejemplo 55.

En un estudio se desea calcular la desviación de medias del tiempo promedio de cocción (en minutos) de 50 productos. Los datos se encuentran en la tabla 24:

Tabla 27

El tiempo promedio de cocción de los productos de envasado

$[L_I - L_S)$	Y_i	f_i	$\lvert Y_i - \bar{y} \rvert$	$f_i \lvert Y_i - \bar{y} \rvert$
49.5 – 50.5	50	1	4.1	4.1
50.5 – 51.5	51	3	3.1	9.3
51.5 - 52.5	52	5	2.1	10.5
52.5 – 53.5	53	9	1.1	9.9
53.5 – 54.5	54	12	0.1	1.2
54.5 – 55.5	55	10	0.9	9
55.5 – 56.5	56	5	1.9	9.5
56.5 – 57.5	57	3	2.9	8.7
57.5 – 58.5	58	2	3.9	7.8
Total		50		70

Solución

$$D_M = \frac{\sum f_i \lvert Y_i - \bar{y} \rvert}{n} = \frac{70}{50} = 1.4$$

3.3.6 VARIANZA

La varianza es una medida estadística que indica cuán dispersos o alejados están los valores de un conjunto de datos con respecto a su media. En otras palabras, evalúa la magnitud de las diferencias entre los datos y su promedio, midiendo la variabilidad de los valores. Una varianza alta indica que los datos están muy dispersos, mientras que una varianza baja refleja que los datos están más cerca de la media.

3.3.6.1 VARIANZA POBLACIONAL

Sea $x_1, x_2, x_3, ..., x_n$, se define como la media aritmética del cuadrado de las desviaciones de cada observación con respecto a su media y se denota por δ_x^2o $V[x]$

3.3.6.1.1 CÁLCULO DE LA VARIANZA

- **Para datos no tabulados**

Está dada por la siguiente formula:

$$\delta_x^2 = \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

Donde:

μ: Media poblacional;

Se sabe que: $\mu = \bar{x}$

x_i: Datos observados

n: Total de observaciones

Ejemplo 56.

Un grupo de docentes investigadores en salud, recolectaron muestras sobre el número de cirugías realizadas en 5 días. Los datos fueron:

x_i: 10; 3; 5; 6; 4

Hallar la varianza poblacional

Solución

Calculando el promedio:

$$\mu = \bar{x} = \frac{\sum x_i}{n} = \frac{10 + 3 + 5 + 6 + 4}{5} = 5,6$$

La varianza:

$$\delta_x^2 = \frac{(10-5{,}6)^2 + (10-5{,}6)^2 + \cdots + (10-5{,}6)^2}{5}$$

$$\delta_x^2 = \frac{29{,}2}{5} = 5{,}84 \text{ cirugías}^2$$

- **Para datos tabulados**

  Se define a través de la siguiente fórmula:

  $$\delta_x^2 = \frac{\sum_{i=1}^{n} f_i (Y_i - \mu)^2}{n}$$

  Donde:

  Y_i: Marca de clase

Ejemplo 57.

Un Tecnólogo realiza un estudio del número de errores detectados en sistemas médicos informatizados. Para dicho estudio se evalúa 20 historias clínicas electrónicas. Los datos se encuentran en la tabla 28:

Tabla 28

Distribución de historias clínicas electrónicas según el número de errores detectados en sistemas médicos informatizados

Y_i	f_i	$Y_i f_i$	$f_i (Y_i - \bar{x})^2$
1	3	3	15.188
2	5	10	7.813
3	2	6	0.125
4	4	16	2.25
5	6	30	18.375
Total (n)	20	65	43.75

Hallar la varianza poblacional

Solución

Calculando ***el promedio:***

$$\mu = \bar{x} = \frac{\sum x_i}{n} = \frac{65}{20} = 3{,}25$$

Varianza ***poblacional:***

$$\delta_x^2 = \frac{\sum_{i=1}^{n} f_i (Y_i - \mu)^2}{n} = \frac{43{,}75}{20} = 2{,}1875\ errores^2$$

3.3.6.2 VARIANZA MUESTRAL

La variancia de una muestra $x_1, x_2, x_3, \ldots, x_n$ de una variable o característica x que abreviadamente escribimos "S_x^2" se define como la media del cuadrado de las desviaciones de las observaciones con respecto de la media aritmética de esos datos.

3.3.6.2.1 CÁLCULO DE LA VARIANZA

- **Para datos no tabulados**

Se expresa mediante la siguiente fórmula:

$$S_x^2 = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}$$

Ejemplo 58.

Se realizó un estudio del número de unidades producidas (por hora) en una fábrica de alimentos.

x_i: 10; 3; 5; 6; 4

Hallar la varianza muestral

Solución

Calculando el promedio:

$$\mu = \bar{x} = \frac{\sum x_i}{n} = \frac{10 + 3 + 5 + 6 + 4}{5} = 5{,}6$$

Calculando la varianza***:***

$$S_x^2 = \frac{(10 - 5{,}6)^2 + (10 - 5{,}6)^{2\prime} \ldots + (10 - 5{,}6)^2}{5 - 1}$$

$S_x^2 = 7{,}3 \text{ unid. prod}^2$

❖ **Para datos no tabulados**

Está representada por la siguiente fórmula:

$$S_x^2 = \frac{\sum_{i=1}^{n} f_i(Y_i - \mu)^2}{n}$$

Ejemplo 59.

En un estudio realizado a 20 internos de la escuela profesional de Tecnología Médica de la Universidad Nacional de Jaén se evaluó el tiempo que emplean en un laboratorio para el procesamiento de las pruebas diagnósticas. Los datos se encuentran en la siguiente tabla:

Tabla 29

Distribución de internos de Tecnología Médica según el tiempo que emplean para el procesamiento de las pruebas

$[LI - LS)$	Y_i	f_i	$Y_i f_i$	$f_i(Y_i - \bar{x})^2$
1200 - 1500	1350	4	5400	1512900
1500 - 1800	1650	5	8250	496125
1800 - 2100	1950	3	5850	675
2100 - 2400	2250	2	4500	162450
2400 - 2700	2550	6	15300	2053350
Total (n)		20	39300	4225500

Hallar la varianza muestral

Solución

El promedio es:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{39300}{20} = 1965$$

Varianza muestral:

$$S_x^2 = \frac{\sum_{i=1}^{n} f_i(Y_i - \bar{x})^2}{n-1} = \frac{4225500}{19} = 222394{,}737 min^2$$

3.3.7 DESVIACIÓN TÍPICA O DESVIACIÓN ESTÁNDAR

Es una medida de la dispersión o variabilidad de un conjunto de datos en relación con su media. Indica cuánto se alejan, en promedio, los valores de los datos respecto a la media del conjunto. En otras palabras, mide cuánto varían los datos entre sí.

Matemáticamente, la desviación estándar se calcula tomando la raíz cuadrada de la varianza, que es el promedio de los cuadrados de las diferencias entre cada dato y la media. Una desviación estándar baja significa que los datos están muy cercanos a la media, mientras que una desviación estándar alta indica que los datos están más dispersos.

Formula:

Desviación estándar poblacional $\delta_x = \sqrt[2]{\delta_x^2}$

Desviación estándar muestral $S_x = \sqrt[2]{S_x^2}$

Ejemplo 60.

Del ejemplo 56 (Número de cirugías) y de la tabla 28 (Distribución de historias clínicas electrónicas según el número de errores detectados en sistemas médicos informatizados) Hallar e interpretar desviación estándar poblacional.

Solución

- Del ejemplo 56 (Número de cirugías) tememos:

$$\delta_x = \sqrt[2]{\delta_x^2}$$

$$\delta_x = \sqrt[2]{5{,}84} = 2{,}417\ cirugías$$

Interpretación

La variabilidad que existe en el número de cirugías es de 2,417

- De la tabla 28 (Distribución de historias clínicas electrónicas según el número de errores detectados en sistemas médicos informatizados)

$$S_x = \sqrt[2]{S_x^2}$$

$S_x = \sqrt[2]{1,303} = 1,518\,errores$

Interpretación

La variabilidad que existe en el número de errores detectados en sistemas médicos informatizados es de 1,518.

Ejemplo 61.

Del ejemplo 58 (Número de unidades producidas (por hora) en una fábrica de alimentos) y de la tabla 29 (Distribución de internos de Tecnología Médica según el tiempo que emplean para el procesamiento de las pruebas) Hallar e interpretar desviación estándar muestral.

Solución

- Del ejemplo 58 (Número de unidades producidas (por hora) en una fábrica de alimentos) tememos:

  $S_x = \sqrt[2]{S_x^2}$

  $S_x = \sqrt[2]{7,3} = 2,702\,unidades\,producidad$

Interpretación

La variabilidad que existe en el número de unidades producidas (por hora) en una fábrica de alimentos es de 2,702

- De la tabla 29 (Distribución de internos de Tecnología Médica según el tiempo que emplean para el procesamiento de las pruebas)

  $S_x = \sqrt[2]{S_x^2}$

  $S_x = \sqrt[2]{222394,737} = 471,587\,min$

Interpretación

La variabilidad que existe en el tiempo que emplean para el procesamiento de las pruebas es de $471,587 min$

3.3.7.1 VENTAJAS Y DESVENTAJAS

- **Ventajas:**
  - Interpretación clara
  - Útil para comparaciones: Es utilizada en diversas disciplinas
  - Aplicabilidad**:** Se utiliza en muchos métodos estadísticos
  - Sensibilidad a los cambios**:** Captura cambios pequeños en la variabilidad

- **Desventajas:**
  - Sensibilidad a valores atípicos.
  - **Requiere normalidad**: Es más adecuada para distribuciones simétricas o normales.

3.3.8 COEFICIENTE DE VARIACIÓN

Se utiliza para expresar la variabilidad relativa de una serie de datos en relación con su media. Se calcula como el cociente entre la desviación estándar (o desviación típica) y la media aritmética, se expresa en porcentaje. Es útil para comparar la dispersión de datos entre diferentes conjuntos, especialmente cuando las unidades de medida son diferentes o las escalas de los datos varían significativamente. Cuanto mayor sea el coeficiente de variación, mayor será la variabilidad relativa de los datos en relación con su media. Su fórmula es:

$$CV_x = \frac{S_X}{\bar{x}} \times 100\%$$

Ejemplo 62.

En la tabla 30 se presenta los datos de dos grupos de mujeres de 11 y 25 años

Tabla 30

Grupos de mujeres según su edad

Edad (Años)	Peso Medio (Kg)	Desviación Típica (Kg)
11	40	2
25	50	2

Calcular e interpretar: Cuál de las dos edades tiene mayor variación.

Solución

Para determinar cuál de las edades tiene mayor variación, se compara los coeficientes de variación de ambas pruebas.

Sea: A: El grupo de mujeres de 11 años

B: El grupo de mujeres de 25 años

$$CV_A = \frac{S_A}{\bar{x}} \times 100 = \frac{2}{40} \times 100 = 5\%$$

$$CV_B = \frac{S_B}{\bar{x}} \times 100 = \frac{2}{50} \times 100 = 4\%$$

Interpretación:

El grupo de mujeres de 11 años presenta mayor variación.

4 MEDIDAS DE FORMA DE LA DISTRIBUCION

4.1 INTRODUCCIÓN

En el análisis de datos, es importante calcular medidas que no solo revelen cuán dispersos están los datos, sino también cómo se distribuyen con respecto a su punto central. Estas características se conocen como asimetría, que describe la falta de simetría en la distribución y curtosis o apuntamiento, que indica la concentración de los datos en torno a su valor central. A diferencia de las medidas de dispersión, se centran en la magnitud de las variaciones, estas medidas proporcionan información sobre la dirección en la que ocurren esas variaciones.

4.2 MEDIDAS DE ASIMETRÍA

Las medidas de asimetría son indicadores estadísticos que muestran el grado de simetría o falta de ella en una distribución de datos. Se utilizan para describir la forma de una distribución en relación a su media. Existe varias medidas de asimetría, una de ellas es el *coeficiente de Pearson.*

4.2.1 COEFICIENTE DE PEARSON

Evalúa la asimetría de la distribución en relación con la media. Si el valor es positivo, indica que la distribución está inclinada hacia la izquierda (con sesgo positivo), mientras que un resultado negativo señala que la distribución tiene sesgo hacia la derecha.

i. Teniendo en cuenta que la media aritmética y la moda coincide en una distribución simétrica. Se expresa como:

$$A_s = \frac{\overline{x} - M_o}{s_x}$$

ii. En distribuciones simétricas se verifica: $\overline{x} - M_o \cong 3(\overline{x} - M_e)$, otra manera de expresar es:

$$A_S = \frac{3(\overline{x} - M_e)}{s_x}$$

Interpretación

- Si la distribución es simétrica, entonces $A_S = 0$, en este caso el orden de los promedios es:

$$\bar{x} = M_e = M_o$$

- Si la distribución es asimétrica positiva (sesgada) a la derecha $A_s > 0$. En este caso el orden de los promedios es:

$$\bar{x} > M_e > M_o$$

- Si la distribución es asimétrica negativa (sesgada) a la izquierda $A_s < 0$. En este caso el orden de los promedios es:

$$\bar{x} < M_e < M_o$$

Nota

iii. Si la distribución es asimétrica positiva (sesgada) a la derecha $A_s > 0$. En este caso el orden de los promedios es:

$$A_s = \frac{nM^3}{(n-1)(n-2)s^3}$$

- Si es simétrica, entonces: $A_s = 0$
- Si es asimétrica positiva entonces: $A_s > 0$
- Si es asimétrica negativa entonces: $A_s < 0$

iv. El coeficiente de asimetría en función de los cuantiles

a. El coeficiente de asimetría cuartílica, está dado en función de los cuarteles.

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

b. El coeficiente de asimetría en función de los percentiles (10 – 90) está dado por:

$$A_S = \frac{P_{90} - 2P_{50} + P_{10}}{P_{90} - P_{10}}$$

La interpretación es la misma que los otros coeficientes de asimetría.

4.2.1.1 REPRESENTACIÓN GRÁFICA DEL COEFICIENTE DE ASIMETRÍA

Figura 20

Distribución Simétrica

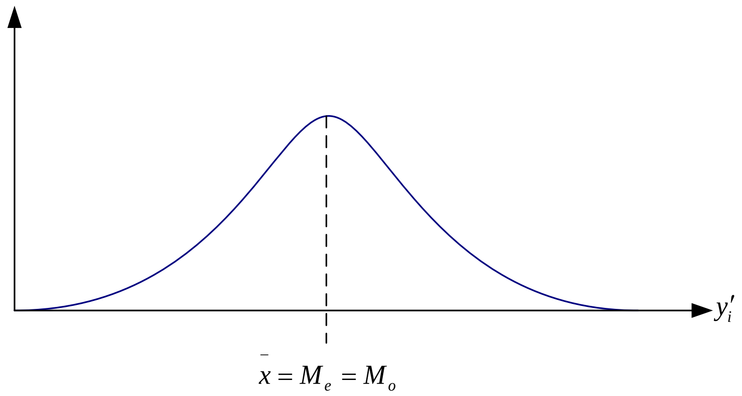


Figura 21

Distribución Asimétrica Positiva

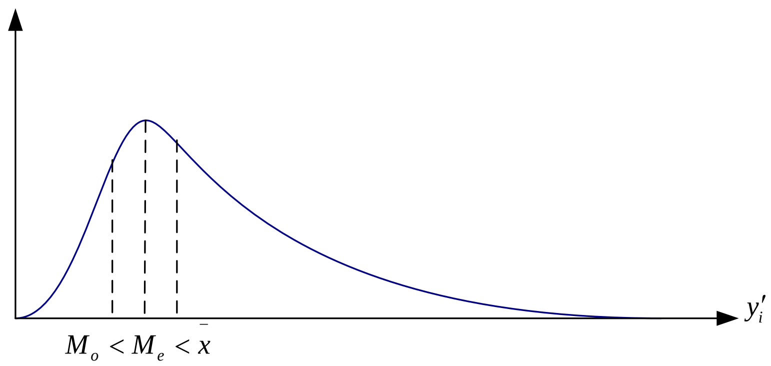
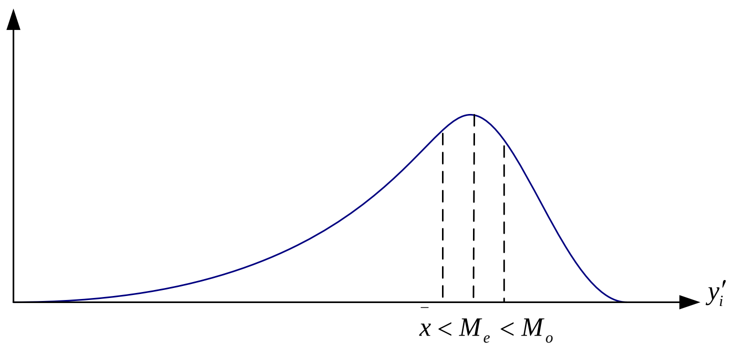


Figura 22

Distribución Asimétrica Negativa



Ejemplo 63.

Se presenta la tabla 31 de una muestra de 50 tiempos en horas de vida útil de los componentes electrónicos:

Tabla 31

Distribución de los componentes electrónicos según su vida útil

$[L_i - L_S]$	Y_i	fi	F_i	$y_i \times f_i$	$y_i^2 f_i$
600 - 800	700	8	8	5600	3920000
800 - 1000	900	16	24	14400	12960000
1000 - 1200	1100	11	35	12100	13310000
1200 - 1400	1300	10	45	13000	16900000
[1400- 1600]	1500	5	50	7500	11250000
Total		**50**		**52600**	**58340000**

Calcular e interpretar el coeficiente de asimetría.

Solución:

Realizando el cálculo de los estadísticos

Media

$$\overline{x} = \frac{\sum y_i \times f_i}{n} = \frac{52600}{50} = 1052$$

Varianza

$$S^2 = \frac{\sum_{i=1}^{m} y_i^2 \times f_i}{n} - \overline{x}^2 = \frac{58340000}{50} - 1052^2 = 60096$$

Desviación estándar

$$S = \sqrt{60096} = 245.15$$

Mediana:

$$M_e = 1000 + 200 \times \frac{[25 - 24]}{11} = 1018.18$$

Reemplazando en la fórmula de Asimetría:

$$A_k = 3\frac{(\overline{x} - Me)}{S} = 3\frac{(1052 - 1018.18)}{245.15} = 0.414$$

Interpretación:

La distribución de los datos presenta una asimetría positiva.

4.3 MEDIDAS DE APUNTAMIENTO O CURTOSIS

Las **medidas de apuntamiento** o **curtosis** indican el grado de concentración que tienen los datos alrededor de la media en una distribución. Miden qué tan "puntiaguda" o "achatada" es una distribución comparada con una distribución normal.

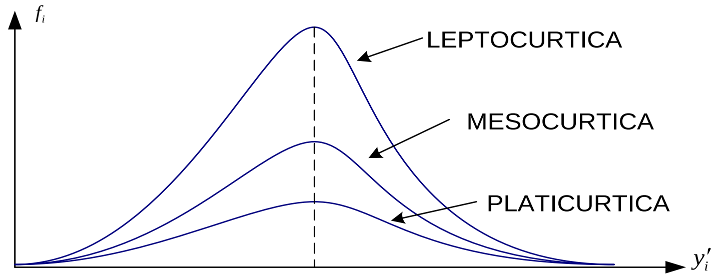
4.3.1 TIPOS DE CURTOSIS

Existen tres tipos principales de curtosis*:*

- Mesocúrtica: es similar a una distribución normal, presentando un nivel de concentración medio alrededor de los valores en torno al centro de la variable.
- Leptocúrtica: muestra una alta concentración de valores alrededor del centro de la variable.
- Platicúrtica: indica una baja concentración de valores en torno al centro de la variable.

4.3.1.1 REPRESENTACIÓN GRÁFICA DE LA CURTOSIS

Figura 29. Diferencias entre leptocúrtica, mesocúrtica, platicúrtica.



4.3.2 CÁLCULO DE LA CURTOSIS

- **Curtosis en función de percentiles.** El coeficiente está dado por:

$$K_U = \frac{(P_{75} - P_{25})}{2(P_{90} - P_{10})}$$

Interpretación

Si es normal, K tiende a 0,5

Si es leptocúrtica, K tiende a 1.0

Si es platicúrtica, K tiende a cero.

- **Curtosis en función de cuantiles.**

$$K = \frac{Q_3 - Q_1}{2\left(P_{90} - P_{10}\right)}$$

Interpretación

Si K = 0.263 la distribución es mesocurtica

Si K > 0.263 la distribución es más puntiaguda o leptocurtica.

Si K < 0.263 la distribución es achatada o platicurtica.

Ejemplo 64.

De la tabla 31 (Distribución de los componentes electrónicos según su vida útil). Calcular e interpretar la curtosis.

Solución:

Calculamos los percentiles:

$$P_{75} = L\inf + C\left[\frac{\frac{k(n)}{100} - F_{i-1}}{f_i}\right] = 1200 + 200\left[\frac{37.5 - 35}{10}\right] = 1250$$

$$P_{25} = L\inf + C\left[\frac{\frac{k(n)}{100} - F_{i-1}}{f_i}\right] = 800 + 200\left[\frac{12.5 - 8}{16}\right] = 856.25$$

$$P_{90} = L\inf + C\left[\frac{\frac{k(n)}{100} - F_{i-1}}{f_i}\right] = 1200 + 200\left[\frac{45 - 35}{10}\right] = 1400$$

$$P_{10} = L\inf + C\left[\frac{\frac{k(n)}{100} - F_{i-1}}{f_i}\right] = 600 + 200\left[\frac{5 - 0}{8}\right] = 725$$

Representamos en la fórmula de kurtosis

$$K_U = \frac{(1250 - 856.25)}{2(1400 - 725)} = 0.2917$$

5 DISTRIBUCIONES BIDIMENSIONALES

5.1 INTRODUCCIÓN

Con frecuencia se investiga dos o más características de cada elemento. Por ejemplo, se puede analizar el porcentaje de humedad y calorías; venta y número cliente. Si se asocia las características de las variables (x e y) están relacionadas entre sí y se llamará variable estadística bidimensional. En muchas situaciones, estas características están interrelacionadas, por lo que es crucial estudiarlas conjuntamente.

La variable estadística bidimensional (x, y) puede clasificarse según la naturaleza: cualitativas, cuantitativas discretas y cuantitativas continuas. Los tipos de distribución son:

- Dos variables cualitativas. Ejemplo: el rendimiento laboral y satisfacción del cliente.
- Uno cualitativo y otro cuantitativo. Ejemplo: puede ser discreto, como rendimiento laboral y números de clientes o continuo como rendimiento laboral y ventas.
- Dos variables cuantitativas, que pueden ser ambos discretos como número de mesas y número de clientes. Uno discreto y otro continuo como número de mesas y ventas. Ambos continuos como ventas y ganancias.

Cuando se observan más de dos características, se habla de variables estadísticas n dimensionales, lo que implica un análisis multivariante, aunque este aspecto no será abordado aquí.

El enfoque principal del estudio de dos características es el análisis descriptivo de series estadísticas bidimensionales. Esto abarca la presentación de tablas estadísticas bidimensionales, representación gráfica de estas tablas y descripción numérica de las series estadísticas de dos variables.

5.2 TABLAS ESTADÍSTICAS BIDIMENSIONALES

Se toma en cuenta un conjunto de n observaciones que se describen simultáneamente mediante las variables "x" e "y", las cuales tienen k modalidades o valores para la variable "x" y k modalidades o valores para la variable "y". La cantidad de observaciones que comparten simultáneamente una modalidad o valor específico de la variable "x" y otra de la variable "y" se representa como. La suma de las frecuencias absolutas es igual al total de observaciones.

$$\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij} = n$$

5.2.1 FRECUENCIAS ABSOLUTAS

La tabla estadística que describe las n observaciones se denomina tabla de distribución de frecuencia bidimensional. Esta tabla organiza en filas las modalidades de la característica "x" y en las columnas la característica "y", formando una estructura de K filas y L columnas. Se asume que las frecuencias absolutas de la fila (columna) no se anulan simultáneamente. Si esto sucede, se puede optar por no considerar la modalidad o valor correspondiente de x (y), o incluso agruparlo con otra modalidad.

Tabla 32

Distribución bidimensional de frecuencias absolutas

Modalidades o valores de X \ Modalidades o valores de Y	Y_1	Y_2	…	Y_J	…	Y_K	**Totales Horizontales** $= \sum_{j=1}^{l} n_{ij} = n_{i.}$
X_1	n_{11}	n_{12}	…	n_{1j}	…	n_{1l}	$n_{1.}$
X_2	n_{21}	n_{22}		n_{2j}		n_{2l}	$n_{2.}$
⋮	…	…	…	…	…	…	…
X_i	n_{i1}	n_{i1}	…	n_{ij}	…	n_{il}	$n_{i.}$
⋮	…	…	…	…	…	…	…
X_K	n_{k1}	n_{k2}		n_{kj}		n_{kl}	$n_{k.}$
Totales Verticales $= \sum_{i=1}^{K} n_{ij} = n.j$	$n_{.1}$	$n_{.2}$	…	$n_{.j}$	…	$n_{.l}$	$n.. = \sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij} = n$

Dónde: se designa por un punto el total según el índice *i* o el índice *j*. Es decir; $n_i.$ es la suma del total de las frecuencias absolutas n_{ij} según el índice j, esto es

$$n_{i.} = \sum_{j=1}^{l} n_{ij},$$

$n_{.j}$ es la suma total de las frecuencias absolutas n_{ij} según el índice i, o sea;

$$n._j = \sum_{i=1}^{K} n_{ij},$$

$n_{..}$ (igual a **n**) es la suma total de las frecuencias absolutas n_{ij} según los índices *i* y *j*, así como también la suma total de los totales n_i según *j* o de los totales $n_{.j}$ esto es

$$n\cdot\cdot = \sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij} = \sum_{i=1}^{k} n_{i.} = \sum_{j=1}^{l} n_{.j} = n$$

La frecuencia absoluta $n_{i.}$ es el número de observaciones que presentan la modalidad o valor de x_i de la característica x, independiente de las modalidades o valores de la característica y. Análogamente, $n_{.j}$ es el número de observaciones que corresponden a la modalidad o valor y_i de la característica e independiente de las modalidades o valores de x.

Asimismo, si la variable x (y/o la variable y) es continua x_i(y/o y_i) representará la marca de clase número i(o j), de tal manera que se reducirá el caso continuo al caso discreto como al estudiar las variables estadísticas de una dimensión.

Si las variables x e y son cualitativas la tabla de distribución bidimensional se llama Tabla de Contingencia.

5.2.2 FRECUENCIAS RELATIVAS

Se llama frecuencia relativa o simplemente frecuencia de la pareja o valores o modalidades x_i e y_j (o bien frecuencia total) a la proporción de observaciones que presentan simultáneamente los valores o modalidades x_i e y_j; es decir:

$$h_{ij} = \frac{n_{ij}}{n}$$

Tabla 33

Distribución bidimensional de frecuencias relativas

Modalidades o valores de X \ Modalidades o valores de Y	Y_1	Y_2	…	Y_J	…	Y_K	**Totales Horizontales** $\sum_{j=1}^{l} h_{ij} = h_{i.}$
X_1	h_{11}	h_{12}	…	h_{1j}	…	h_{1l}	$h_{1.}$
X_2	h_{21}	h_{22}		h_{2j}		h_{2l}	$h_{2.}$
⋮	…	…	…	…	…	…	…
X_i	h_{i1}	h_{i1}	…	h_{ij}	…	h_{il}	$h_{i.}$
⋮	…	…	…	…	…	…	…
X_K	h_{k1}	h_{k2}		h_{kj}		h_{kl}	h_k
Totales Verticales $\sum_{i=1}^{K} n_{ij} = n_{.j}$	$h_{.1}$	$h_{.2}$	…	$h_{.j}$	…	$h_{.l}$	$h_{..} = \sum_{i=1}^{k}\sum_{j=1}^{l} h_{ij} = 1$

La suma de las frecuencias relativas de todos los pares de valores o modalidades posibles es igual a la unidad. Es decir,

$$\sum_{i=1}^{k}\sum_{j=1}^{l} h_{ij} = 1\,.$$

Las sumas parciales se designan igualmente por un punto (.) en lugar de índice, que hace la función de la sumatoria.

$$h_{i.} = \sum_{j=1}^{l} h_{ij} = \frac{n_{i.}}{n}$$

$$h_{.j} = \sum_{i=1}^{k} h_{ij} = \frac{n.j}{n}$$

$$\sum_{i=1}^{k} h_{i.} = \sum_{j=1}^{l} h_{.j} = 1.$$

En ciertas situaciones podrá definirse las frecuencias relativas con relación al total de cada columna o de fila:

$$\frac{n_{ij}}{n_{i.}} \; y \; \frac{n_{ij}}{n_{.j}} \;, \quad i = 1,2,\ldots,\; k \quad ; \quad j = 1,2,\ldots,l$$

5.2.3 DISTRIBUCIONES MARGINALES

Al estudiar una distribución bidimensional, se puede enfocar el estudio en el comportamiento de una de las variables, sin considerar como se comporta la otra. Este enfoque se conoce como análisis de una distribución marginal.

En las distribuciones bidimensionales se pueden obtener **dos** distribuciones marginales: una para la variable "*x*", y otra para la variable "*y*".

- **Distribución marginal de X**

Si se observa la columna marginal (columna de los totales horizontales) de la tabla. Las frecuencias absolutas $n_{i.}$ definen la distribución marginal de la variable x. Esta es una distribución de una sola característica:

$$n_{i.} = \sum_{j=1}^{l} n_{ij}$$

La frecuencia relativa marginal del valor $x_{i.}$ es igual $h_{i.}$,

$$h_{i\cdot} = \frac{n_{i\cdot}}{n}$$

Asimismo, la suma de las frecuencias absolutas marginales es n:

$$\sum_{i=1}^{k} n_{i.} = n_{..} = n$$

La suma de las frecuencias marginales es igual a uno: $\sum_{i=1}^{k} h_{i\cdot} = 1$

Tabla 34

Valores o modalidades de X

Valores o modalidades de x	Frecuencias absolutas $n_{i.}$	Frecuencia relativa $h_{i.}$
X_1	$n_{1.}$	$h_{1.}$
X_2	$n_{2.}$	$h_{2.}$
:	…	…
X_i	$n_{i.}$	$h_{i.}$
:	…	…
X_K	$n_{k.}$	$h_{k.}$
Total	$n_{..}$	1.00

- **Distribución marginal de Y**

  Análogamente, la distribución marginal de la característica "y" está definida por las frecuencias absolutas marginales $n_{.j}$

  La frecuencia marginal relativa del valor o modalidad y_j es:

$$h_{\cdot j} = \frac{n_{\cdot j}}{n}$$

Tabla 35

Valores o modalidades de Y

Valores o modalidades de y	Frecuencias absolutas $n_{.j}$	Frecuencia relativa $h_{.j}$
Y_1	$n_{\cdot 1}$	$h_{\cdot 1}$
Y_2	$n_{\cdot 2}$	$h_{\cdot 2}$
:	…	…
Y_i	$n_{\cdot i}$	$h_{\cdot i}$
:	…	…
Y_K	$n_{\cdot j}$	$h_{\cdot k}$
Total	$n_{..}$	1.00

Las variables marginales se comportan como variables unidimensionales, por lo que pueden ser representadas en tablas de frecuencias.

5.2.4 DISTRIBUCIÓN CONDICIONAL

La distribución de frecuencias absolutas de la variable x condicionada a la variable y, toma el valor fijo $y_j (y = y_j), \quad j = 1,2,..., l$ se denota por $n_{\frac{x}{y}=y_j}$

$n_{\frac{x}{y}=y_j} = n_{ij} \quad , i = 1,2,\ldots,k$

Existe distribuciones condicionadas de la variable "x" que le corresponden un valor y_j de la variable "y".

La distribución de frecuencias relativas de la variable "x" condicionada a la variable "y" tome el valor fijo $y_j\left(y = y_j\right), con \ j = 1,2, \ldots, l$ es igual a:

$$n_{\frac{x}{y}=y_j} = \frac{n_{ij}}{n_{.j}}, i = 1,2, \ldots, k$$

La tabla estadística de la distribución de x condicionada a que la variable y tome el valor de y_j tiene la siguiente forma:

Tabla 36

Distribución de X condicionada a la variable Y

Valores o modalidades de x	Frecuencias absolutas de x condicionada por $y = y_i$ $n x/_{y=y_i}$	Frecuencia relativa de x condicionada por $y = y_i$ $h x/_{y=y_i}$
X_1	n_{1j}	$\frac{n_{1j}}{n_{.j}}$
X_2	n_{2j}	$\frac{n_{2j}}{n_{.j}}$
:	…	…
X_i	n_{ij}	$\frac{n_{ij}}{n_{.j}}$
:	…	…
X_K	n_{kl}	$\frac{n_{kj}}{n_{.j}}$
Total	$n_{.j}$	1.00

En forma similar se define la distribución de “y” condicionada a que “x” tome el valor $x_i (x = x_i)$ los $n_{i.}$ elementos poseen el valor o modalidad $x_{i.}$ de la variable x se reparten según la variable y por las frecuencias absolutas de la i-ésima fila de la tabla de dos entradas. La frecuencia relativa de “y” condicionada por $x = x_i$ es:

$$n_{\frac{y}{x}=x_j} = \frac{n_{ij}}{n_{i.}}$$

Se tiene *k* distribuciones según la variable “y”. La tabla estadística de la distribución de “y” condicionada por $x = x_i$ tiene la forma:

Tabla 37

Distribución de Y condicionada a la variable X

Valores o modalidades de y	Frecuencias absolutas de Y condicionada por $x = x_i$ $n_{y/x=x_i}$	Frecuencia relativa de x condicionada por $x = x_i$ $h_{y/x=x_i}$
Y_1	n_{i1}	$\frac{n_{i1}}{n_{i}.}$
Y_2	n_{i2}	$\frac{n_{i2}}{n_{i}.}$
:	...	...
Y_i	n_{ij}	$\frac{n_{ij}}{n_{i}.}$
:	...	...
Y_K	n_{il}	$\frac{n_{il}}{n_{i}.}$
Total	$n_i.$	1.00

Ejemplo 59.

Un establecimiento de salud desea relacionar el tipo de atención que brinda (x) (Excelente, Buena, Regular, Malo. Pésimo) y los ingresos semanales percibidos (en miles) (y). Se obtuvo una muestra de 20 pacientes, los resultados se dan a continuación:

X:	R	R	B	B	B	B	B	B	E	E
Y:	18.9	19.0	20.8	22.7	23.9	25.2	25.7	27.4	40.0	35.6

X:	E	E	E	M	P	P	R	R	R	P
Y:	29.3	30.8	31.6	15.3	10.0	11.6	13.2	15.3	12.7	10.5

A partir de la información:

a. Elabore la tabla de distribución de frecuencias absolutas simples.
b. Elabore la tabla de distribución de frecuencias relativas simples. Interprete.
c. Elabore las tablas de distribuciones marginales (x e y). Interprete.
d. La distribución de x condicionada a que $y = [10 - 16)$ $y = [28 - 34)$
e. b. La distribución de y condicionada por $x = \mathrm{Re}\,gular$ $x = Pésimo$
f. Graficar

Solución

Trabajando para cada variable:

- **Para X:**

  V.E: Tipo de atención

  T.V: Cualitativa/Ordinal

  $R_x = \{Excelente, Bueno, Re\,g\,ular, Malo, Pésimo\}$

- **Para Y**:

  V.E: Ingresos semanales

  T.V: Cuantitativo Continuo/Razón

  Calcular el rango

  R = 40 – 10 = 30

  Número de intervalos

  m = 1 + 3.33 log (20) = 5.33

  m = 5 intervalos

  Amplitud intervália

  $C = \frac{30}{5} = 6$

Realizando el conteo en la tabla tenemos:

El primer punto (R ; 18.9)

El punto onceavo (E ; 29.3)

Tabla 38

Tabla de conteo

Y X	[10 16)	[16 22)	[22 28)	[28 34)	[34 40]	**TOTAL**
Excelente				||| (E; 29.3)	|	**5**
Buena		|	~~||||~~			**6**
Regular	|||	|| (R;18.9)				**5**
Malo	|					**1**
Pésimo	|||					**3**
TOTAL	**7**	**3**	**5**	**3**	**2**	**20**

Tabla 39

Distribución bidimensional de frecuencias absolutas simples

Ingresos semanales Tipo de atención	[10 16)	[16 22)	[22 28)	[28 34)	[34 40]	TOTAL
Excelente				3	2	5
Buena		1	5			6
Regular	3	2				5
Malo	1					1
Pésimo	3					3
TOTAL	7	3	5	3	2	20

Interpretación:

$n_{2,3} = 5$. Cinco pacientes expresan que la atención del establecimiento es buena y los ingresos semanales oscila de 2200 a más, pero menos de 2800 soles.

$n_{4,1} = 1$ Un paciente manifiesta que la atención del establecimiento es mala y los ingresos semanales oscila de 1000 a más pero menos de 1600 soles.

$n_{2.} = 6.$ Seis pacientes expresan que la atención que brinda el establecimiento es buena.

$n_{.1} = 7.$ El ingreso semanal que percibe el establecimiento es de 1000 a más pero menos de 1600 soles.

Tabla 40

Distribución bidimensional de frecuencias relativas simples

Ingresos semanales / Tipo de atención	[10 16)	[16 22)	[22 28)	[28 34)	[34 40]	**TOTAL**
Excelente				3/20 =0.15	2/20=0.10	**0.25**
Buena		1/20 =0.05	5/20 =0.25			**0.30**
Regular	3/20 =0.15	2/20 =0.1				**0.25**
Malo	1/20 =0.05					**0.05**
Pésimo	3/20 =0.15					**0.15**
TOTAL	**0.35**	**0.15**	**0.25**	**0.15**	**0.10**	**1**

Interpretación:

$h_{2,3} = 0.25.$ Proporción de pacientes que indican que la atención del establecimiento es buena y los ingresos semanales es de 2200 a más pero menos de s/.2800

$h_{4,1} = 0.05.$ Proporción de pacientes que expresan que la atención del establecimiento es mala y los ingresos semanales es de 1000 a más, pero menos de s/.1600

$h_{2.} = 0.3$ Proporción de pacientes que manifiestan que la atención del establecimiento es buena.

$h_{.1} = 0.35.$ La proporción del ingreso semanal del establecimiento es de 1000 a más pero menos de s/1600

Tabla 41

Distribución marginal de X

Tipo de atención	$n_{i.}$	$h_{i.}$	$\%h_{i.}$
Excelente	5	0.25	25%
Buena	6	0.30	30%
Regular	5	0.25	25%
Malo	1	0.05	5%
Pésimo	3	0.15	15%
TOTAL	**20**	**1.00**	**100%**

Interpretación:

El 30% de los pacientes manifiestan que la atención del establecimiento es buena mientras que el 5% indican que es malo.

Tabla 42

Distribución marginal de Y

$[L_I - L_S)$	$n_{.j}$	$h_{.j}$	$\%h_{.j}$
[10 16)	7	0.35	35%
[16 22)	3	0.15	15%
[22 28)	5	0.25	25%
[28 34)	3	0.15	15%
[34 40]	2	0.10	10%
TOTAL	**20**	**1.00**	**100**%

Interpretación:

El 35% de los ingresos semanales del establecimiento oscila de 1000 a más pero menos de 1600 soles semanales, mientras que el 10% es de 3400 a más pero menos de 4000 soles.

Tabla 43

La tabla de distribución x condicionada por y = [10 - 16)

Valores de x	Excelente	Bueno	Regular	Malo	Pésimo	Total
$n_{x/y} = [10 - 16)$	0	0	3	1	3	7
$h_{x/y} = [10 - 16)$	0	0	3/7	1/7	3/7	1

Tabla 44

Distribución de x condicionada por y=[28-34)

Valores de x	Excelente	Bueno	Regular	Malo	Pésimo	Total
$n_{x/y} = [28-34)$	3	0	0	0	0	3
$h_{x/y} = [28-34)$	3/3	0	0	0	0	1

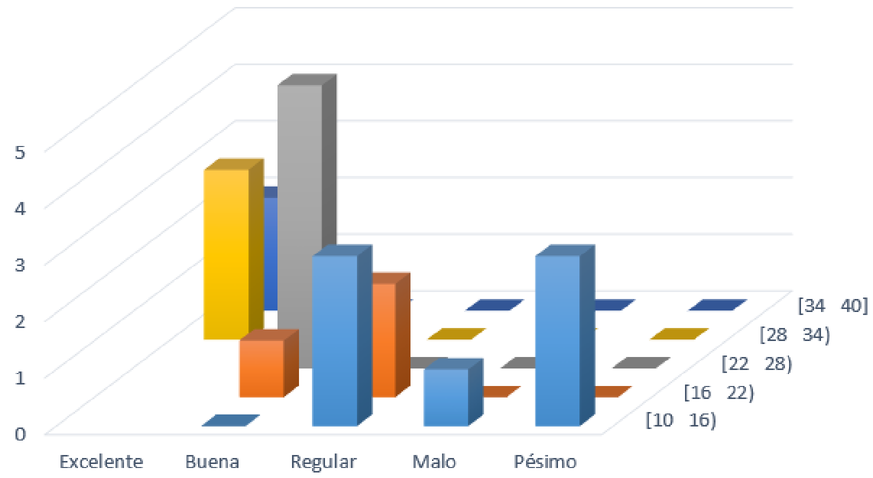
Tabla 45

La distribución de frecuencias absolutas de y condicionado por X=regular

Valores de y	[10 - 16)	[16 - 22)	[22 - 28)	[28 - 34)	[34 - 40]	Total
$ny/_x = regular$	3	2	0	0	0	5
$hy/_x = regular$	3/5	2/5	0	0	0	1

Figura 23

Tipo de atención del establecimiento y el ingreso semanal



Interpretación:

Podemos inferir que cinco pacientes indican que la atención del establecimiento es buena y que el ingreso semanal es de 2200 a más, pero menos de 2800 soles mientras que un paciente indica que la atención es buena y que el ingreso semanal es de 1600 a más pero menos de 2200 soles.

Ejemplo 60.

Se obtuvo los resultados de recuperación y mortalidad de los accidentes ocurridos en el centro de trabajo y trasladados a los hospitales. Los datos se presentan en la siguiente tabla:

Tabla 46

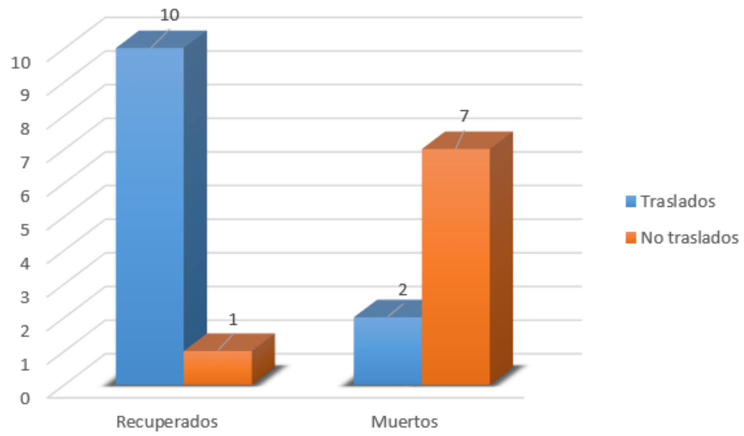
Proceso de recuperación y fallecimiento de los trabajadores

Y / X	Recuperados	Muertos	Total
Traslados	10	2	12
No traslados	1	7	8
Total	11	9	20

Solución

Figura 24

Proceso de recuperación y fallecimiento de los trabajadores



Interpretación:

El gráfico indica que los trabajadores que fueron trasladados tuvieron una mayor tasa de recuperación (10 recuperados y 2 fallecidos), mientras que los no trasladados tuvieron una tasa de fallecimiento más alta (7 fallecidos y solo 1 recuperado). Esto podría indicar que el traslado tuvo un impacto positivo en la recuperación de los trabajadores.

6 CORRELACIÓN Y REGRESIÓN LINEAL

6.1 INTRODUCCIÓN

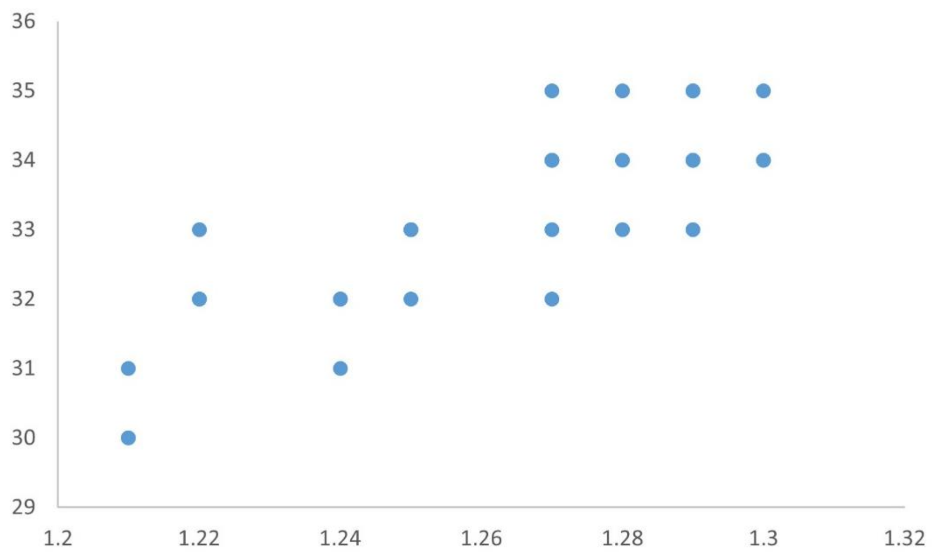
La correlación y la regresión lineal son dos métodos estadísticos utilizados para analizar la relación entre dos variables cuantitativas. Aunque están estrechamente conectados, cada uno se enfoca en estrategias de análisis distintas. El coeficiente de correlación mide el grado de asociación lineal entre las variables (X, Y) sin asumir ninguna direccionalidad específica en la relación entre ambas. Por otro lado, la regresión lineal investiga las relaciones entre variables y permite predecir los valores de una variable dependiente Y a partir de los valores de una variable independiente X, bajo la suposición de que existe una relación entre ellas.

6.2 DIAGRAMA DE DISPERSIÓN

Un diagrama de dispersión es una representación gráfica que muestra la relación entre dos variables(x,y) en un conjunto de datos. Este diagrama es útil para identificar patrones, tendencias.

Figura 25

Diagrama de dispersión



Interpretación.

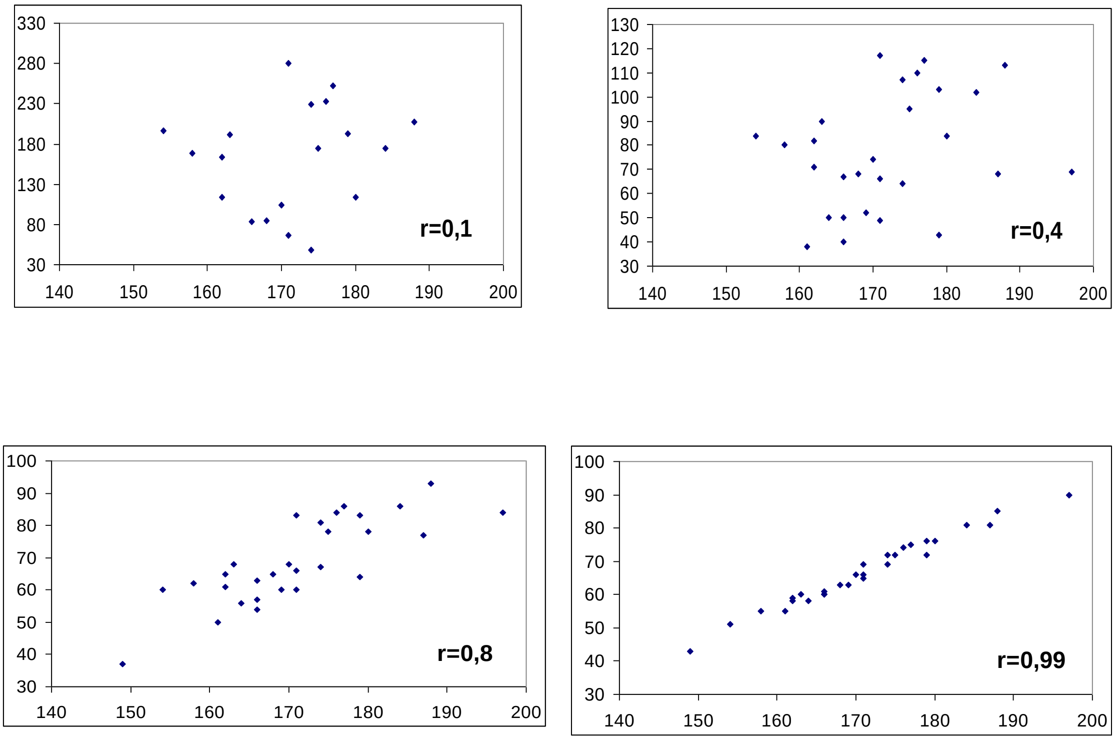
Si analizamos la distribución de los puntos en este gráfico de dispersión, se puede ver claramente una relación lineal entre los valores de las variables x e y

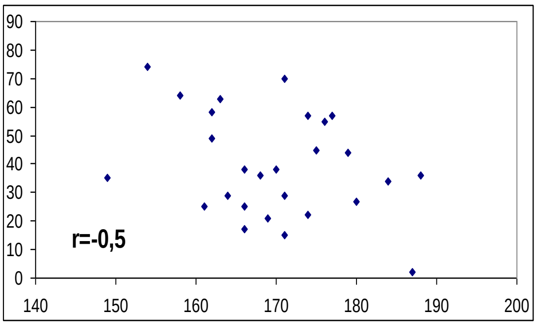
6.3 CORRELACIÓN LINEAL

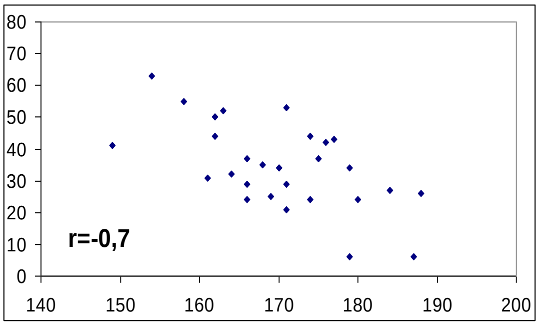
Es una medida estadística que evalúa la relación lineal entre dos variables cuantitativas, de manera que se puede expresar mediante una línea recta. La correlación lineal, conocida coeficiente de correlación de Pearson, no implica causalidad; simplemente indica la fuerza y la dirección de la relación lineal entre las dos variables, se representa por el símbolo "r" y puede variar en un rango de -1 a 1. Los valores de -1 y 1 indican una relación perfectamente negativa o positiva, respectivamente, mientras que 0 indica la ausencia de correlación lineal. Cuanto más cercano sea el valor de r a -1 o 1, más fuerte será la relación lineal, y cuanto más cercano a 0, menos fuerte será la relación.

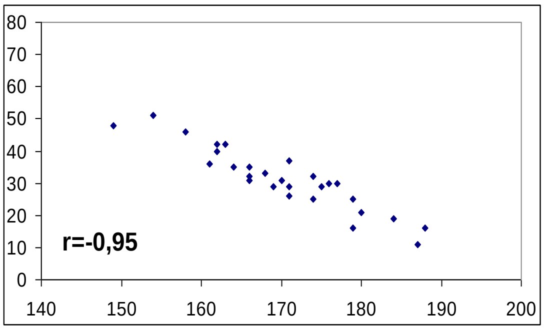
Tabla 47

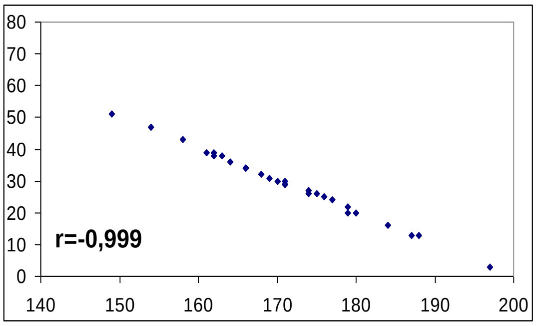
Nubes de puntos y fuerza de la asociación entre las dos variables











Para calcular el coeficiente de correlación lineal se utiliza la fórmula siguiente:

$$r = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{\sqrt{n\sum X^2 - \left(\sum X_i\right)^2} \times \sqrt{n\sum Y^2 - \left(\sum Y_i\right)^2}}$$

Donde:

n : Número de pares de datos

X e Y : variables (independiente) y (dependiente)

Tabla 48

Interpretación coeficiente de correlación de Pearson

Valor	Significado
r = -1	Correlación negativa perfecta
-0,90 ≤ r <-1	Correlación negativa muy alta
-0,75 ≤ r <-0,90	Correlación negativa alta
-0,50 ≤ r <-0,75	Correlación negativa moderada
-0,20 ≤ r <-0,50	Correlación negativa baja
0 < r <-0,20	Correlación negativa muy baja
r = 0	Correlación nula
0 < r ≤ 0,20	Correlación positiva muy baja

$0{,}20 < r \leq 0{,}50$	Correlación positiva baja
$0{,}50 < r \leq 0{,}75$	Correlación positiva moderada
$0{,}75 < r \leq 0{,}90$	Correlación positiva alta
$0{,}90 < r < 1$	Correlación positiva muy alta
$r = 1$	Correlación positiva perfecta

Ejemplo 61.

Se desea estudia la relación, de una empresa quiere predecir el valor anual de sus ventas totales en un país determinado, basándose en la relación entre estas ventas y la renta nacional.

Ventas	189	190	208	227	239	252	257	274	293	308	316
Renta	402	404	412	425	429	436	440	447	458	469	469

Con base en esta información:

a. Elabore un diagrama de dispersión. Comentar.

b. Calcule el coeficiente de correlación lineal

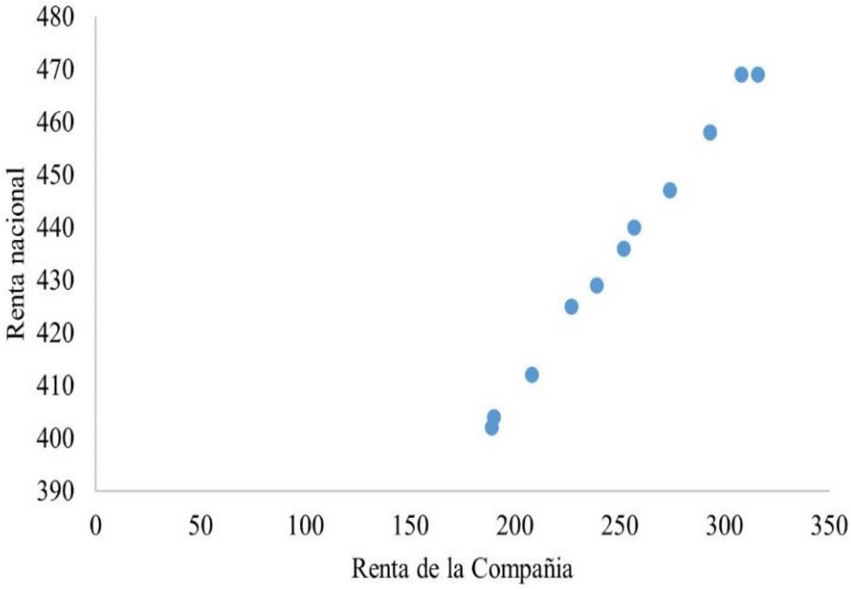
Solución

Realizando los cálculos:

	X: Ventas	Y: Renta	X^2	Y^2	XY
	189	402	35721	161604	75978
	190	404	36100	163216	76760
	208	412	43264	169744	85696
	227	425	51529	180625	96475
	239	429	57121	184041	102531
	252	436	63504	190096	109872
	257	440	66049	193600	113080
	274	447	75076	199809	122478
	293	458	85849	209764	134194
	308	469	94864	219961	144452
	316	469	99856	219961	148204
$\sum$	**2753**	**4791**	**708933**	**2092421**	**1209720**

Figura ***26***

Diagrama de dispersión la renta nacional y las ventas de la compañía



Interpretación:

En el diagrama de dispersión podemos observar el comportamiento de los datos de forma general, esto nos indica que cuando las ventas anuales se incrementa la renta nacional tiende a incrementarse, por lo que el coeficiente de correlación o grado de asociación entre las variables es positivo.

Calcule el coeficiente de correlación lineal

$$r = \frac{11(1209720) - 2753(4791)}{\sqrt{11(708933) - (2753)^2} \times \sqrt{11(2092421) - (4791)^2}} = 0.998$$

Interpretación:

Esto nos indica que existe una correlación positiva muy alta entre las ventas anuales y la renta nacional.

6.4 COEFICIENTE DE DETERMINACIÓN

Es una medida estadística que indica qué tan bien se ajusta un modelo de regresión a los datos observados. Representa la proporción de la variación total de la variable dependiente que es explicada por la variación de la variable independiente en el modelo.

$$0 \leq R \leq 1$$

Para *calcular el coeficiente de determinación lineal* se utiliza la fórmula siguiente:

$$R = r^2$$

Lo cual el porcentaje de variación de la variable de respuesta (Y) es explicado por la variable predictora (X).

Un R^2 de 1 indica que el modelo explica perfectamente todas las variaciones de los datos, mientras que R^2 de 0 indica que el modelo no explica ninguna de las variaciones en los datos.

Ejemplo 57

los datos corresponden a las ventas anuales y la renta nacional. Hallar e interpretar el coeficiente de determinación.

Solución

Se tiene el coeficiente de correlación lineal ($r = \quad 0{,}998$) entonces:

$$R = (0{,}998)^2 = 0.996$$

$$\%R = 99.6\%$$

Interpretación:

El 99.6% de la renta nacional es explicado por las ventas totales (anuales).

6.5 REGRESIÓN LINEAL

La regresión lineal es una técnica estadística utilizada para modelar la relación entre una variable dependiente y una o más variables independientes. El objetivo es encontrar la línea recta que mejor se ajusta a los datos, minimizando la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por la línea.

El modelo es:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

Donde:

Y	:	Variable independiente
$\beta_0 y \beta_1$	:	Son dos constantes desconocidas que representan el punto de intersección y la pendiente respectivamente.
ε	:	Es el error aleatorio (función de pérdida)

La Fórmula para calcular pendiente e intersección:

$$\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} \qquad \beta_0 = \frac{\sum y - \hat{\beta}_1 \sum x}{n}$$

Para calcular la línea de regresión, aplicaremos el método de mínimos cuadrados, que implica reducir al mínimo la suma de los cuadrados de los errores.

Ejemplo 64:

Del ejemplo 57, los datos corresponden a las ventas anuales y la renta nacional.

a. Hallar la ecuación de regresión lineal.
b. En el diagrama de dispersión trazar la recta de la ecuación de regresión lineal.
c. Si las ventas totales es 350 ¿Cuál es la renta nacional?
d. Si la renta nacional es de 450 ¿Cuál es la venta total?

Solución

Realizando los cálculos:

$$\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{11(1209720) - 2753(4791)}{11(708933) - 2753^2} = 0.535$$

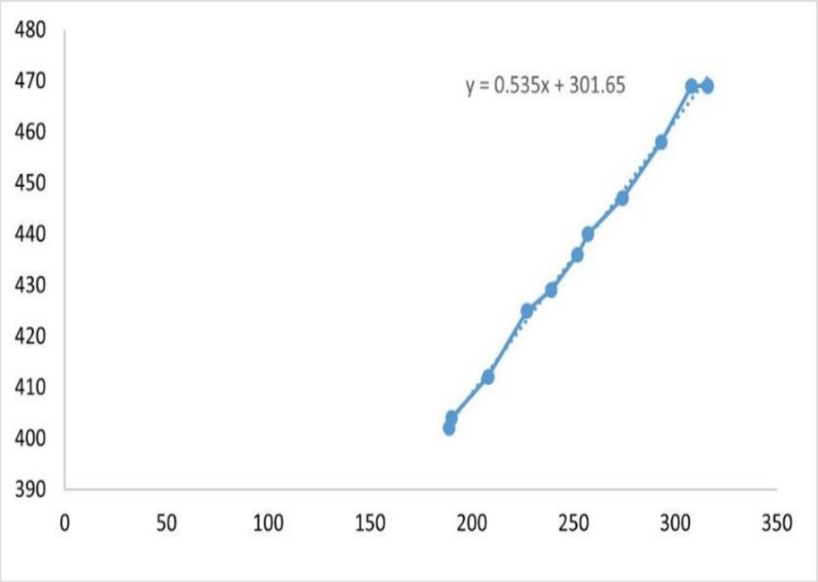
$$\beta_0 = \frac{\sum y - \hat{\beta}_1 \sum x}{n} = \frac{4791 - 0.535(2753)}{11} = 301.654$$

Por lo tanto***, la ecuación de regresión es:***

$$Y = 301.654 - 0.535x$$

Figura 27

Diagrama de dispersión



Si las ventas totales es 350 ¿Cuál es la renta nacional?

$$Y = 301.654 - 0.535(350) = 488.9$$

Si la renta nacional es de 450 ¿Cuál es la venta total?

$$450 = 301.654 - 0.535x$$

$$x = \frac{450 - 301.654}{0.535} = 277.3$$

7 EJERCICIOS RESUELTOS

APLICACIONES ESTADÍSTICAS UTILIZANDO EL SOFTWARE JAMOVI

7.1 CAPÍTULO 1

1. Identificar si cada una de las situaciones representa una población (P) o una muestra (M)
   - Habitantes distrito de Jaén e el año 2024....................(*P*)
   - Números de transacciones que realiza la empresa "Coca cola" en el año 2020....................(P)
   - Cantidad de habitantes que se encuentran en el Perú en el año 2020.................(P)
   - El 50% de estudiantes egresados de la Universidad Nacional de Jaén.........(M)
   - Cantidad de internos que se encuentran en el penal de Jaén........(P)
   - Estudio del 30% de los internos procesados en la ciudad de Jaén..............(M)
   - Se estudió antecedentes judiciales de 20 mujeres cuyas edades se encuentran entre 20 a 30 años que desean trabajar en la fábrica de espárragos en la ciudad de Trujillo en el año 2022................(*M*)

2. Un investigador recopila información de los ingresos anuales y el número de años de experiencia laboral de los trabajadores de una empresa de galletas "San Jorge". En relación a este estudio, identificar cada una de las propiedades como constante (C) o variable (V)
   - Sexo................(V)
   - Ingresos anuales.................(V)
   - Nacionalidad....................(C)
   - Profesión.....................(V)
   - Número de años de experiencia................(V)

- Producción por trabajador................(V)
- Horas extras que realizan cada trabajador................(V)
- Número de hijos................(V)
- Estado Civil................(V)
- Empresa donde laboran....................(C)

3. Identificar cada una de las variables de acuerdo a su naturaleza:
   - Número de trabajadores del "Banco de la Nación" (**Cuantitativa discreta**).
   - Distancia que puede recorrer los estudiantes que estudian en la Universidad Nacional de Jaén (**Cuantitativa continua**).
   - Las minas que explotan el mineral en el Perú (**Cualitativa nominal**).
   - Tiempo que demoran los operarios para encender la máquina (**Cuantitativa continua**).
   - Número de cursos que llevan los estudiantes en la maestría de "Gestión Ambiental" de la Universidad Nacional de Jaén (**Cuantitativa discreta**).
   - Temperatura en la costa medido en grados centígrados (**Cuantitativa continua**).

4. Determinar el nivel de medición de las variables:
   - Personas que ingresan al Banco de "Crédito" a realizar una transacción (ordinal, razón)
   - Temperatura de la ebullición del agua medida en la escala de Fahrenheit (intervalo)
   - Ingreso familiar de los docentes del colegio "Jaén de Bracamoros" (bajo, medio, alto) (ordinal)
   - Categoría ocupacional de los pobladores del distrito de Chirinos (empleado, obrero, independiente) (nominal)
   - Sueldo de los empleados del Banco "Continental" (más de 500 y menos de 2000) (razón)

5. Si a las unidades estadísticas E1 y E2 de una población se les asignan los valores 4 y 16 respectivamente en una escala determinada, esto indica que la escala utilizada está proporcionando una medición cuantitativa de alguna característica de E_1 y E_2

a. Nominal

b. Ordinal

c. De razón

Solución:

Se puede decir que E_1, es diferente que E_2

Además 4<16, en la diferencia se establece un orden

$\frac{16}{4} = 4$, se estable un a razón con los valores que toma la variable, asimismo de una diferencia y de un orden

- ❖ En la escala de medición nominal dos o más valores de una variable, se sabe que son válidas las relaciones de igualdad y no igualdad $(=)\ (\neq)$

$$Como\ \ E_1 = 5, E_2 = 20 \Rightarrow E_1 \neq E_2$$

- ❖ En la escala de medición ordinal dos o más valores de una variable, pueden ordenarse ya sea en forma ascendente o descendente, también son válidas las relaciones de igualdad $(=)$, no igualdad $(\neq)$, de orden $(\leq)$

$$Como\ \ E_1 = 5, E_2 = 20 \Rightarrow E_1 \leq E_2 ,\ 5 < 20$$

- ❖ En la escala de medición de razón dos o más valores de una variable, se pueden ordenar, también son válidas las relaciones de igualdad $(=)$, no igualdad $(\neq)$, de orden $(\leq)$ y todas las operaciones matemáticas.

En la escala de razón permanece invariante ante la transformación. $Y = ax + b$ donde a es una constante.

$Como\ E_1 = 5\ ,\ E_2 = 20$

$\Rightarrow 20 = a(5)$

$\Rightarrow a = \frac{20}{5} = 4$

$\therefore$ E_2 es igual a 4 veces la medida de E_1.

6. Cierta variable asigna los valores 1, 4 y 9 a las unidades estadísticas E_1, E_2, E_3, respectivamente en una escala de intervalos, si en la misma escala se le asigna 1 a E_1 y -8 a E_2 ¿Qué valor se le asigna a E_3?

   Solución:

❖ En la escala de medición por intervalos, se puede elegir una unidad de escala y comprobar las formas diferentes entre dos valores de la escala, es decir: dados X_1, X_2, X_3, se verifica la relación:

$$x_3 - x_1 = c(x_2 - x_1) \;\; o \;\; \frac{x_3 - x_1}{x_2 - x_1} = c$$

Donde C es constante.

Una escala de intervalo, permanece invariante ante la transformación. Donde a y b son constantes aleatorias.

Sea $E_1 = 1$ $E_2 = 4$ $E_3 = 9$

$E_1 = -1$ $E_2 = -8$ $E_3 = ?$

$$\Rightarrow \frac{9-1}{4-1} = c$$

$$\Rightarrow c = \frac{8}{3}$$

$$\Rightarrow \frac{8}{3} = \frac{E_3 - 1}{-8-1} = \frac{E_3 - 1}{-9}$$

Entonces: $\frac{8}{3}(-9) = E_3 - 1$

$$\Rightarrow -24 = E_3 - 1$$

$$\therefore \quad E_3 = -23$$

7. Al medir una determinada característica en una población, a las unidades estadísticas E1, E2 y E3 se les asignan los valores 2, 5 y 17, respectivamente, utilizando una escala A. Sin embargo, cuando se emplea una escala B, a E2 y E3 se les asignan los valores 5 y 29, respectivamente:

a. Se puede afirmar que A y B son la misma escala de razón.

b El valor de E_1 usando la escala B ¿se puede aseverar que ambas escalas son nominales? ¿Son ordinales? Es la misma de intervalo.

Solución:

a. Desarrollando la alternativa se tiene:

Sea:

$E_1=2$ $E_2=5$ $E_3=17$ escala A

E_1=? E_2=5 E_3=29 escala B.

$$x_3 = cx_3 \Rightarrow c = \frac{x_3}{x_2}, x_2 \neq 0$$

$$17 = c(5) \quad \Rightarrow c = \frac{17}{5}$$

$$29 = c(5) \quad \Rightarrow c = \frac{29}{5}$$

$\frac{17}{5} \neq \frac{29}{5}$ $\therefore$ No es la misma escala de razón

b. Desarrollando la alternativa (b) se tiene:

Si se sabe que son nominales entonces se tiene que:

$E_1 \neq E_2 \neq E_3$ ó $2 \leq 5 \neq 17$ ó $E_1 \neq 5 \neq 29$ (Nominal)

Si se sabe que son Ordinales entonces se tiene que:

$E_1 < E_2 \Rightarrow E_1 < 5$ (Ordinal)

Si se sabe que son de medición de intervalo entonces se tiene que:

$$\frac{E_3 - E_1}{E_2 - E_1} = c \qquad \Rightarrow c = \frac{17-2}{5-2} = \frac{15}{3} = 5$$

$$\frac{29 - E_1}{5 - E_1} = c \qquad \Rightarrow 29 - E_1 = 5(5 - E_1)$$

$$4 = -4E_1$$

$$\therefore E_1 = -1 \quad \text{(Intervalo)}$$

8. Si, $X_1 = 0$ e $Y_1 = 32$ son dos valores asignados para medir la temperatura y, $X_2 = 100$ e $Y_2 = 212$ corresponden a dos valores asignados a la temperatura de otro elemento. hallar la ecuación de regresión lineal

Solución:

Hallando la relación entre X e Y en la escala de intervalos:

$$Y = ax + b$$
$$32 = a(o) + b$$
$$\Rightarrow b = 32$$

$$Y = ax + b$$
$$212 = a(100) + 32$$
$$a = \frac{180}{100}$$
$$\Rightarrow a = \frac{9}{5}$$

9. Al medir cierta característica en una población, las escalas A y B asignan valores X e Y a un mismo elemento. Si la relación entre los valores

   $Y = \frac{X}{3} + 4$

   ¿Son ambas escalas A y B

   a. De intervalo
   b. de razón.

   Solución:

   a . $X = 3Y - 12$ Si es de intervalo

   b. $Y = \frac{x}{3} + 4$ Si es de razón

10. Clasifique las variables e indique el tipo de escala que están medidas las características:

   - Preferencia de un producto (cualitativa nominal)
   - Especialidad en la carrera de Ingeniería Mecánica y eléctrica (cualitativa nominal)
   - Número de ventas de una tienda comercial (cuantitativa discreta)
   - Número de turistas que arriban en San Ignacio (cuantitativa discreta)
   - Dirección (cuantitativa Discreta)
   - Rentas (cuantitativa continua)
   - Categorización en los alumnos de la Universidad Nacional de Jaén (Cualitativa ordinal)
   - Ingreso Mensual familiar promedio (cuantitativa continua)
   - Número de D.N.I (cuantitativa discreta)
   - Grupos Políticos (cualitativa nominal)
   - Variación en la demanda de un producto (cuantitativa continua)
   - Nivel de corrupción en el congreso (cualitativa ordinal)
   - Ingreso per. cápita de los ciudadanos de Jaén (cuantitativa continua)
   - Rendimiento laboral (cuantitativa continua)
   - Vida útil (cuantitativa continua)
   - Número de empleados de una fábrica (cuantitativa discreta)
   - Número de goles marcados por un equipo de fútbol (cuantitativa discreta)
   - La velocidad de un vehículo (cuantitativa continua)

- El estado civil (cualitativa nominal)
- El color de ojos de los empleados (cualitativa nominal)
- La tienda de ropa preferida por los habitantes de una ciudad (Cualitativa nominal)
- Longitud de tornillos producidos en una empresa (cuantitativa continua)
- Número de empresas eléctricas visitadas por los estudiantes (cuantitativa discreta)
- Tiempo para arreglar un motor eléctrico (cuantitativa continua)
- Número de motores (cuantitativa discreta)
- Número de plantas en una huerta (cuantitativa discreta)
- Temperaturas de la ciudad de Jaén: 19°C, 25°C, 26°C (cuantitativa continua)
- Pesos de las varillas de construcción: 1.53 kg, 2.43 kg, 4.31kg, etc. (cuantitativa continua)
- La profesión de un grupo de docentes: Contador, Abogado, Administrador, Educador, Ingeniero (cualitativa nominal)
- La marca de un producto (cualitativa nominal)
- Calidad de un material (cualitativa ordinal)
- El lugar de nacimiento (cualitativa nominal)
- Nivel de instrucción (cualitativa ordinal)
- Nivel Socioeconómico (cualitativa ordinal)

7.2 CAPÍTULO 2

1. Construir el número de intervalos:

   a. Datos enteros $X_{min} = 10,\quad X_{max} = 74,\qquad m = 8$

   b. Datos con dos decimales $X_{min} = 2.75,\quad X_{max} = 4.06,\qquad m = 8$

   c. Datos con tres decimales $X_{min} = 0.282,\quad X_{max} = 0.655,\qquad m = 6$

   Solución:

   a. Rango: R = 74 – 10 = 64

   m = 8

   $$C = \frac{64}{8} = 8$$

   Sus intervalos son:

10 - 18	42 - 50
18 - 26	50 - 58
26 - 34	58 - 66
34 - 42	66 - 74

   b. Rango: R = 4.06 – 2.75 = 1.31

   m = 7

   $$C = \frac{1.31}{7} = 0.19$$

   Sus intervalos son:

2.74 - 2.93	3.50 - 3.69
2.93 - 3.12	3.69 - 3.88
3.12 - 3.31	3.88 - 4.07
3.31 - 3.50	

   c. Rango: R = 0.655 – 0.282 = 0.373

   m = 6

   $$C = \frac{0.373}{6} = 0.063$$

   Sus intervalos son:

   0.282 - 0.345 0.471 - 0.534

0.345	-	0.408	0.534	-	0.597
0.408	-	0.471	0.597	-	0.660

2. Las inversiones anuales en miles de soles, de un conjunto de 40 empresas de la ciudad de Jaén durante el año 2024 fueron:

31	17	27	20	28	10	34	25	4	24
15	39	18	30	41	26	12	46	18	23
36	19	29	37	33	27	27	24	26	31
25	28	33	28	22	23	31	29	35	21

a. Construir la tabla e interprete.
b. Determinar el porcentaje de empresas con una inversión se encuentre entre 14 mil y menos de 20 mil dólares.

Solución:

Variable en estudio: Inversión

Tipo de variable: Cuantitativa Continúa

Construcción de las tablas de frecuencias

Rango

$$R = X_{max} - X_{min}$$

$$R = 46 - 4 = 42$$

Numero de intervalos

$$m = 1 + 3.33\log{(n)}$$

$$m = 7$$

Amplitud interválica

$$C = \frac{R}{m}$$

$$C = \frac{42}{7} = 6$$

Figura 28

Ingreso de la variable al software Jamovi

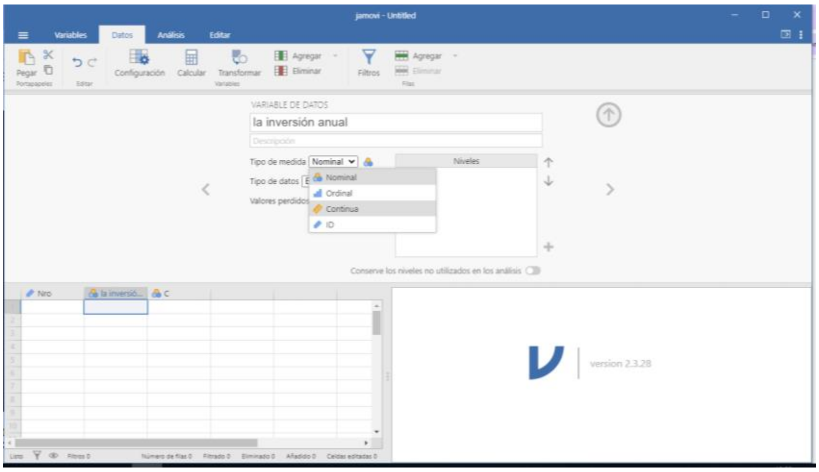


Figura 29

Ingreso de la variable dar clic en tabla de frecuencias

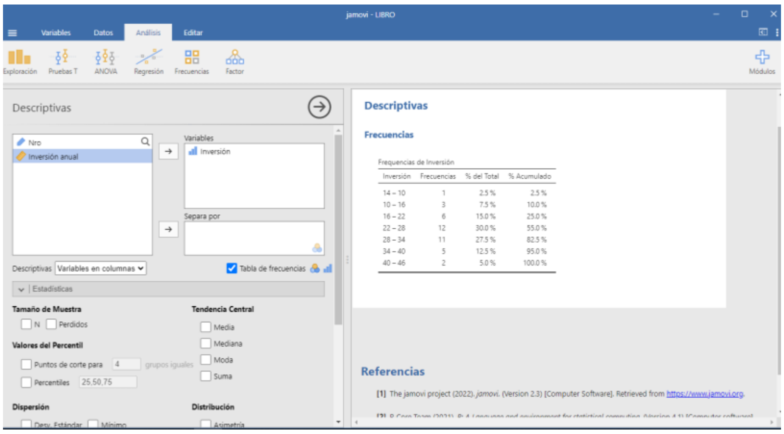
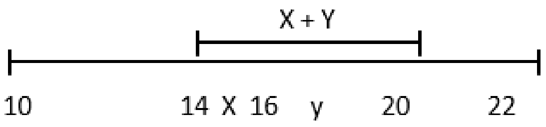


***Tabla* 49**

Distribución de la inversión anual, en miles de dólares realizada por 40 pequeñas empresas de la ciudad de Jaén durante el año 2024

Inversión	Frecuencias	% del Total	% Acumulado
10 – 14	1	2.5 %	2.5 %
14 – 16	3	7.5 %	10.0 %
16 – 22	6	15.0 %	25.0 %
22 – 28	12	30.0 %	55.0 %
28 – 34	11	27.5 %	82.5 %
34 – 40	5	12.5 %	95.0 %
40 – 46	2	5.0 %	100.0 %

Determinando el porcentaje de empresas con una inversión entre 14 mil y menos de 20 *mil dólares.*



Trabajando con regla de tres simple se tiene:

6C	3	6 C	6
2C	X	4 C	Y
$\Rightarrow X = 1$		$\Rightarrow Y = 4$	

Entonces se tiene:

X + Y = 5

Calculando el porcentaje se tiene:

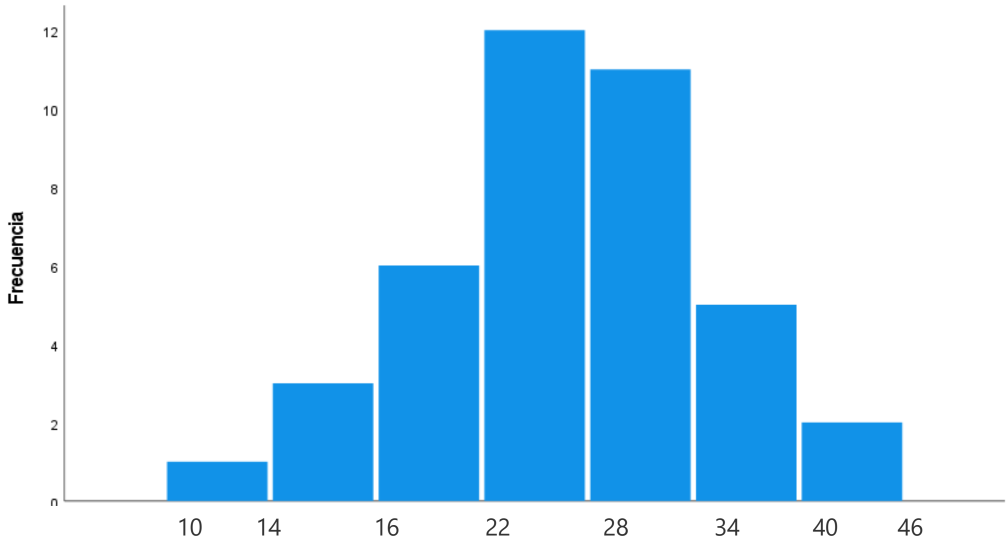
$$\% = \frac{5}{40}(100) = 12.5$$

Otra forma de calcular sería:

$$\left[\frac{(16-14)}{(16-10)}(0.075) + \frac{(16-14)}{(16-10)}(0.150)\right] \times 100 = 12.5\%$$

Figura 30

Distribución de la inversión anual, en miles de dólares realizada por 40 pequeñas empresas de la ciudad de Jaén durante el año 2024



3. Se toma nota de la cantidad de tiempo en minutos que 30 estudiantes de la Universidad Nacional de Jaén, emplean durante el segundo semestre del 2023 para completar una tarea, y estos son los datos:

21.3	15.8	18.4	22.7	19.6	15.8	26.4	17.3	16.2	23.9
26.8	22.7	18.0	20.5	11.0	18.5	23.0	24.6	20.1	16.2
08.3	21.9	12.3	22.3	13.4	17.9	12.2	13.4	15.1	19.1

a. Elaborar una tabla que muestre la distribución de frecuencias dividido los datos en 6 intervalos de igual amplitud.
b. Determinar el tiempo que corresponde al 25% inferior de las tareas en función de la tarea.

Solución:

Calculamos

Construcción de las tablas de frecuencias

Rango

$R = 26.8 - 8.3 = 18.5$

Numero de intervalos

$m = 6$

Amplitud interválica

$$C = \frac{18.5}{6} = 3.1$$

Tabla 50

Distribución del tiempo (en minutos) de los estudiantes de la Universidad Nacional de Jaén en el segundo semestre del 20223 para completar una tarea

Tiempo	Frecuencias	% del Total	% Acumulado
8.3 – 11.4	2	6.7 %	6.7 %
11.4 – 14.5	4	13.3 %	20.0 %
14.5 – 17.6	6	20.0 %	40.0 %
17.6 – 20.7	8	26.7 %	66.7 %
20.7 – 23.8	6	20.0 %	80.0 %
23.8 – 26.9	4	13.3 %	100.0 %

Del intervalo (8.3 11.4) se tiene el 10%

Del intervalo (11.4 14.5) se tiene el 14%

Faltaría el 1%. Calculando se tiene:

17% 3.1C

1% X

X = 0.18235C

Ahora se tiene que sumar las amplitudes que se encuentran por debajo del 25%. (Y)

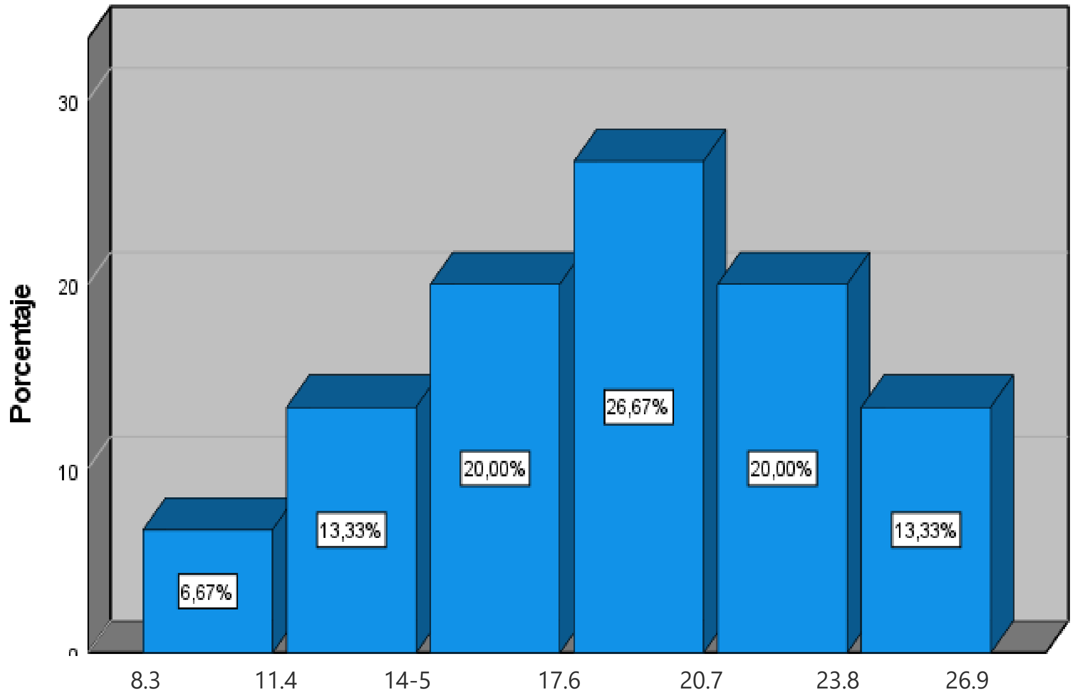
Y = 3.1C +3.1C + 0.18235C = 6.38C

Tomando el tiempo de inicio (8,3) se tiene:

Y = 8.3 + 6.38 =14.68

Figura 31

Distribución del tiempo (en minutos) de los estudiantes de la Universidad Nacional de Jaén en el segundo semestre del 20223 para completar una tarea



4. Los resultados de notas del examen parcial de matemáticas se presentan mediante la distribución de frecuencias.
   a. Agregar los valores faltantes en la distribución de frecuencias.
   b. Representar gráficamente la ojiva mayor.

c. Estimar el porcentaje de notas se sitúan aproximadamente dentro del rango [8.14]

Tabla 51

Notas del examen parcial de matemáticas

Intervalo	Marca de Clases	Frecuencia Relativo	Frec. Relativa acumulada
[– [	-	0.15	-
6 – [	-	-	0.45
[– [	-	-	0.70
[– [	13.5	-	-
[–]	-	0.10	-
Total	-	1.00	-

Solución:

Hallando la amplitud y el primer intervalo:

$$Y_i + 2.5C = 13.5$$
$$6 + 2.5C = 13.5$$
$$2.5C = 7.5$$
$$C = \frac{7.5}{2.5} = 3$$
$$Y_i - C = Y_0$$

Reemplazando se tiene:

6 – 3 = 3 $\Rightarrow C = 3$

❖ Hallando las frecuencias relativas:

$h_1 = H_1 = 0.15$

$H_1 + h_2 = H_2$

$0.15 + h_2 = 0.45$

$h_2 = 0.45 - 0.15$

$h_2 = 0.30$

$H_2 + h_3 = H_3$

$0.45 + h_3 = 0.70$

$h_3 = 0.70 - 0.45$

$h_3 = 0.25$

Además:

$h_1 + h_2 + h_3 + h_4 + h_5 = 1.00$

$0.15 + 0.30 + 0.25 + h4 + 0.10 = 1$

$h_4 = 1 – 0.80$

$h_4 = 0.20$

Ahora completando el cuadro con los resultados trabajados se tiene:

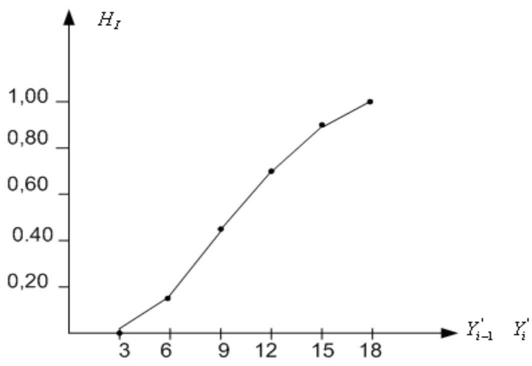
Tabla 52

Notas del examen parcial de matemáticas

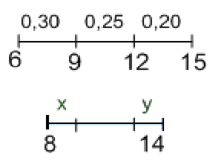
Y_{i-1} Y_i	Y_i	h_i	H_i
3 – 6	4.5	0.15	0.15
6 – 9	7.5	0.30	0.45
9 – 12	10.5	0.25	0.70
12 – 15	13.5	0.20	0.90
15 – 18	16.5	0.10	1.00
Total	-	1.00	-

Figura 32

Ojiva mayor



Sea “Z” el porcentaje de notas que se sitúan aproximadamente dentro del rango [8 ; 14)



Calculando

3 C 0.30 | 3 C 0.20
C X | 2 C Y

$\Rightarrow$ X = 0.10 | $\Rightarrow$ Y = 0.133

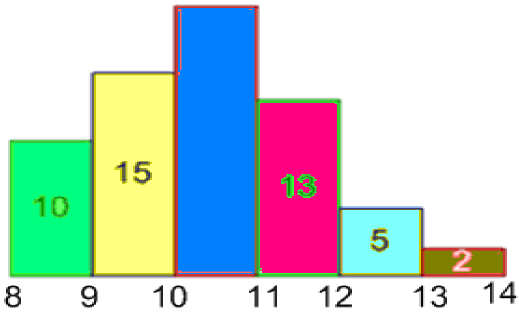
Por lo tanto:

Z = 0.10 + 0.133 + 0.25 = 0.483 = 48.3%.

5. ¿Qué porcentaje de las 65 personas que realizaron una prueba de aptitud utilizó un tiempo entre 9 y 11,5 minutos, representada en el histograma?

Figura 33

Tiempo que utilizan las 65 personas en una prueba de aptitud



Solución

Del enunciado se tiene que son 65 personas que realizaron la prueba de aptitud, entonces:

n = 65

Pero n = $f_1 + f_2 + f_3 + f_4 + f_5 + f_6$

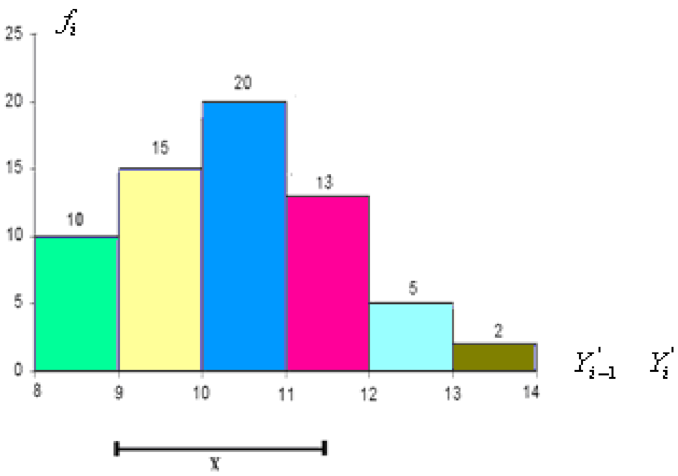
Reemplazando se tiene:

n = 10 +15 + X + 13+ 5 + 2 = 65

X = 20

Figura 34

Tiempo que utilizan las 65 personas en una prueba de aptitud



C 13
0.5C Y
⇒ Y = 6.5

Por lo tanto:
$\therefore z = (15 + 20 + 6.5) / 65 = 0.6385$
Z = 63.85%.

6. En una empresa, se registran sueldos mínimos y máximos de 200 empleados, que oscilan entre S/. 150 y S/. 300. Estos sueldos se organizan en una distribución de frecuencias con 5 intervalos de igual amplitud. Se tiene información específica sobre cuántos empleados gana en diferentes rangos salariales: 20 empleados ganan al menos S/.150, pero menos de S/.180, 60 ganan menos de S/.210, 110 ganan menos de S/. 240, 180 ganan menos de S/ 270. y el 10% restante de empleados gana a lo más S/. 300; reconstruir la distribución de frecuencias con base en estos resultados.

Solución:

n = 200
$X_{max} = 300$ $X_{min} = 150$
m = 5

$$C = \frac{300 - 150}{5} = 30$$

Hallando las frecuencias:

$f_1 = 20,\quad f_1 + f_2 = 60 \quad \Rightarrow f_2 = 60 - 20 = 40$

$f_1 + f_2 + f_3 = 110 \quad \Rightarrow f_3 = 110 - 20 - 40 = 50$

$f_1 + f_2 + f_3 + f_4 = 180 \quad \Rightarrow f_4 = 180 - 20 - 40 - 50 = 70$

$El\ 10\%(200) = 20$

Tabla 53

Registro de los sueldos de los empleados de una empresa

Y_{i-1} Y_i	Y_i	f_i	F_i
150 - 180	165	20	20
180 – 210	195	40	60
210 – 240	225	50	110
240 – 270	255	70	180
270 – 300	285	20	200
Total	-	200	-

7. Calcular la distribución de frecuencias absolutas de un conjunto de datos sobre la vida útil de una batería, que se dividió en 5 intervalos de igual tamaño, con frecuencias relativas acumuladas de 0.10, 0.25, 0.55, 0.80 y 1.00. Además, se sabe que la tercera frecuencia absoluta acumulada es igual a 11, que la segunda marca de clase es 6 y que el límite inferior del cuarto intervalo es 12.

Solución:

Del problema, los siguientes datos se tiene:

$n = 5$

$H_1 = 0.10$	$H_2 = 0.25$
$H_3 = 0.55$	$H_4 = 0.80$
$H_5 = 1.00$	$Y_2 = 6$
$F_3 = 11$	$Y_4' = 12$

Ahora hallando los intervalos:

- De $Y_4' = 12$

$$Y_0 + 3C = 12$$
$$Y_0 = 12 - 3C \ldots\ldots(1)$$

- De $Y_2 = 6$

$$2Y_0 + 3C = 12$$
$$2Y_0 + 3C = 12 \ldots\ldots(2)$$

Reemplazando (1) en (2) se tiene:

$$2 \times (12 - 3C) + 3C$$
$$24 - 6C + 3C = 12$$
$$3C = 12$$
$$C = \frac{12}{3} = 4$$
$$Y_0 = 12 - 3(4) = 0$$
$$\therefore Y_0 = 0$$

❖ Hallando las frecuencias:

$$\frac{F_i}{n} = H_i \quad \Rightarrow n = \frac{F_i}{H_i}$$

$$n = \frac{F_i}{H_i} = \frac{11}{0.50} = 20$$

Entonces:

$F = n \times H_i$

$F_1 = 20(0.10) = 2$

$F_2 = 20(0.25) = 5$

Además:

$F_1 = f_1 = 2$

$F_2 = f_1 + f_2$

$5 = 2 - f_2$

$f_2 = 3$

Tabla 54

Distribución de frecuencias de la vida útil de una batería

$Y_{i-1} \quad Y_i$	Y_i	f_i	h_i	F_i	H_i
0 – 4	2	2	0.10	2	0.10
4 – 8	6	3	0.15	5	0.25
8–12	10	6	0.25	11	0.55
12–16	14	5	0.25	16	0.80
16–20	18	4	0.20	20	1.00
Total	**-**	**20**	**1.00**	**-**	**-**

7.3 CAPÍTULO 3

1. Una fábrica de bebidas gaseosas está considerando introducir un nuevo sabor en el mercado. Para evaluar la aceptación de este nuevo sabor, se llevó a cabo un estudio en el año 2024 en el cual se encuestó a una muestra de 30 niños del colegio "Sagrado Corazón de Jesús". En la investigación se utilizó una escala de calificación de 10 puntos, para medir el nivel de aceptación, los puntajes obtenidos de los 30 fueron:

2	8	7	4	10	5	7	7	3	2
6	8	5	5	8	6	6	7	7	7
6	6	4	7	6	7	6	2	8	7

Encontrar:

a. La medida aritmética
b. La mediana
c. El cuantil tres
d. Decil cinco
e. El percentil setenta y cinco
f. Moda

Solución:

Identificando la variable, tipo de variable y su escala tenemos:

Variable : Nivel de aceptación

Tipo de variable : Cuantitativa discreta/ Ordinal

Tabla 55

Distribución de bebidas gaseosas según el nivel de aceptación del nuevo sabor de una marca de gaseosa en los niños del colegio "Sagrado Corazón de Jesús" en el año 2024

Nivel de aceptación	Frecuencias	% del Total	% Acumulado
2	3	10.0 %	10.0 %
3	1	3.3 %	13.3 %
4	2	6.7 %	20.0 %
5	3	10.0 %	30.0 %
6	7	23.3 %	53.3 %
7	9	30.0 %	83.3 %
8	4	13.3 %	96.7 %
10	1	3.3 %	100.0 %

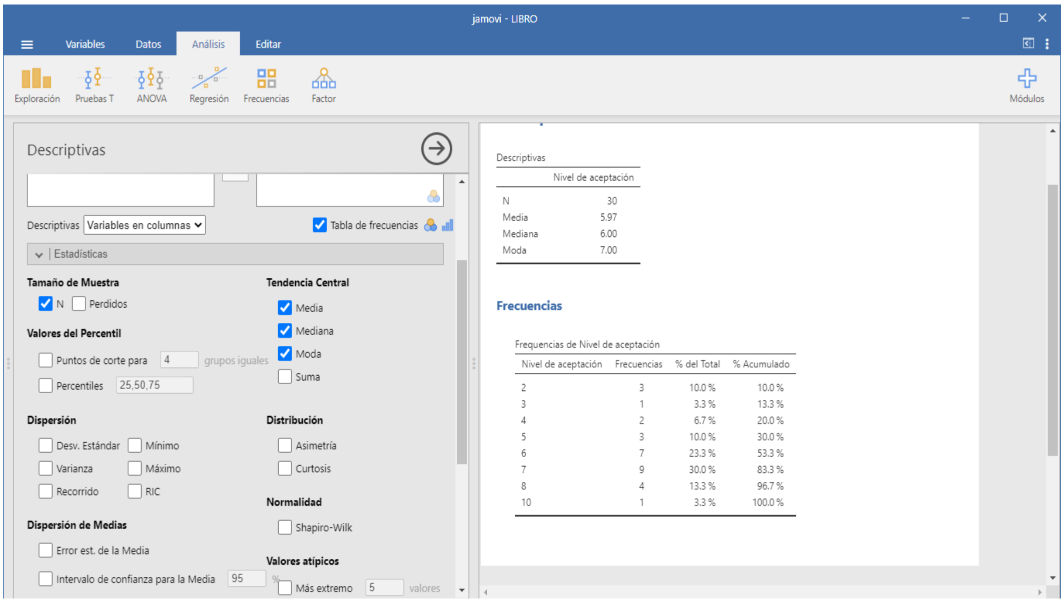
Tabla 56

Resultados descriptivos de la distribución de bebidas gaseosas según el nivel de aceptación del nuevo sabor de una marca de gaseosa en los niños del colegio "Sagrado Corazón de Jesús" en el año 2024

	Nivel de aceptación
N	30
Media	5.97
Mediana	6.00
Moda	7.00

Figura 35

Ventana de trabajo para resultados descriptivos y tablas de distribución de frecuencias



Calculando el promedio aritmético se tiene:

$$\bar{x} = \frac{\sum_{i=1}^{30} y_i f_i}{30} = \frac{179}{30} = 5{,}96$$

$$\bar{x} = 6$$

Interpretación

El promedio de aceptación por los niños del colegio "Sagrado Corazón de Jesús" de la nueva marca de gaseosas según su sabor se encuentra en la escala de 6 puntos.

- Hallando la mediana se tiene:

  Para calcular la mediana, se obtiene la suma de todos los datos y se divide entre dos.

  $\frac{n}{2} = \frac{30}{2} = 15$

  Cumple la primera condición.

  $Me = Y_{j-1} = 6$

  Interpretación

  El 50% de aceptación por los niños del colegio "Sagrado Corazón de Jesús" de la nueva marca de gaseosas según su sabor es menor o igual a 6 puntos, mientras que el otro 50% supera a dicho número.

- Determinando el cuartil tres se tiene: K =3

  $\frac{nk}{4} = \frac{30 \times 3}{4} = 22.5$

  Cumple la primera condición:

  $Q_3 = Y_{j-1} = 7$

  Interpretación

  El 75% de aceptación por los niños del colegio "Simón Bolívar" de la nueva marca de gaseosas según su sabor es menor o igual a 7 puntos mientras que el otro 25% supera a dicho número.

- Calculando el Decil cinco tenemos: K = 5

  $\frac{nk}{10} = \frac{30 \times 5}{10} = 15$

  Cumple la primera condición:

  $D_5 = Y_{j-1} = 6$

  Interpretación

  El 50% de aceptación por los niños del colegio "Simón Bolívar" de la nueva marca de gaseosas según su sabor es menor o igual a 6 puntos mientras que el otro 50% supera a dicho número.

- Hallando el percentil sesenta y cinco K = 65

  $$\frac{nk}{100} = \frac{30 \times 65}{100} = 19.5$$

  Cumple la primera condición.

  $P_{65} = Y_{j-1} = 7$

  Interpretación

  El 65% de aceptación por los niños del colegio “Simón Bolívar” de la nueva marca de gaseosas según su sabor es menor o igual a 7 puntos mientras que el otro 35% supera a dicho número.

- Para hallar la moda debemos tomar la mayor frecuencia en este caso es 9 por lo tanto $M_o = 7$.

  Interpretación

  La mayoría de los niños del colegio “Simón Bolívar” aceptan la nueva bebida gaseosa en escala de 7 puntos.

2. A un analista le interesa supervisar los gastos (en miles de soles) que realiza la empresa en publicidad para promover un producto. La analista registra los gastos de 48 meses seleccionados al azar de la base de datos.

30,	10	15	20	35	40	50	80	90	100
110	70	41	36	55	58	55	102	55	78
46	28	19	25	15	21	25	40	55	75
85	95	105	95	75	46,	41	60	93	59
107	60	84	51	33	24	69	84		

Encontrar:

a. La media aritmética
b. La mediana
c. El cuantil uno
d. Decil tres
e. El percentil cincuenta
f. Moda

Solución:

Variable : Gastos en la publicidad

Tipo de variable : Cuantitativa continúa/ Razón

Para poder trabajar se debe primero construir el cuadro de frecuencias absolutas simples y acumuladas.

Rango:

R = 110 – 10 = 100

Número de intervalo.

m = 1 + 3.33 log (48) = 7 intervalos

Amplitud interválica.

$$C = \frac{100}{7} = 14.38 \approx 15$$

Hallando la nueva amplitud:

$$C' = \frac{105}{7} = 15$$

Encontrado la diferencia del rango con el nuevo rango:

D = R’ – R

D = 105 – 100 = 5

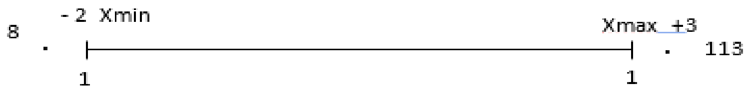


Tabla 57

Gastos de la empresa en publicidad para promover un producto

F	Frecuencias	% del Total	% Acumulado
23 – 38	8	16.7 %	16.7 %
08 – 23	6	12.5 %	29.2 %
38 – 53	9	18.8 %	47.9 %
68 – 83	6	12.5 %	60.4 %
83 – 98	7	14.6 %	75.0 %
98 – 113	5	10.4 %	85.4 %
53 – 68	7	14.6 %	100.0 %

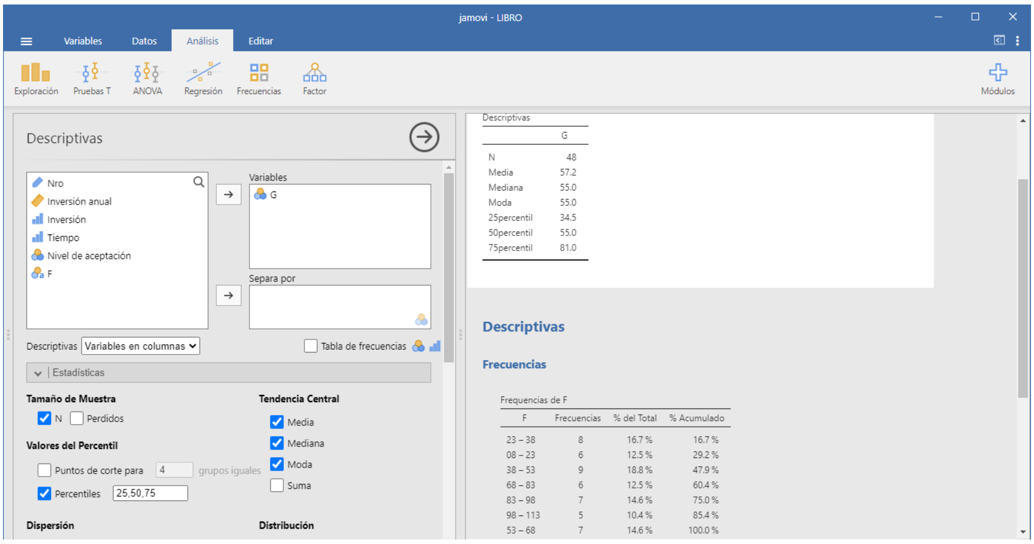
Tabla 58

Medidas descriptivas de Gastos de la empresa en publicidad para promover un producto

	Gastos
N	48
Media	57.2
Mediana	55.0
Moda	55.0
25percentil	34.5
50percentil	55.0
75percentil	81.0

Figura 36

Ventana de trabajo para resultados descriptivos y tablas de distribución de frecuencias



- Hallando el promedio aritmético se tiene:

$$\bar{x} = \sum Y_i h_i = 57.811$$

Interpretación

El promedio de gastos que realizó la empresa hace dos años en publicidad es de 57.811 soles.

- Hallando la mediana se tiene:

Para ubicar la mediana dividimos el total de los datos entre dos.

$$\frac{n}{2}=\frac{48}{2}=24$$

Cumple la segunda condición:

$$M_e = 53 + 15\left(\frac{24-23}{7}\right) = 55.143$$

Interpretación:

El 50% de los gastos que realizó la empresa hace dos años en publicidad es de 55.143 soles mientras que el otro 50% supera a dicho gasto.

- Hallando el cuartil uno tenemos: K = 1

Para ubicar el cuartil uno se selecciona el total de datos y se divide entre cuatro.

$$\frac{nk}{4}=\frac{48\times 1}{4}=12$$

Cumple la segunda condición:

$$Q_1 = 23+15\left(\frac{12-6}{8}\right)=34.25$$

Interpretación:

El 25% de los gastos que realizó la empresa hace dos años en publicidad es de 34.25 soles mientras que el otro 75% supera a dicho gasto.

- Hallando el Decil cinco tenemos: K = 5

Para ubicar el cuartil uno se selecciona el total de los datos y se divide entre diez.

$$\frac{nk}{10}=\frac{48\times 3}{10}=14.4$$

Cumple la segunda condición:

$$D_3 = 38+15\left(\frac{14.4-14}{9}\right)=38.667$$

Interpretación:

El 30% de los gastos que realizó la empresa hace dos años en publicidad es de 38.667 soles mientras que el otro 70% supera a dicho gasto

- Hallando el percentil cincuenta K =50

  Para ubicar el cuartil uno se toma el total de los datos y se divide entre cien.

$$\frac{nk}{100} = \frac{48 \times 50}{100} = 24$$

  Cumple la segunda condición:

$$P_{50} = 53 + 15\left(\frac{24-23}{7}\right) = 55.143$$

  Interpretación:

  El 50% de los gastos que realizó la empresa hace dos años en publicidad es de 55.143 soles mientras que el otro 50% supera a dicho gasto.

- Para hallar la moda debemos tomar la mayor frecuencia en este caso es 9 por lo tanto:

$$M_o = y'_{i-1} + C\left(\frac{\Delta_1}{\Delta_1 + \Delta_2}\right)$$

  donde:

  $\Delta_1 = 9 - 8 = 1$

  $\Delta_2 = 9 - 7 = 1$

$$M_0 = 38 + 15\left(\frac{1}{1+1}\right) = 45.5$$

  Interpretación

  La mayoría de gastos que realizó la empresa hace dos años en publicidad es de 45.500 soles.

3. En estudios forestales, se mide la cantidad de árboles por unidad de superficie. Se sabe que distribución de frecuencias simétricas de 5 intervalos de igual amplitud, además se tiene los siguientes datos, calcular la mediana.

   N = 110. $f_4 - f_5 = 10$, $f_4 - f_3 - f_1 = 0$ $f_3 = 10$

   $y'_4 f_4 = 975$ $y'_0 = 12{,}5$

   Solución:

   El problema nos indica que la distribución es simétrica por lo tanto se tiene:

   $f_1 = f_5$

$f_2 = f_4$

También se sabe que:

$f_3 = 10$

Se sabe que $\sum f_i = n = 110$ entonces trabajando esta parte se tiene:

$f_1 + f_2 + f_3 + f_4 + f_5 = 110$

$f_1 + f_2 + f_3 + f_2 + f_1 = 110$

$2f_1 + 2f_2 + f_3 = 110$

$2f_1 + 2f_2 + 10 = 110$

$2f_1 + 2f_2 = 100$

$f_1 + f_2 = 50$ .... (*)

Se tiene como dato que:

$f_4 – f_5 = 10$

$f_4 + f_5 = -10$ .... (**)

De * y ** se tiene:

$f_1 + f_2 = 50$

$f_4 + f_5 = -10$

De esto se tiene: $2f_1 = 40$

$f_1=20$ $f_5 = 20$

$f_1=30$ $f_4 = 30$

De los datos se podrá trabajar:

$$Y_4^{'} f_4 = 975$$

$$Y_4^{'}(30) = 975$$

$$Y_4^{'} = \frac{975}{30}$$

$$Y_4^{'} = 32.5$$

Se sabe que:

$$Y_4^{'} = Y_0^{'} + 4C = 32.5$$

$12.5 + 4c = 32.5$

$4c = 32.5 – 12.5$

$4c = 20$ $c = 5$

Tabla 59

Cantidad de árboles por unidad de superficie

$[LI - LS$	Y_i	f_i	F_i
12.5 – 17.5	15	20	20
17.5 – 22.5	20	30	50
22.5 – 27.5	25	10	60
27.5 – 32.5	30	30	90
32.5 – 37.5	35	20	110
TOTAL		110	

Calculando la mediana:

$$\frac{n}{2} = \frac{110}{2} = 55$$

En este caso se da: $F_{i-1} > \frac{n}{2}$

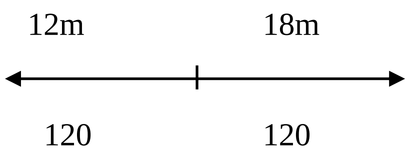
$$M_e = 22.5 + 5\left(\frac{55-50}{10}\right) = 25$$

Interpretación:

El 50% de la cantidad de árboles por unidad de superficie es menor o igual a 25 árb/km² mientras que el otro 50% supera a dicha superficie.

4. Calcular la tasa de productividad diaria promedio para el trabajo completo, considerando que un grupo de trabajadores inicialmente excava los primeros 120 metros a una velocidad de 12 metros por día, mientras que los siguientes 120 metros los excavan a una velocidad de 18 metros por día.

Solución



$$M_a = \frac{n}{\sum \frac{1}{x_i}} = \frac{2}{\frac{1}{12}+\frac{1}{18}} = \frac{72}{5} = 14.4$$

Interpretación

El promedio diario de productividad del trabajo es de 14.4 metros.

5. Un estudio se llevó a cabo para analizar los ingresos de 240 trabajadores de una empresa. Se ha confirmado que la distribución es simétrica y contiene 6 intervalos. Además, se disponen de ciertos datos relevantes:

   C = 5 Mo = 20 F_4 = 180 f_6 = 25

   a. Hallar el cuartil uno
   b. El 30% de los empleados con ingresos más altos tienen un ingreso mínimo
   c. ¿Qué porcentaje de empleados se verá favorecida si el sindicato solicita logra establecer un salario mínimo de s/.12. 00, en el nuevo acuerdo colectivo?
   d. ¿Qué fracción de los empleados se considera altamente calificada si ganan más de s/. 28.00 por día?
   e. El ingreso medio por día de los trabajadores de la empresa.
   f. El ingreso mediano de la distribución.

Solución

Primero se debe trabajar con los datos disponibles para construir la distribución. La distribución es simétrica:

$$f_1 = f_6 = 25$$

$$f_2 = f_5$$

$$f_3 = f_4$$

Se tiene 6 intervalos (m = 6 es par) y se puede obtener el límite inferior a partir de este dato y de la moda.

$$M_o = y'_{\frac{m}{2}} = y'_{\frac{6}{2}} = y'_3 = 20$$

Donde:

$y'_3 =$ límite superior de la tercera clase

Pero: C =5 entonces se podría construir los intervalos.

$$y'_2 = 20 - 5 = 15 \qquad y'_4 = 20 + 5 = 25$$

$$y'_1 = 15 - 5 = 10 \qquad y'_5 = 25 + 5 = 30$$

$$y'_0 = 10 - 5 = 5 \qquad y'_6 = 30 + 5 = 35$$

Construyendo las frecuencias se tiene: $F_6 = 240$

Pero

$F_6 = F_5 + f_6$ 240 = F_5 + 25 F_5 = 215

Como dato tenemos: $F_4 = 180$ $F_5 = 215$

entonces:

$$F_5 = F_4 + f_5$$
$$f_5 = 215 - 180$$
$$f_5 = 35 \quad \Rightarrow f_2 = 35$$

También se sabe que:

$$\sum f_i = n = 240$$
$$f_1 + f_2 + f_3 + f_4 + f_5 + f_6 = 240$$
$$25 + 35 + 2f_3 + 25 + 35 = 240$$
$$2f_3 = 120$$
$$f_3 = 60 \quad \Rightarrow f_4 = 60$$

Tabla 60

Ingresos de los trabajadores de una empresa

[	⟩	f_i	F_i
5	10	25	25
10	15	35	60
15	20	60	120
20	25	60	180
25	30	35	215
30	35	25	240
TOTAL		240	-

- Hallando el Q_1

  $$\frac{nk}{4} = \frac{240 \times 1}{4} = 60$$

  Cumple la primera condición:

  Entonces

$$Q_1 = 15 + 5\left(\frac{60 - 25}{60}\right) = 17.92$$

Interpretación:

El 25% de los ingresos de los trabajadores es menor o igual a 17.92 soles mientras que el otro 75% supera a dicho ingreso.

- El que supera al 30% es el P_{70} por lo tanto se tiene que calcular el percentil setenta. Para ubicar el cuartil uno se selecciona el total de los datos y se lo divide entre cien.

$$\frac{nk}{100} = \frac{240 \times 70}{100} = 168$$

Cumple la segunda condición:

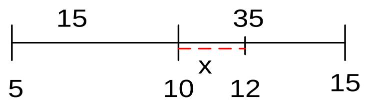
$$P_{50} = 20 + 5\left(\frac{168 - 120}{60}\right) = 24$$

Interpretación:

El 70% de los ingresos de los trabajadores es menor o igual a 24.00 soles mientras que el otro 30% supera a dicho gasto.

- Se beneficiarán todos los trabajadores que tengan un ingreso por debajo de 1200

Para esto se calcula cuantos trabajadores se encuentran por debajo de su sueldo.



5C	35
2C	X

$$X = \frac{2 \times 35}{5} = 14$$

Por lo tanto, los trabajadores que tengan un ingreso por debajo de 1 200 es:

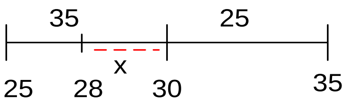
Y = 25 + 14 = 39 trabajadores

Calculando el porcentaje de trabajadores que tengan un ingreso por debajo de 1 200

$$p = \frac{29}{240} \times 100 = 12.083\%$$

Con el nuevo pacto colectivo se beneficiarán el 12.08% de los trabajadores.

- El número de trabajadores que pertenecen al grupo muy calificado



5C	35
2C	X

$$X = \frac{2 \times 35}{5} = 14$$

Los trabajadores que pertenecen al grupo muy calificado

Y = 25 + 14 = 39 trabajadores

Calculando el porcentaje de trabajadores que tengan un ingreso por debajo de 1200

$$p = \frac{39}{240} \times 100 = 16.25\%$$

Con el nuevo pacto colectivo se beneficiarán el 16.25% de los trabajadores.

Como la distribución es simétrica entonces $\bar{x} = M_e = M_o$

Entonces $\bar{x} = 20$

Por lo anterior se tiene que la mediana es: $M_e = 20$

6. Los costos de fabricación en soles, de diez objetos son los siguientes:

9.35 9.46 9.20 9.80 9.77 9.00 9.99 9.36 9.50 9.60

Calcular el beneficio promedio por cada objeto, dado que el precio de venta de cada uno es tres veces su costo de fabricación, reducido en 5 soles

Solución

Calculando el promedio se tiene:

$$\bar{x} = \frac{\sum x_i}{n} = \frac{9.35 + 9.46 + \ldots + 9.50 + 9.60}{10} = 9.503$$

El precio de venta está dado por la ecuación:

$$Y_i = 3X_i - 5$$

Calculando el promedio de del precio de venta:

$$\bar{Y} = 3\bar{X} - 5$$

$$\bar{Y} = 3(9.503) - 5 = 23.509$$

Calculando la utilidad media de cada objeto se tiene:

$$\bar{U} = \bar{Y} - \bar{X}$$

$$\bar{U} = 23.509 - 9.503 = 14.006$$

7. Se ha realizado un estudio en la "Universidad Nacional de Jaén" con el personal administrativo, para ello se toma una muestra de cinco secretarias para medir en que velocidad escriben y así ver si son eficientes. Los datos fueron: 50, 63, 28, 45, 53 palabras por minuto. Si todos ellos escriben el mismo texto, determine la velocidad media.

Solución

Como el tiempo y la velocidad están en relación inversa, la media armónica dará el promedio correcto

$$\bar{X}_A = \frac{5}{\frac{1}{50} + \frac{1}{63} + \frac{1}{28} + \frac{1}{45} + \frac{1}{53}} = 44.374$$

Interpretación

La velocidad media en que escriben las secretarias es de 44 palabras por minuto.

8. Los salarios aumentaron en los últimos 4 años en 28%, 23%, 27%, 25% determinar:
   a. Determinar la tasa de crecimiento en los últimos 5 años
   b. Determinar e interpretar, la media anual de crecimiento
   c. Calcular, la media geométrica anual de crecimiento

Solución

Suponga que el salario que percibe en es de s/. 100 entonces se tiene:

- Después de revisar las tasas de crecimiento registradas en los últimos cuatro años según la tabla, calcular la tasa de crecimiento para el quinto año.

Tabla 61

Tasa de crecimiento de los salarios en los últimos 5 años

Años	Aumento en los últimos años (%)	Salarios	Tasas
1	28	100 + 0.28(100) =128	128/100=1.28
2	23	128+ 0.23(128) =157.44	157.44/100=1.5744
3	27	157.44+0.27 (15744) = 199.949	199.949/100=1.999
4	25	199.949+0.25(199.949)=249.936	249.936/100=2.499

.

El factor de crecimiento promedio anual es:

$$\bar{x}_G = \sqrt[4]{1.28 \times 1.5744 \times 1.99949 \times 2.49936} = 1.78143$$

$$\bar{x} = 1 + \frac{78.143}{100}$$

El porcentaje promedio de crecimiento es de 78.143%

Estimando el salario para el quinto año:

249.9362 + (0.78143 x 249.9362) = 445.2438

La tasa para el quinto año será:

445.2438/100 = 4.452438

- La media de crecimiento anual es:

$$\bar{x} = \frac{1.28 + 1.5744 + 1.99949 + 2.49936 + 4.45244}{5} = 2.3611$$

$$\bar{x} = 1 + \frac{136.11}{100}$$

El porcentaje promedio de crecimiento es de 136.11%

- La media geométrica anual de crecimiento

$$\bar{x}_G = \sqrt[5]{1.28 \times 1.5744 \times 1.99949 \times 2.49936 \times 4.452438} = 2.1396$$

$$\bar{x} = 1 + \frac{113.96}{100}$$

El porcentaje promedio de crecimiento es de 113.96%

7.4 CAPÍTULO 4

1. En la tabla se pueden observar los promedios de las desviaciones estándar en el proceso de llenado de gaseosa en la empresa Coca-Cola durante el año 2023.

1.4	1.8	2	2.2	2.4
1.4	1.9	2	2.2	2.5
1.5	1.9	2.1	2.2	2.6
1.6	2	2.1	2.3	2.6
1.6	2	2.2	2.3	2.9
1.7	2	2.2	2.4	3.5

Obtenga:

a. La diferencia entre el valor máximo y mínimo (rango)

b. Desviación de medias

c. La Desviación cuartílica

d. Varianza

e. Desviación estándar

f. El Coeficiente de variación.

Solución

- **Rango**

$$R = X_{máx} - X_{\min} = 3.5 - 1.4 = 2.1$$

Tabla 62

Distribución de frecuencias según las desviaciones típicas del llenado de gaseosa en la empresa coca cola en el 2023

Llenado de botella	Frecuencias	% del Total	% Acumulado
1.3 - 1.7	5	16.7 %	16.7 %
1.7 - 2.1	9	30.0 %	46.7 %
2.1 - 2.5	11	36.7 %	83.3 %
2.5 2.9	3	10.0 %	93.3 %
2.9 3.3	1	3.3 %	96.7 %
3.3 3.7	1	3.3 %	100.0 %

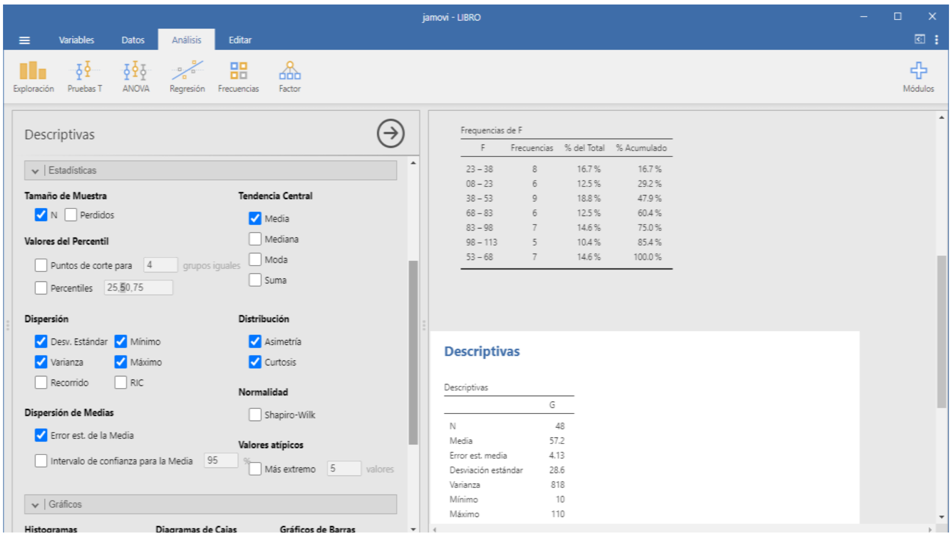
Figura 37

Medidas descriptivas de las desviaciones típicas del llenado de gaseosa en la empresa coca cola en el 2023

	G
N	48
Media	57.2
Error est. media	4.13
Desviación estándar	28.6
Varianza	818
Mínimo	10
Máximo	110
Asimetría	0.208
Error est. asimetría	0.343
Curtosis	-1.08
Error est. curtosis	0.674

Figura 38

Ventana de trabajo para resultados descriptivos y tablas de distribución de frecuencias



Trabajando manualmente tenemos la siguiente tabla

Tabla 63

Distribución de frecuencias según las desviaciones típicas del llenado de gaseosa en la empresa coca cola en el 2023

〈	]	Y_i	f_i	F_i	$Y_i f_i$	$\lvert Y_i - \bar{y}\rvert$	$\lvert Y_i - \bar{y}\rvert \times f_i$	$f_i \times (Y_i - \bar{y})^2$
1.3	1.7	1.5	5	5	7.5	0.653	3.267	2.134
1.7	2.1	1.9	9	14	17.1	0.253	2.280	0.578
2.1	2.5	2.3	11	25	25.3	0.147	1.613	0.237
2.5	2.9	2.7	3	28	8.1	0.547	1.640	0.897
2.9	3.3	3.1	1	29	3.5	1.347	0.947	0.896
3.3	3.7	3.5	1	30	3.1	0.947	1.347	1.814
TOTAL			30	-	64.6		11.093	6.555

- **Desviación de medias**

Primero se calcula el promedio:

$$\bar{y} = \frac{\sum y_i f_i}{n} = \frac{64.6}{30} = 2.153$$

$$DM = \frac{\sum_{i=1}^{m} f_i \lvert y_i - \bar{y}\rvert}{n} = \frac{11.093}{30} = 0.369$$

- **Desviación cuartílica**

Calculando el cuartil 1

$$\frac{nk}{4} = \frac{30(1)}{4} = 7.5$$

Asimismo, la menor frecuencia absoluta acumulada que supera es $F_2 = 14$

$$Q_1 = 1.7 + 4\left(\frac{6.5-5}{9}\right) = 2.81$$

Calculando el cuartil 3

$$\frac{nk}{4} = \frac{30(3)}{4} = 22.5$$

Por lo tanto, la menor frecuencia absoluta acumulada que supera es $F_3 = 25$

$$Q_1 = 2.1 + 4\left(\frac{22.5-14}{11}\right) = 5.19$$

$$Q_D = \frac{Q_3 - Q_1}{2} = \frac{5.19 - 2.81}{2} = 1.19$$

- **Varianza**

$$V(x) = S_x'^2 = \frac{\sum_{i=1}^{n} f_i \left(y_i - \bar{y} \right)^2}{n}$$

$$V(x) = \frac{6.555}{30} = 0.2185$$

- **Desviación típica**

$$D(x) = S_X = \sqrt{V(x)} = \sqrt{S_X^2} = \sqrt{0.2185} = 0.467$$

- **Coeficiente de variación**

$$CV_X = \frac{S_X}{\bar{y}} \times 100$$

$$CV_X = \frac{0.467}{2.153} \times 100 = 21.69\% \ \mathbf{> 15\%}$$

$$\therefore \text{La distribucion de los datos no es HOMOGENEO}$$

2. Se tienen registrados la cantidad de días que 15 estudiantes de primer año estuvieron ausentes en la escuela durante todo el año escolar.

1	8	2	2	3	9	2	3
5	2	4	7	9	8	5	

Calcule lo siguiente:

a. Calcular e interpretar Rango
b. Desviación de medias
c. La Desviación cuartílica
d. Varianza
e. Desviación estándar
f. El Coeficiente de variación.

Solución

Figura 39

Ventana de trabajo para resultados descriptivos

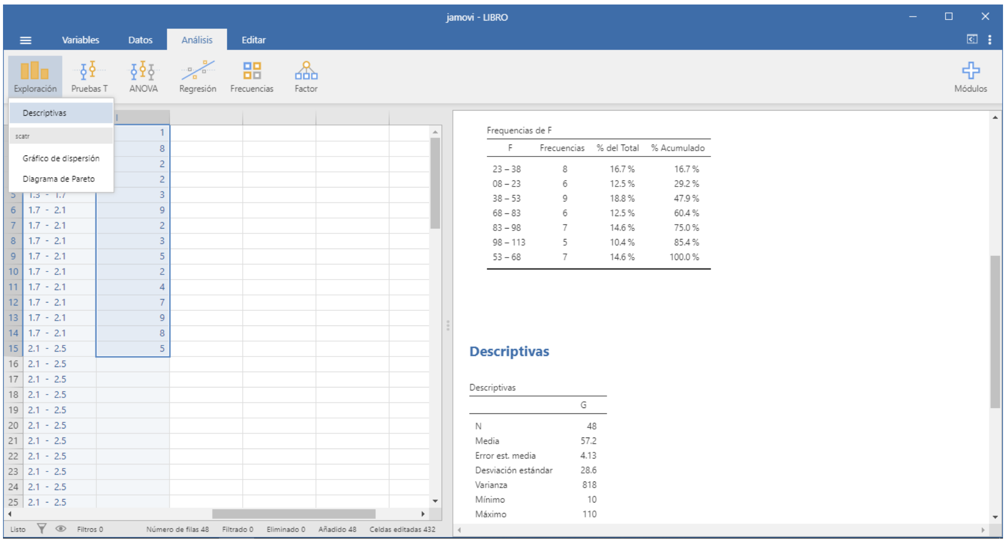


Tabla 64

Medidas descriptivas de los registros de días de ausencia de los estudiantes de primer año a lo largo del año escolar

	llenado de botella
N	15
Perdidos	33
Media	4.67
Mediana	4.00
Desviación estándar	2.85
Mínimo	1.00
Máximo	9.00

Trabajando manualmente:

- **Rango**

$$R = X_{\max} - X_{\min} = 1 - 9 = 8$$

- **Desviación de medias**

$$DM = \frac{\sum_{i=1}^{m} f_i \left| y_i - \bar{y} \right|}{n} = 2.444$$

- **Desviación cuartílica:** Ordenando los datos:

1	2	2	2	2	3	3	4
5	5	7	8	8	9	9	

Hallando los cuartiles:

$$\frac{(n+1)k}{4} = \frac{15(1)}{4} = 3.75$$

$$Q_1 = 2 + (2-2)\times 0.75 = 2$$

$$\frac{(n+1)k}{4} = \frac{15(3)}{4} = 11.25$$

$$Q_1 = 8 + (8-8)\times 0.25 = 8$$

$$Q_D = \frac{Q_3 - Q_1}{2} = \frac{8-2}{2} = 3$$

- **Varianza**

$$V(x) = S_x'^2 = \frac{\sum_{i=1}^{n} f_i\left(y_i - \bar{y}\right)^2}{n}$$

$$V(x) = 7.556$$

- **Desviación estándar**

$$D(x) = S_X = \sqrt{V(x)} = \sqrt{S_X^2} = \sqrt{7.556} = 2.749$$

- **Coeficiente de Variación**

$$CV_X = \frac{2.749}{4.667}\times 100 = 166.038\% > \mathbf{15\%}$$

$\therefore$ La distribucion de los datos no es HOMOGENEO

3. Una compañía de micro computadoras personales (PC) cuentan con 530 vendedores a nivel nacional; con el ranking de ventas durante los últimos cinco años, calcular la varianza utilizando el método abreviado. Los datos se muestran en cuadro de frecuencias.

Tabla 65

Distribución de vendedores a nivel nacional de según el ranking de ventas durante los últimos cinco años

〈]	Y_i	f_i	Z_i	Y_iZ_i	$z_i^2 f_i$
15 20	17.5	20	-3	-60	180
20 25	22.5	50	-2	-100	200
25 30	27.5	90	-1	-90	90
30 35	32.5	150	0	375	0
35 40	37.5	100	1	100	100
40 45	42.5	85	2	170	340
45 50	47.5	35	3	105	315
Total	**-**	**530**	**-**	**125**	**1225**

Solución

$$S_x^2 = V(x) = \frac{c^2}{n-1}\left[\sum z_i'^2 f_i - \frac{\left(\sum z_i' f_i\right)^2}{n}\right] = V(x) = \frac{5^2}{(530-1)}\left[1225 - \left(\frac{125}{530}\right)^2\right]$$

$$V(x) = 57.890$$

4. Supongamos que estamos evaluando un área forestal de "n" hectáreas de los árboles presentes en parcelas de la muestra estudiada. Calcule la desviación estándar con un tamaño de muestra de 50 a partir de la distribución de frecuencias

Clases	0 - 3.5	3.5 - 5.5	5.5 - 8.5	8.5 - 12.5	12.5 – 16.5	16.5 – 17.5
Hi	0.04	0.08	0.32	0.16	0.10	0.30

Solución

Tabla 66

Evaluación de un área forestal de "n" hectáreas de los árboles presentes en parcelas de la muestra estudiada

〈]	Y_i	h_i	$Y_i h_i$	$(Y_i - \bar{x}_i)^2 \times h_i$
0.0 - 3.5	1.75	0.04	0.07	33.489
3.5 - 5.5	4.5	0.08	0.36	1.62
5.5 - 8.5	7.0	0.32	2.24	15.68
8.5 – 12.5	10.5	0.16	1.68	17.64
12.5 – 16.5	14.5	0.10	1.45	21.025
16.5 – 17.5	17.0	0.30	5.1	86.7
Total	-	-	10.9	146.014

Calculando el promedio y la desviación estándar

$$\overline{x} = \sum x_i h_i = 10.9$$

$$S'_x = \sqrt{\sum_{i=1}^{n} h_i \left(x_i - \bar{x}\right)^2} = 12.084$$

5. El dueño de una estación de gasolina informa, que un cliente particular consume durante los meses de invierno un promedio de 105 galones por semana con una desviación estándar de 12 galones. Encuentre la media y desviación estándar del consumo del cliente por día.

Solución

Al utilizar las propiedades para la media aritmética como para la desviación estándar, se tiene:

$$M\left[k\bar{x}\right] = kM\left[\bar{x}\right]$$

$$M\left[\frac{1}{7}(105)\right] = \frac{1}{7}(105) = 15$$

Utilizar un promedio de 15 galones por día.

$$S_y = S_{ax} = |a|\, S_x$$

$$S_y = S_{\frac{1}{7}x} = \left|\frac{1}{7}\right| S_x = \frac{1}{7}(12) = 1.7143$$

$\Rightarrow$ La desviación estándar por día es de 1.7143

6. Se clasifica los sueldos en miles de nuevos soles de los empleados de la empresa "YAMÓN" durante el año 2022, calcular el coeficiente de variación de los empleados de la empresa y se han obtenido los resultados:

Tabla 67

Sueldos de los empleados de la empresa "YAMÓN"

[SUELDOS ›	f_i
150 - 160	10
160 - 170	30
170 - 180	20
Total	60

Solución

Para resolver este ejercicio, es necesario determinar el promedio y la desviación estándar

$$\bar{x} = \frac{\sum f_i x_i}{n} = \frac{10000}{60} = 166.67$$

$$S_x^2 = \frac{\sum_{i=1}^{n} f_i \left(x_i - \bar{x}\right)^2}{n} = \frac{2833.333}{60} = 47.222$$

$$S_x^2 = \sqrt{47.222} = 6.8718$$

Calculando el coeficiente de variación de La empresa "YAMÓN"

$$CV_X = \frac{6.8718}{166.67} \times 100 = 4.1230\%$$

7. Una población de alumnos tiene una estatura media de 160 cm con desviación estándar de 16 cm. Estos mismos estudiantes, tienen un peso medio de 70 kg con una desviación estándar de 14 kg. ¿Cuál de las dos variables, estatura o peso, muestra mayor dispersión o variabilidad?

Estatura

$\mu_E = 160m \wedge \sigma_E = 16m$

$$CV_E = \frac{\sigma_E}{\mu_E} = \frac{16\,cm}{160\,cm} = \frac{1}{10} = 0.1 = 10\%$$

Peso

$\mu_p = 160m \wedge \sigma_E = 16m$

$$CV_p = \frac{S_p}{\overline{X}_p} = \frac{14\,kg}{70\,kg} = \frac{1}{5} = 0.2 = 20\%$$

Interpretación : Se puede observar que el Coeficiente de variación del peso es mayor que el Coeficiente de Variación de la Estatura ($CV_P > CV_E$, lo que indica, que el peso de esta población de alumnos tiene una mayor variabilidad relativa en comparación con su estatura.

7.5 CAPÍTULO 5

1. En la tabla se muestra los resultados evaluados por un ingeniero forestal acerca de la temperatura que afecta la tasa de descomposición de la materia orgánica en los suelos forestales.

Intervalo	[- 40 -30 ›	[-30 -20 ›	[-20 -10 ›	[-10 0 ›	[0 10 ›
Frecuencia	7	11	15	15	24

Intervalo	[10 20 ›	[20 30 ›	[30 4 0 ›	[40 50 ›	[5 0 60 ›
Frecuencia	41	26	17	7	3

Hallar los coeficientes de asimetría y apuntamiento.

Solución

Tabla 68

Temperatura que afecta la tasa de descomposición de la materia orgánica en los suelos forestales

[›	Y_i	f_i	$Y_i f_i$	$(Y_i - \overline{x})^2 f_i$	F_i
[- 40 - 30 ›	-35	7	-245	13306.72	7
[-30 - 20 ›	-25	11	-275	12418.56	18
[-20 - 10 ›	-15	15	-225	8354.4	33
[-10 - 0 ›	-5	24	-120	4439.04	57
[0 - 10 ›	5	49	245	635.04	106
[10 - 20 ›	15	41	615	1679.36	147
[20 - 30 ›	25	26	650	6992.96	173
[30 - 4 0 ›	35	17	595	11848.32	190
[40- 50 ›	45	7	315	9274.72	197
[5 0- 60 ›	55	3	165	6458.88	200
TOTAL		200	1720	75408	-

Calculando los estadígrafos:

- Promedio

$$\overline{x} = \frac{\sum f_i x_i}{n} = \frac{1720}{200} = 8.60$$

- Calculando la varianza:

$$S_x{}'^2 = \frac{\sum_{i=1}^{n} f_i\left(x_i - \bar{x}\right)^2}{n} = \frac{75408}{200} = 377{,}040$$

- Calculando la desviación estándar

$$S_x{}' = \sqrt{377.040} = 19.418$$

- Moda

$$M_o = 0 + 10\left(\frac{25}{25+8}\right) = 7.576$$

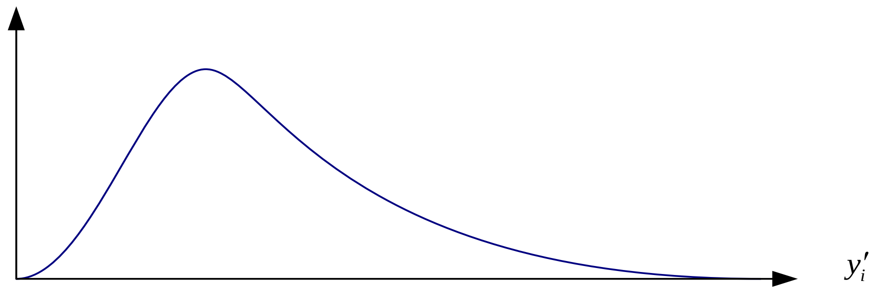
- Coeficiente de asimetría: Pearson

$$A_s = \frac{\bar{x} - M_o}{S_x} = \frac{8.60 - 7.576}{19.418} = 0.053$$

$\therefore$ Es una distribución Asimétrica Positiva

Figura 40

Gráfica del coeficiente de asimetría de la temperatura que afecta la tasa de descomposición de la materia orgánica en los suelos forestales



- Calculando los percentiles y cuartiles para la curtosis

$$Q_1 = -10 + 10\left(\frac{50-33}{24}\right) = -2.917 \qquad Q_3 = 20 + 10\left(\frac{150-147}{26}\right) = 21.154$$

$$P_{10} = -20 + 10\left(\frac{20-18}{15}\right) = -18.667 \qquad P_{90} = 30 + 10\left(\frac{180-173}{17}\right) = 34.118$$

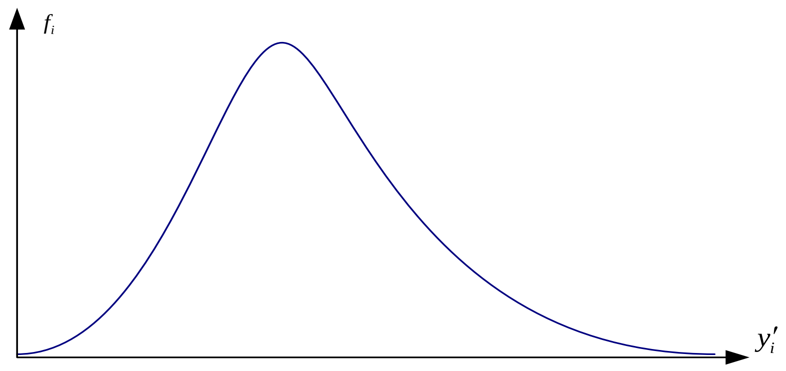
- Kurtosis

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{21.154 + 2.917}{2 \times (34.118 + 18.667)} = 0.59$$

$\therefore$ Es una distribución leptocúrtica porque es mayor que 0.263

Figura 41

Distribución leptocúrtica de la temperatura que afecta la tasa de descomposición de la materia orgánica en los suelos forestales



2. Una empresa decide evaluar a sus trabajadores de acuerdo con el número de ventas que realiza en el día. Los datos se encuentran resumidos en una tabla de frecuencias.

Tabla 69

Número de ventas que realiza en el día los trabajadores

Y_i	fi	$Y_i f_i$	$(Y_i - \bar{x})^2 f_i$	F_i
1	1	1	18.4041	1
2	1	2	10.8241	2
3	21	63	110.1261	23
4	52	208	86.5332	75
5	51	255	4.2891	126
6	31	186	15.6271	157
7	18	126	52.6338	175
8	14	112	102.8174	189
9	5	45	68.8205	194
10	6	60	133.1046	200
Total	200	1058	603.18	

Calculando los estadígrafos:

Tabla 70

Medidas descriptivas de ventas que realiza en el día los trabajadores

	A
N	200
Media	5.29
Mediana	5.00
Moda	4.00
Desviación estándar	1.75
Varianza	3.07
Mínimo	1
Máximo	10
Asimetría	0.706
Error est. asimetría	0.172
Curtosis	0.380
Error est. curtosis	0.342
25percentil	4.00
50percentil	5.00
75percentil	6.00

- Promedio

$$\overline{x} = \frac{\sum f_i x_i}{n} = \frac{1058}{200} = 5.26$$

- Desviación estándar

$$S_x{'}^2 = \frac{\sum_{i=1}^{n} f_i\left(x_i - \overline{x}\right)^2}{n} = \frac{603,18}{200} = 3,0159$$

$$S_x{'} = \sqrt{3,0159} = 1,737$$

- Moda

$$M_o = 4$$

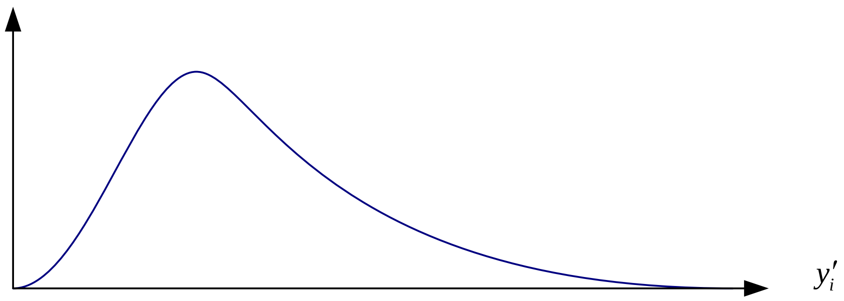
- Coeficiente de asimetría: Pearson

$$A_s = \frac{\overline{x} - M_o}{S_x} = \frac{5.29 - 4}{1.737} = 0.7426$$

$\therefore$ Es una distribución Asimétrica Positiva

Figura 42

Coeficiente de asimetría de Número de ventas que realiza en el día por sus rabajadores



.

- Calculando los percentiles y cuartiles para la curtosis

$$\frac{nk}{4} = \frac{200 \times 1}{4} = 50$$
$$Q_1 = 4$$

$$\frac{nk}{4} = \frac{200 \times 3}{4} = 150$$
$$Q_3 = 6$$

$$\frac{nk}{100} = \frac{200 \times 10}{100} = 20$$
$$P_{10} = 3$$

$$\frac{nk}{100} = \frac{200 \times 90}{100} = 100$$
$$P_{90} = 5$$

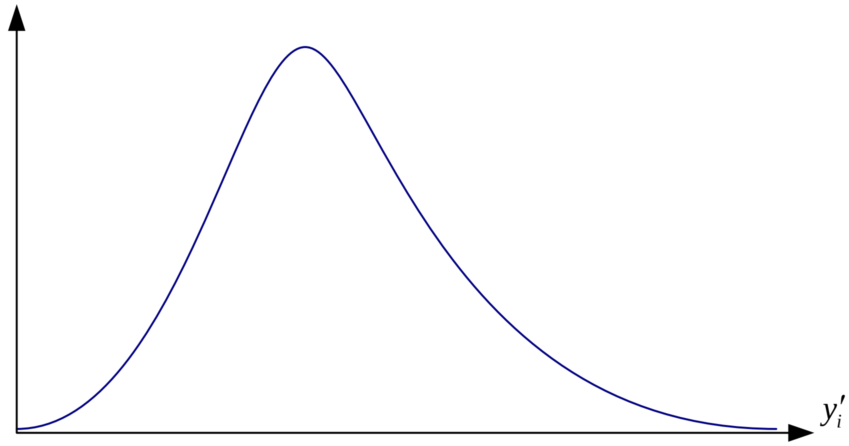
- Kurtosis

$$K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})} = \frac{6-4}{2 \times (5-3)} = 0.50$$

$\therefore$ Es una distribución leptocúrtica porque es mayor que 0.263

Figura 43

Distribución leptocúrtica de Número de ventas que realiza en el día por sus rabajadores



.

3. La resistencia representa una propiedad crucial en los materiales empleados en las casas prefabricadas. En el proceso de evaluación, se sometieron a pruebas de esfuerzo rigurosas los 50 elementos de placa prefabricada, y se documentó la amplitud máxima (en milímetros) de las grietas resultantes. Los datos recopilados para la muestra en análisis son los siguientes:

Tabla 71

Pruebas de esfuerzo rigurosas de las placas prefabricadas según la amplitud máxima de las grietas resultantes.

Ancho máximo de la grieta (mm) $[L_i - L_S]$	**N° de elementos de placa** f_i	F_i
0.4 – 0.5	4	4
0.5 – 0.6	15	20
0.6 – 0.7	11	31
0.7 – 0.8	8	39
0.8 – 0.9	5	44
[0.9 – 1]	6	50
Total	50	-

Calcular y analizar el coeficiente de curtosis relacionado con los anchos máximos (en milímetros) de las grietas desarrolladas en la muestra de 50 elementos de placas prefabricadas.

Solución:

Tabla 72

Tabla de percentiles

Percentil	Posición	Fórmula	Valor
P_{75}	37.5	$P_{75} = 0.7 + 0.1 * \frac{[37.5 - 31]}{8} = 0.78$	0.78
P_{25}	12.5	$P_{25} = 0.5 + 0.1 * \frac{[12.5 - 6]}{16} = 0.54$	0.54
P_{90}	45	$P_{90} = 0.9 + 0.1 * \frac{[45 - 44]}{6} = 0.92$	0.92
P_{10}	5	$P_{10} = 0.5 + 0.1 * \frac{[5 - 4]}{16} = 0.51$	0.51

Reemplazando en la fórmula de curtosis:

$$K_U = \frac{(P_{75} - P_{25})}{2(P_{90} - P_{10})} = \frac{0.78 - 0.54}{2(0.92 - 0.51)} = 0.293$$

Interpretación:

Los datos presentan una distribución PLATICÚRTICA

1. Las puntuaciones de la productividad (X) y calidad del trabajo (Y) de 40 trabajadores son las siguientes:

X:	20	30	20	30	40	30	30	40	40	40	50	50	50
Y:	25	31	38	38	59	45	51	35	42	49	67	46	55
fi	6	4	3	3	3	2	1	1	3	2	2	3	7

Calcular:

a. Las frecuencias absolutas simples y acumuladas

b. Las frecuencias relativas simples y acumuladas

c. Los porcentajes simples y acumuladas

d. La marginal de *x* e *y*

e. Promedio de *x*, Promedio de *y*, varianza muestral de *x*, varianza muestral de *y*, desviación estándar de *x* y desviación estándar de *y*.

f. La $Cov\left(x,y\right)$

g. $V\left(x+y\right)$, $V\left(x-y\right)$

Solución

Tabla 73

Ventana de trabajo para resultados de la tabla de distribución de frecuencias absolutas simples

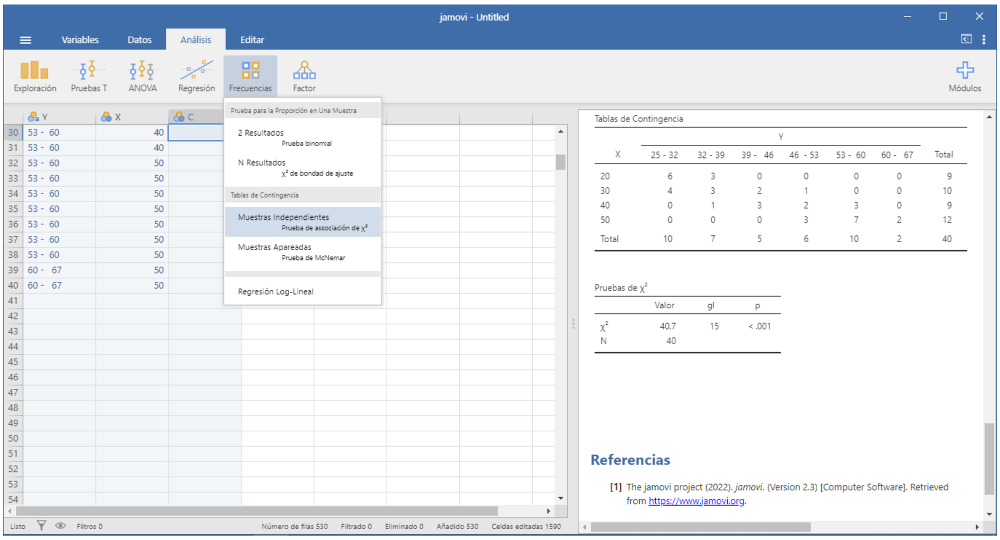


Tabla 74

Distribución bidimensional de frecuencias absolutas simples

	Y						
X	25 - 32	32 - 39	39 - 46	46 - 53	53 - 60	60 - 67	Total
20	6	3	0	0	0	0	9
30	4	3	2	1	0	0	10
40	0	1	3	2	3	0	9
50	0	0	0	3	7	2	12
Total	10	7	5	6	10	2	40

Tabla 75

Distribución bidimensional de frecuencias absolutas s y porcentajes simples

		Y						
X		25 - 32	32 - 39	39 - 46	46 - 53	53 - 60	60 - 67	Total
20	Observado	6	3	0	0	0	0	9
	% del total	15.0 %	7.5 %	0.0 %	0.0 %	0.0 %	0.0 %	22.5 %
30	Observado	4	3	2	1	0	0	10
	% del total	10.0 %	7.5 %	5.0 %	2.5 %	0.0 %	0.0 %	25.0 %
40	Observado	0	1	3	2	3	0	9
	% del total	0.0 %	2.5 %	7.5 %	5.0 %	7.5 %	0.0 %	22.5 %
50	Observado	0	0	0	3	7	2	12
	% del total	0.0 %	0.0 %	0.0 %	7.5 %	17.5 %	5.0 %	30.0 %
Total	Observado	10	7	5	6	10	2	40
	% del total	25.0 %	17.5 %	12.5 %	15.0 %	25.0 %	5.0 %	100.0 %

- ***Para X:***

  V.E: Puntuaciones de la productividad

  T.V: Cuantitativa continua / Razón

  $R_x = \{20; 30; 40; 50\}$

  En algunos casos una variable cuantitativa continua se debe tratar como una variable cuantitativa discreta cuando presenta una constante (k). En este caso (k=10)

- ***Para Y:***

  V.E: Puntuaciones calidad de trabajo

  T.V: Cuantitativo Continuo/Razón

  Calcular el rango

  R = 67 – 25 = 42

  Número de intervalos

  m = 1 + 3.33 log (40) = 6.33

  m = 6 intervalos

Tabla 76

Distribución bidimensional de frecuencias absolutas simples

$[L_I - L_S)$ / x_i	[25 - 32)	[32 - 39)	[39 - 46)	[46 - 53)	[53 - 60)	[60 - 67]	**TOTAL**
20	6	3					**09**
30	4	3	2	1			**10**
40		1	3	2	3		**09**
50				3	7	2	**12**
TOTAL	**10**	**07**	**05**	**06**	**10**	**02**	**40**

Tabla 77

Distribución bidimensional de frecuencias absolutas acumuladas

$[L_i - L_s)$ / x_i	[25 32)	[32 39)	[39 46)	[46 53)	[53 60]	[60 67]
20	6	9	9	9	9	9
30	10	16	18	19	19	19
40	10	17	22	25	28	28
50	10	17	22	28	38	40

Tabla 78

Distribución bidimensional de frecuencias absolutas acumuladas

$[L_i - L_s)$ / x_i	[25 32)	[32 39)	[39 46)	[46 53)	[53 60]	[60 67]
20	6	9	9	9	9	9
30	10	16	18	19	19	19
40	10	17	22	25	28	28
50	10	17	22	28	28	40

Tabla 79

Distribución bidimensional de frecuencias relativas simples

$[L_i - L_s)$ / χ_i	[25 32)	[32 39)	[39 46)	[46 53)	[53 60]	[60 67]	**TOTAL**
20	0.15	0.075					0.225
30	0.1	0.075	0.05	0.025			0.25
40		0.025	0.075	0.05	0.075		0.225
50				0.075	0.175	0.05	0.3
TOTAL	0.25	0.175	0.125	0.15	0.25	0.05	1

Tabla 80

Distribución bidimensional de frecuencias relativas acumuladas

$[L_i - L_s)$ / χ_i	[25 32)	[32 39)	[39 46)	[46 53)	[53 60]	[60 67]
20	0.15	0.225	0.225	0.225	0.225	0.225
30	0.25	0.40	0.45	0.475	0.475	0.475
40	0.25	0.425	0.55	0.625	0.7	0.7
50	0.25	0.425	0.55	0.7	0.7	1

Tabla 81

Distribución bidimensional de porcentajes simples

x_i \ $[L_i - L_s)$	[25 32)	[32 39)	[39 46)	[46 53)	[53 60]	[60 67]	**TOTAL**
20	15%	7.5%					**22.5%**
30	10%	7.5%	5%	2.5%			**25%**
40		2.5%	7.5%	5%	7.5%		**22.5%**
50				7.5%	17.5%	5%	**30%**
TOTAL	**25%**	**17.5%**	**12.5%**	**15%**	**25%**	**5%**	**100%**

Tabla 82

Distribución bidimensional de porcentajes acumuladas

x_i \ $[L_i - L_s)$	[25 32)	[32 39)	[39 46)	[46 53)	[53 60]	[60 67]
20	15%	22.5%	22.5%	22.5%	22.5%	22.5%
30	25%	40%	45%	47.5%	47.5%	47.5%
40	25%	42.5%	55%	62.5%	70%	70%
50	25%	42.5%	55%	70%	70%	100%

Tabla 83

Distribución marginal de X

x_i	$n_{i.}$	$h_{i.}$	$\%h_{i.}$
20	09	0.225	22.5%
30	10	0.25	25%
40	09	0.225	22.5%
50	12	0.3	30%
TOTAL	**40**	**1,00**	**100%**

Tabla 84

Distribución marginal de Y

$[L_i - L_s)$	$n_{.j}$	$h_{.i}$	$\%h_{.i}$
[25 32)	10	0.25	25%
[32 39)	7	0.175	17.5%
[39 46)	5	0.125	12.5%
[46 53)	6	0.15	15%
[53 60]	10	0.25	25%
[60 67]	2	0.05	5%
TOTAL	**40**	**1,00**	**100%**

Tabla 85

Distribución marginal de X

x_i	$n_{i.}$	$x_i n_{i.}$	$n_{i.}(x_{i.} - \bar{x})^2$
20	09	180	2304
30	10	300	360
40	09	360	144
50	12	600	2352
TOTAL	**40**	**1440**	**5160**

Calculando el promedio:

$$\bar{x} = \frac{\sum_{i=1}^{k} n_i \cdot x_i}{n} = \frac{1440}{40} = 36$$

Calculando la varianza:

$$s_x^2 = \frac{\sum_{i=1}^{k} n_{i.}(x_i - \bar{x})^2}{n-1} = \frac{5160}{39} = 132.308$$

Calculando la desviación estándar:

$$s_x = \sqrt{132.308} = 11.503$$

Tabla 86

Distribución marginal de Y

$[L_i - L_s)$	$Y_{.j}$	$n_{.j}$	$Y_{.j}.n_{.j}$	$\sum_{j=1}^{k} n_{.j}\left(y_i - \bar{y}\right)^2$
[25 32)	28.5	10	285	2212.65625
[32 39)	35.5	7	248.5	434.109375
[39 46)	42.5	5	212.5	3.828125
[46 53)	49.5	6	297	225.09375
[53 60]	56.5	10	565	1722.65625
[60 67]	63.5	2	127	810.03125
TOTAL	**-**	**40**	**1735**	**5408.375**

Calculando el promedio:

$$\bar{y} = \frac{\sum_{j=1}^{l} n_{.j}.y_j}{n} = \frac{1735}{40} = 43.375$$

Calculando la varianza:

$$s_y^2 = \frac{\sum_{j=1}^{k} n_{.j}\left(y_j - \bar{y}\right)^2}{n-1} = \frac{5408.375}{39} = 138.676$$

Calculando la desviación estándar:

$$s_x = \sqrt{138.676} = 11.776$$

Calculando la covarianza

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{l} n_{ij}x_i y_i - \bar{x}\bar{y}$$

$$Cov(x,y) = \frac{1}{40}[(6 \times 20 \times 28.5) + (4 \times 30 \times 28.5) + (3 \times 20 \times 35.5) + \cdots + (2 \times 50 \times 63.5)] - 36 \times 43.375$$

$$Cov(x = 113.75$$

$V(x+y)$, $V(x-y)$

Calculando la $V(x+y)$

$$V(x+y) = V(x) + V(y) + 2Cov(x,y) = 132.308 + 138.676 + 2(113.75) = 498.484$$

Calculando la $V(x-y)$

$$V(x-y) = V(x) + V(y) - 2Cov(x,y) = 132.308 + 138.676 - 2(113.75) = 43.484$$

1. Supongamos que un ingeniero forestal desea analizar la relación entre la precipitación anual (en mm) y el crecimiento en altura (en metros) del árbol de Caoba. Se recogen los siguientes datos de 5 parcelas experimentales.

Parcela	1	2	3	4	5
Precipitación en (mm) (x)	800	850	900	950	1000
Crecimiento en altura en (mt) (Y)	1.5	1.8	2	2.2	2.4

Con base en esta información:

a. Elabore un diagrama de dispersión. Comentar.

b. Calcule el coeficiente de correlación lineal

c. Hallar e interpretar el coeficiente de determinación.

d. Hallar la ecuación de regresión lineal.

e. En el diagrama de dispersión trazar la recta de la ecuación de regresión lineal.

f. Si las precipitaciones es 870 ¿Cuál es el crecimiento?

g. Si el crecimiento es de 750 ¿Cuál es la precipitación?

Solución

Tabla 87

Primera ventana de trabajo de la relación entre la precipitación anual y el crecimiento en altura del árbol de Caoba

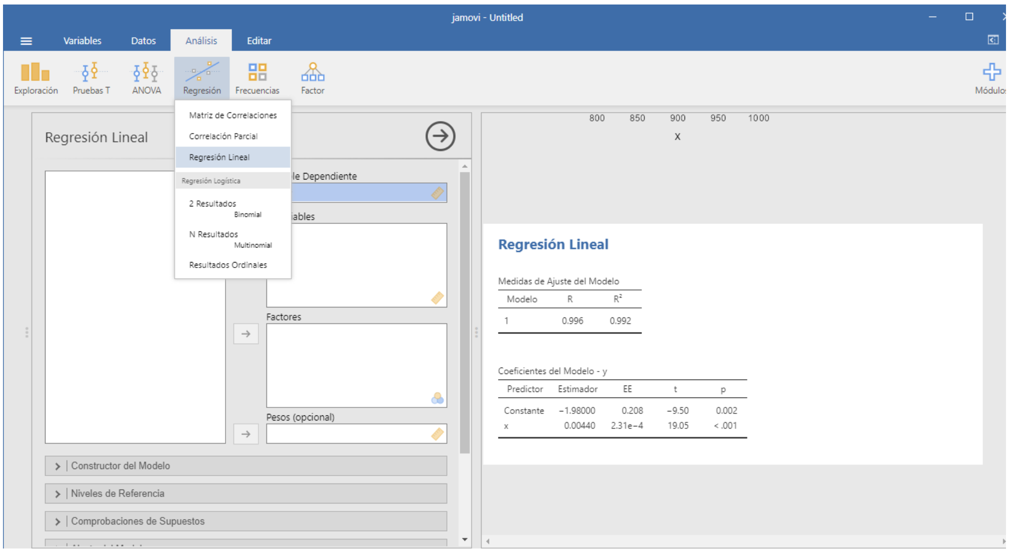


Tabla 88

Segunda fase de análisis de la relación entre la precipitación anual y el crecimiento en altura del árbol de caoba.

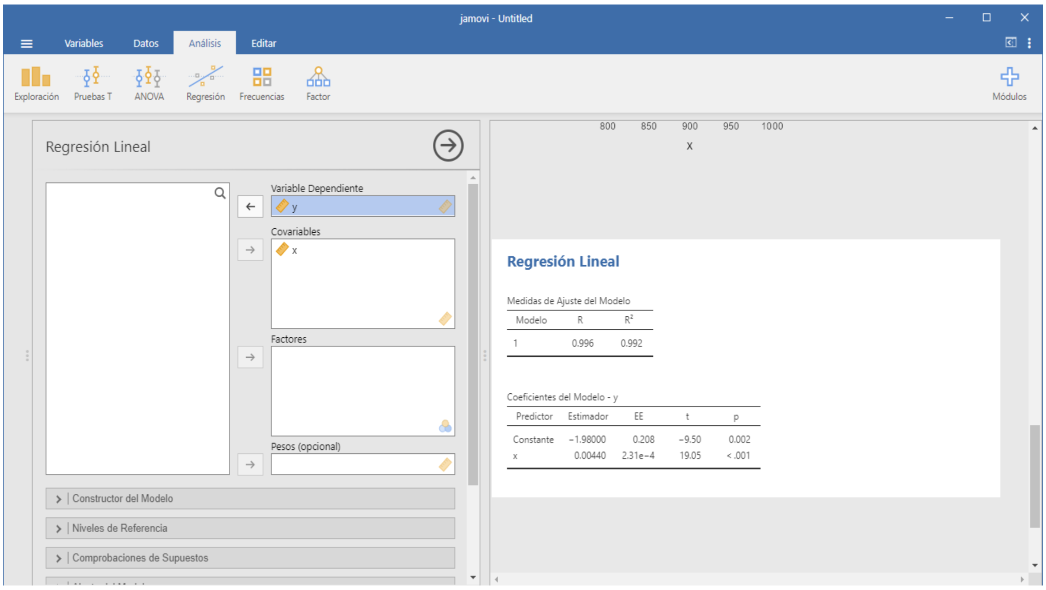


Tabla 89

Medidas de Ajuste del Modelo

Modelo	R	R^2
1	0.996	0.992

Tabla 90

Coeficientes del Modelo - y

Predictor	Estimador	EE	t	p
Constante	-1.98000	0.208	-9.50	0.002
x	0.00440	2.31e-4	19.05	< .001

De la tabla 90 se obtiene el siguiente modelo:

$$Y = -1.98 + 0.0044x$$

Figura 44

Tercera etapa del análisis de la relación entre la precipitación anual y el crecimiento en altura del árbol de caoba.

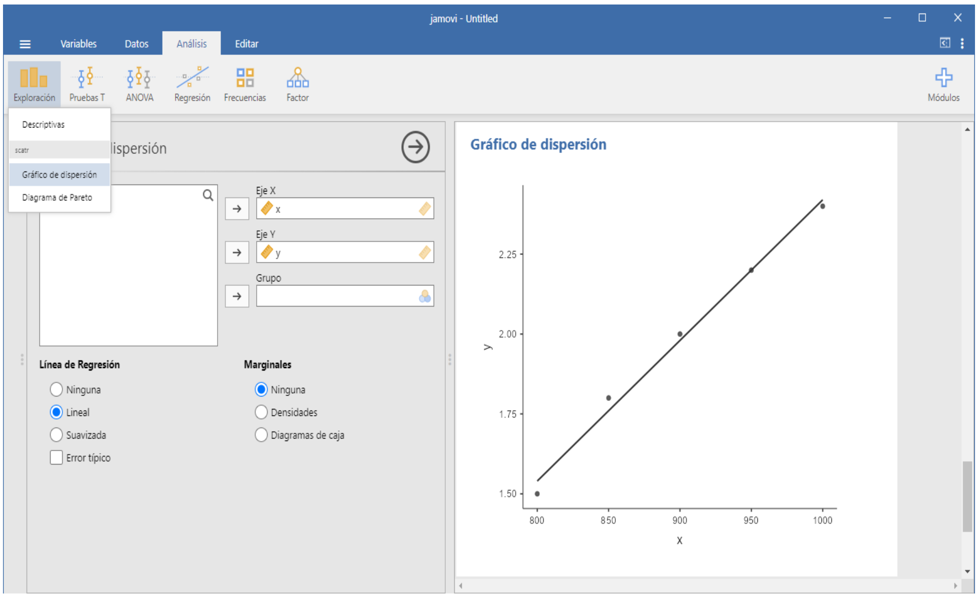
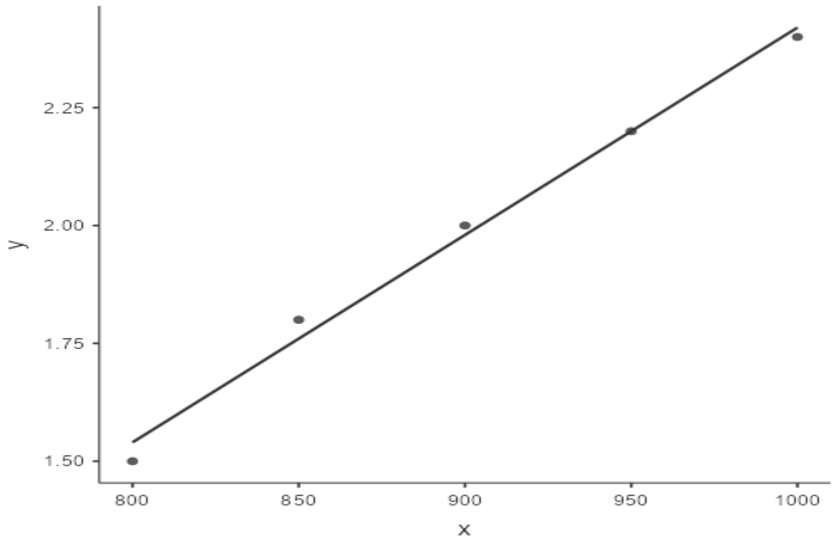


Figura 45

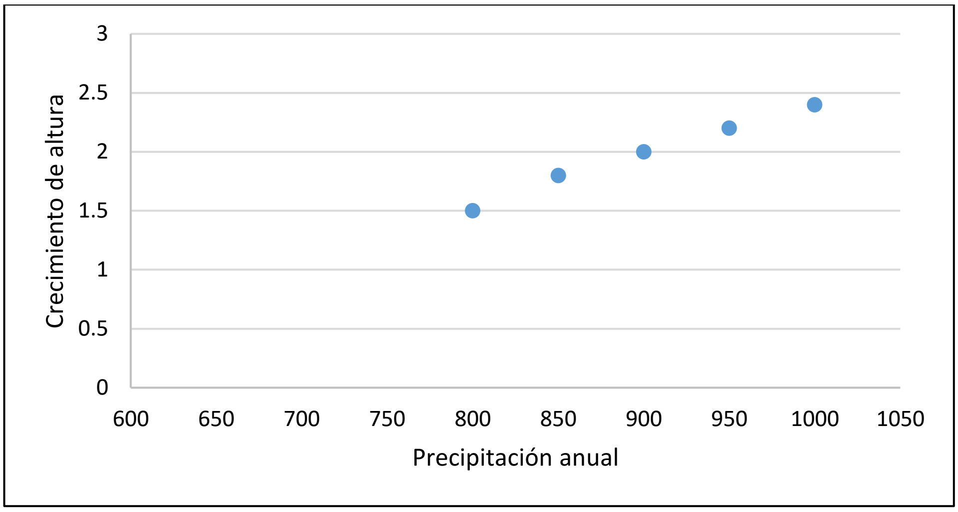
Diagrama de dispersión de las variables: precipitación anual y crecimiento en altura de los árboles de Caoba



Trabajando manualmente tememos:

Figura 46

Diagrama de dispersión de las variables: precipitación anual y crecimiento en altura de los árboles de Caoba



Interpretación.

Si analizamos la distribución de los puntos en este gráfico de dispersión, se puede ver claramente una relación lineal entre precipitación anual y crecimiento en altura de los árboles de Caoba.

Calcular las sumas de las variables

Tabla 91

Tabla de sumas de las variables: precipitaciones y crecimiento en altura de los árboles de Caoba

Parcela	Precipitación en (mm) (X)	Crecimiento (mt) (Y)	X^2	Y^2	XY
1	800	1.5	640000	2.25	1200
2	850	1.8	722500	3.24	1530
3	900	2	810000	4	1800
4	950	2.2	902500	4.84	2090
5	1000	2.4	1000000	5.76	2400
Total	4500	9.9	4075000	20.09	9020

Aplicar la fórmula del **Coeficiente de correlación de Pearson:**

$$r = \frac{n\sum X_iY_i - \sum X_i \sum Y_i}{\sqrt{n\sum X^2 - (\sum X_i)^2} \times \sqrt{n\sum Y^2 - (\sum Y_i)^2}}$$

$$r = \frac{5 \times 9020 - 4500 \times 9.9}{\sqrt{5 \times [4075000 - (4500)^2]} \times \sqrt{5 \times [20{,}09 - (9.9)^2]}} = 0.999$$

Interpretación:

Esto nos indica que existe una correlación positiva muy alta entre la precipitación anual y crecimiento en altura de los árboles de Caoba.

Coeficiente de determinación

$R = r^2$

$R = (0.999)^2 = 0.992$

Interpretación:

El 99.2% del crecimiento en altura de los árboles de Caoba es explicado por la precipitación anual.

Ecuación de regresión lineal

Realizando los cálculos:

$$\beta_1 = \frac{n\sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{5(9020) - 4500(9.9)}{5(4075000) - 4500^2} = 0.0044$$

$$\beta_0 = \frac{\sum y - \beta_1 \sum x}{n} = \frac{9.9 - 0.0044(4500)}{5} = -1.98$$

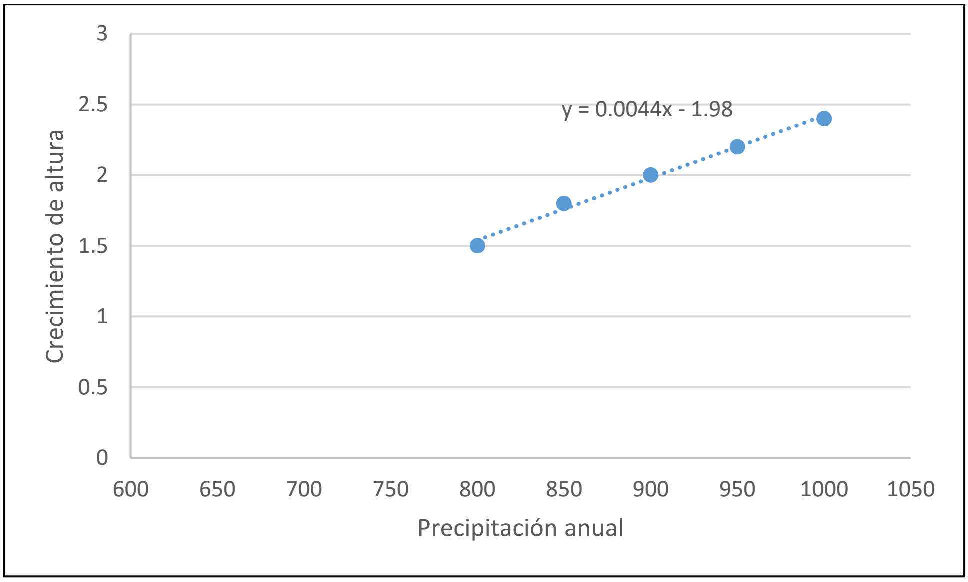
Por lo tanto***, la ecuación de regresión es:***

$Y = -1.98 + 0.0044x$

Trazando la recta de la ecuación de regresión lineal en el diagrama de dispersión

Figura 47

Diagrama de dispersión de las variables: precipitación anual y crecimiento en altura de los árboles de Caoba



Si las precipitaciones es 870 ¿Cuál es el crecimiento?

$$Y = -1.98 + 0.0044(870) = 1.85$$

Si el crecimiento es de 750 ¿Cuál es la precipitación?

$$750 = -1.98 + 0.0044x$$

$$x = \frac{750 + 1.98}{0.0044} = 170.90$$

2. Determinar la línea de regresión utilizando la serie de datos que representa altura y peso de los estudiantes en una clase. En este caso, se considera que la altura es la variable independiente, representada por "x", mientras que el peso es la variable dependiente, representada por "y". Los datos se dan a continuación:

Alumno	1	2	3	4	5	6	7	8	9	10
Estatura	1.25	1.28	1.27	1.21	1.22	1.29	1.3	1.24	1.27	1.29
Peso	32	33	34	30	32	35	34	32	32	35

Alumno	11	12	13	14	15	16	17	18	19	20
Estatura	1.25	1.28	1.27	1.21	1.22	1.29	1.3	1.24	1.27	1.29
Peso	33	35	34	30	33	34	35	32	33	33

Alumno	21	22	23	24	25	26	27	28	29	30
Estatura	1.25	1.28	1.27	1.21	1.22	1.29	1.3	1.24	1.27	1.29
Peso	33	34	34	31	32	34	34	31	35	34

Solución

Figura 48

Análisis de la relación entre el Altura y peso de los estudiantes en una clase

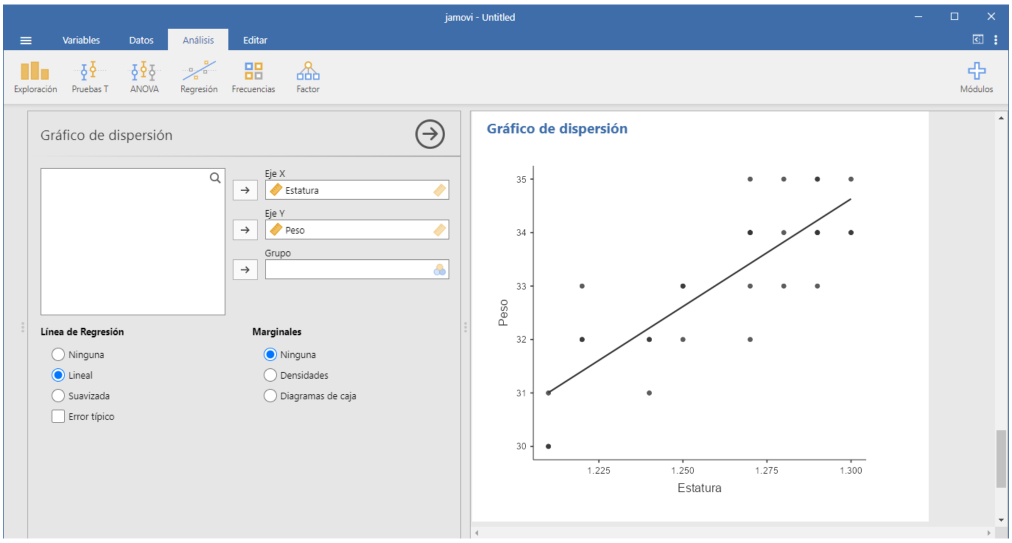


Figura ***49***

Diagrama de dispersión altura y peso de los estudiantes

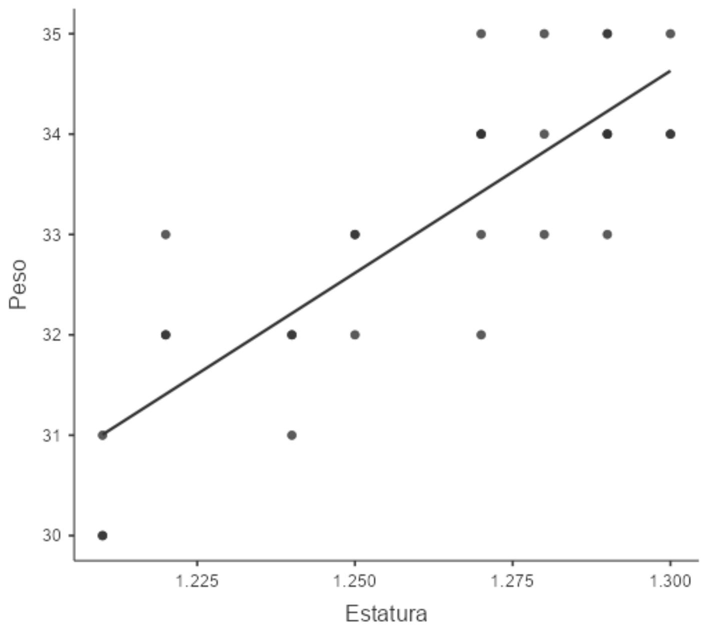


Tabla 92

Medidas de Ajuste del Modelo

Modelo	R	R^2
1	0.828	0.686

Tabla *93*

Coeficientes del Modelo - y

Predictor	Estimador	EE	t	p
Constante	-17.7	6.50	-2.73	0.011
Estatura	40.3	5.15	7.82	< .001

De la tabla 93 se obtiene el siguiente modelo:

$$y = -17.714 + 40.265x$$

Trabajando manualmente:

Tabla 94

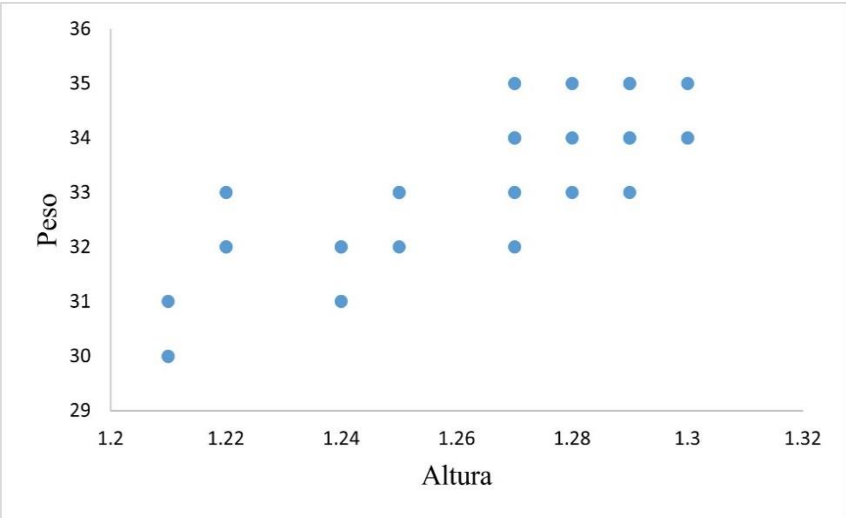
Altura y peso de los estudiantes en una clase

Alumno	Estatura	Peso	x^2	y^2	xy
1	1.25	32	1.56	1024	40
2	1.28	33	1.64	1089	42.24
3	1.27	34	1.61	1156	43.18
4	1.21	30	1.46	900	36.3
5	1.22	32	1.49	1024	39.04
6	1.29	35	1.66	1225	45.15
7	1.3	34	1.69	1156	44.2
8	1.24	32	1.54	1024	39.68
9	1.27	32	1.61	1024	40.64
10	1.29	35	1.66	1225	45.15
11	1.25	33	1.56	1089	41.25
12	1.28	35	1.64	1225	44.8
13	1.27	34	1.61	1156	43.18
14	1.21	30	1.46	900	36.3
15	1.22	33	1.49	1089	40.26
16	1.29	34	1.66	1156	43.86
17	1.3	35	1.69	1225	45.5

18	1.24	32	1.54	1024	39.68
19	1.27	33	1.61	1089	41.91
20	1.29	33	1.66	1089	42.57
21	1.25	33	1.56	1089	41.25
22	1.28	34	1.64	1156	43.52
23	1.27	34	1.61	1156	43.18
24	1.21	31	1.46	961	37.51
25	1.22	32	1.49	1024	39.04
26	1.29	34	1.66	1156	43.86
27	1.3	34	1.69	1156	44.2
28	1.24	31	1.54	961	38.44
29	1.27	35	1.61	1225	44.45
30	1.29	34	1.66	1156	43.86
Total	37.86	993.00	47.805	32929	1254.2

Figura 50

Diagrama de dispersión altura y peso de los estudiantes



Modelo de Regresión Estimada

$$Y = \beta_0 + \beta_1 x$$

$$\beta_1 = \frac{n\sum xy - \sum x \sum y}{n\sum x^2 - (\sum x)^2} = \frac{30(1254.2) - 37.86(993)}{30(47.805) - 37.86^2} = 40.265$$

$$\beta_0 = \frac{\sum y - \hat{\beta}_1 \sum x}{n} = \frac{993 - 40.265(37.86)}{30} = 17.714$$

Por lo tanto, la **recta** que mejor se ajusta a esta serie de datos es:

$y = -17.714 + 40.265x$

Esta recta define los valores estimados del peso, para cada valor de la estatura:

Estatura	**Valore estimados del Peso**
1.20	30.6
1.21	31.0
1.22	31.4
1.23	31.8
1.24	32.2
1.25	32.6
1.26	33.0
1.27	33.4
1.28	33.8
1.29	34.2
1.30	34.6

8 EJERCICIOS PROPUESTOS

8.1 CAPÍTULO 1

1. El término estadística se deriva de la palabra:
   a. Statics
   b. Staat
   c. Status
   d. Stadium
   e. De ninguna de las anteriores
   Rpta: b y c
2. La siguiente información: "Formulación de leyes que simplifiquen la descripción de un gran número de observaciones y expliquen dichas observaciones" constituye una de los aspectos fundamentales de:
   a. La técnica
   b. Toda la tecnología
   c. Todo fenómeno
   d. Toda ciencia
   e. Ninguna de las anteriores
   Rpta: d
3. La definición estadística se considera como:
   a. Una ciencia
   b. Un enfoque metodología general
   c. Un conjunto de prácticas y procedimientos concretos
   d. Un método de análisis
   e. Ninguna de las opciones.
   Rpta: a
4. Durante la época pre-colombina de América, ya existían antecedentes de la estadística. Es una proposición:
   a. Verdadera
   b. Falsa
   Rpta: a

5. La etapa inicial del desarrollo de la estadística se caracteriza por:
   a. Ser poco conocida
   b. Estar vinculada a los censos y registros de datos
   c. Estar circunscritas a Europa
   d. Haberla iniciado los romanos
   e. Estar sustentadas en el cálculo de probabilidades

   Rpta: b
6. Caracteriza a la etapa de sistematización de la estadística:
   a. El registro de los datos
   b. Su relación con las Matemáticas
   c. Las técnicas del muestreo
   d. La realización de los censos
   e. La aparición de las escuelas

   Rpta: e
7. La escuela que busca establecer los principios que gobiernan los eventos políticos y sociales es conocida como la escuela
   a. Administrativa
   b. Alemana
   c. Inglesa
   d. Francesa
   e. Ninguna de las anteriores

   Rpta: c
8. La escuela que incorpora la teoría de la probabilidad como base matemático fundamental de la estadística es conocida como la escuela:
   a. Demografía
   b. Francesa
   c. Inglesa
   d. Alemana
   e. Administrativa

   Rpta: b
9. La etapa …………. avance en estadística se caracteriza por el desarrollo de técnicas que permiten derivar conclusiones generales sobre una población en función del análisis de muestras representativas
   a. Inicial

b. Intermedia
c. Final
d. Actual
e. Ninguna es correcta

Rpta: d

10. La estadística inferencial tiene por objetivo:
    a. Describir colecciones de datos
    b. Obtener medidas de resumen estadístico
    c. Presentar los datos en forma de tablas
    d. Inferir propiedades de una población
    e. Categorizar las variables en estudio

    Rpta: c

11. Para que una característica o propiedad de un objeto se considere como variable debe tender a variar y:
    a. Presentarse siempre del mismo modo
    b. Ser susceptible de medición
    c. Producir diferentes resultados
    d. Ser un atributo del objeto
    e. Medirse numéricamente

    Rpta: b y c

12. El criterio de clasificación de las variables en cualitativas y cuantitativas se basa en:
    a. La escala de medición utilizada por la variable
    b. La relación entre las variables
    c. La categorización
    d. La naturaleza de la variable
    e. N.A.

    Rpta: d

13. Una variable continua es una variable………………que sólo puede tomar…………. valores dentro de un intervalo determinado.
    a. Cualitativa / todos los
    b. Dependiente / ciertos
    c. Interviniente / todos los
    d. Cuantitativa / cualquiera de los

e. Cuantitativa / infinitos

Rpta: e

14. Las variables que establecen categorías con un orden convencional y mantienen diferencias cuantificables entre ellas son conocidas como variables:
    a. Ordinales
    b. Nominales
    c. Categorización
    d. Definición
    e. Conceptualización

    Rpta: a

15. Las diversas clases que define el investigador para clasificar los diversos valores que puede tomar una variable se conocen como:
    a. Categorías
    b. Medidas de resumen
    c. Categorización
    d. Definición
    e. Conceptualización

    Rpta: b

16. En toda tabla o cuadro estadístico se pueden identificar tres elementos fundamentales:
    a. Unidades de análisis – Categorías – Variables
    b. Variables – Categorías – Información
    c. Unidades de análisis – Variables – Datos
    d. Encabezado – Cuerpo – Nota de pie
    e. Número – Título – Cuerpo

    Rpta: e

17. Las de tablas que existen son las tablas de frecuencias y los cuadros de análisis.
    a. Verdadero
    b. Falso

    Rpta: a

18. En un cuadro de análisis, los totales de cada columna se ubican en la……………. línea del cuerpo.
    a. Última
    b. Primera

c. N.A.

Rpta: b

19. La variable principal de un cuadro estadístico está precedida por la palabra……………. y está distribuida en……………………

a. Según / Columnas
b. Por / Filas
c. Según / Filas
d. N.A.

Rpta: d

20. Cuando usa una nota de unidad de medida, en un cuadro de análisis, ésta debe colocarse:

a. Debajo de la fuente
b. Debajo del título
c. Debajo de la nota de elaboración
d. Sobre la fuente
e. En cualquier lugar

Rpta: d

21. Al conjunto formado por todos los objetivos y hechos que existen se le denomina:

a. Existencia
b. Estado
c. Realidad
d. Población
e. Muestra

Rpta: c

22. Los conocimientos constituyen conjuntos de……………acerca de la realidad.

a. Objetos
b. Hechos
c. Elementos
d. Leyes
e. Ideas

Rpta: e

23. Las interrogantes in respuestas, en la relación sujeto realidad, corresponden al aspecto…….

a. Conceptual
b. Lingüístico
c. Formal
d. Fáctico
e. Ideal

Rpta: b

24. Para el trabajo científico, el aspecto más confiable y que constituye el elemento de referencia para contrastar las ideas es el aspecto:

a. Fáctico
b. Conceptual
c. Procesal
d. Lingüístico
e. Formal

Rpta: a

25. Lo que origina un problema es:

a. Un objeto
b. Un hecho
c. Una interrogante
d. Una respuesta
e. Ninguno

Rpta: c

26. La ciencia también permite realizar predicciones

a. Verdadero
b. Falso

Rpta: a

27. En un proceso de investigación en general, la evaluación deba realizarse:

a. Al final del proceso
b. Al término de cada etapa
c. Durante cada etapa
d. Al inicio de cada etapa
e. Solo por expertos

Rpta: c

28. ¿En qué fase del proceso de investigación estadística se deben seleccionar los procedimientos, herramientas y métodos para recopilar y examinar los datos?

a. Etapa de planeamiento o preparación
b. Etapa de recolección de datos
c. Etapa de organización y presentación de datos
d. Análisis e interpretación de datos
e. Formulación de conclusiones y preparación del informe

Rpta: a

29. En la investigación bibliografía correspondiente a la etapa de recopilación de datos en la investigación estadística, los datos se toman de las fuentes denominadas:
    a. Fuentes primarias
    b. Fuentes secundarias
    c. Fuentes terciarias
    d. Unidades de análisis
    e. Ninguna de las anteriores

    Rpta: b
30. La presentación de los datos en forma gráfica se hace:
    a. Antes de la forma tabular
    b. Inmediatamente después de conteo
    c. Como complemento de la forma tabular
    d. En muy raras ocasiones
    e. Solo en los informes finales

    Rpta: c
31. Para que una cualidad, características de un objeto se considere como variable
    a. Presentarse siempre del mismo modo
    b. Ser susceptible de medición
    c. Ser un atributo del objetivo
    d. Producir diferentes resultados
    e. Medirse numéricamente

    Rpta: b
32. Establezca con veracidad o falsedad de cada una de las proposiciones siguientes:
    a. La estadística no estudia los métodos científicos para recopilar, organizar, resumir y analizar datos

b. Es un conjunto de procedimientos y técnicas empleadas para recolectar, organizar y analizar datos.
c. La estadística descriptiva se refiere a la recolección, presentación, descripción, análisis e interpretación de una colección de datos
d. La inferencia estadística se basa en métodos que permiten generalizar o tomar decisiones basadas en información parcial obtenida mediante técnicas descriptivas.
e. La estadística se aplica a todas las áreas científicas y se considera una ciencia auxiliar que respalda la investigación en otras ciencias.
f. El punto de partida de toda investigación suele ser la identificación de un problema a resolver.
g. Las Etapas del Método Científico pueden variar, pero generalmente involucran más de tres etapas, como observación, hipótesis, experimentación, análisis y conclusiones.
h. El método científico de investigación se basa en dos tipos de razonamiento: el deductivo y el inductivo.
i. La recolección adecuada de datos no es de suma importancia para el investigador.
j. Toda información tiene dos aspectos fundamentales: la Fuente de obtención y método para su recolección.
k. Una encuesta es un conjunto de cuestionarios normalizados dirigidos a una muestra representativa de la población.
l. La población se refiere a la colección de todos los elementos que están siendo estudiando.
m. Un estadístico es una característica o medida calculada a partir de una muestra y se utiliza para estimar parámetros poblacionales.
n. La variable es la característica o atributo que se está estudiando en una investigación.
o. La muestra es una parte de la población la cual no es representativa

   Rpta: a (F), b (V), c (V), d (F), e (F), f (V), g (f), h (V), i (f), j (V), k (V), l (V), m (V), n (V), o (f).

33. Identificar la población (P) o muestra (M).
    a. El Producto Bruto Interno del Perú en el año 2022.................... ( )
    b. El ingreso per cápita de Huaraz en el año 2021...........................( )

c. Número de clientes de la empresa "Carsa" en el año 2021......... (   )
d. El 20% de las mujeres que trabajan y estudian en la "Universidad nacional de Jaén"..(   )
e. Se estudió el 30% de los internos de los penales del departamento de Cajamarca..(   )
f. Se realiza un estudio los antecedentes penales de 20 personas en el poder judicial año 2022..(   )

Rpta: a (P), b (M), c (P), d (M), e (M), f (M).

34. Con base a la población de todos los docentes peruanos que trabajan en la Universidad Nacional de Jaén, identificar los siguientes como constante (C) o variable (V)

a. Asociación donde pertenecen...........( )
b. Ingreso económico............................( )
c. Número de hijos................................( )
d. Número de cursos..............................( )
e. Estatura...( )
f. Nacionalidad....................................( )
g. Edad..( )
h. Centro de trabajo................................( )
i. Numero de investigaciones................( )
j. Cargo que desempeñan.....................( )
k. Facultan a donde pertenecen..............( )
l. Categoría...( )
m. AFP donde pertenece……………….( )

Rpta: a (C), b (V), c (V), d (V), e (V), f (C), g (V), h (C), i (V), j (V), k (V), l (V), m (V)

35. Clasificar, si los enunciados son variables cualitativas (nominal u ordinal) o cuantitativas (discretas o continuas):

- Especialidades que ofrecen la universidad "Santiago Antúnez de Manolo" (Cualitativa nominal)
- Notas de evaluación de los estudiantes de la Facultad de Ingeniería (Cuantitativa continua)
- Número de aulas en la ciudad universitaria "Universidad Nacional de Jaén" (Cuantitativa discreta)

- Ingreso económico de un grupo de familias (Cuantitativa continua)
- Tiempo de servicio (Cuantitativa continua)
- Institución de procedencia (Cualitativa nominal)
- Empleado de una empresa según:
  - Ingreso mensual (Cuantitativa continua)
  - Sexo (Cualitativa nominal)
  - Edad (Cuantitativa discreta)
  - Religión (Cualitativa nominal)
  - Grado de instrucción (Cualitativa ordinal)
  - Hora de entrada de una empresa (Cuantitativa continua)

36. Utilizar "A" para las series que están formadas por datos cualitativos, es decir, aquellos que representan atributos o características distintivas, y "V" para las series formadas por datos cuantitativos, que son valores numéricos medibles.
    a. Nacionalidad (A)
    b. Edad (V)
    c. Filiación política (A)
    d. Estado civil (A)
    e. Profesión (A)
    f. Número de pulsaciones por minuto (V)
    g. Gastos mensuales en alimentación (V)
    h. Número de llamadas telefónicas por día (V)
    i. Condición del que habita un inmueble (A)
    j. Puntaje obtenido en la prueba de Estadística (V)
    k. Número de aulas en una academia (V)
37. Identificar cada una de las propiedades como constante (C) o variables (V) en relación con el estudio de recopilación de información de ingresos anuales y el número de años de los profesores peruanos.
    a. Sexo (V)
    b. Nacionalidad (C)
    c. Profesión (C)
    d. Número de años de experiencia (V)
38. Un fabricante de medicamentos está interesado en determinar qué fracción de personas con hipertensión arterial puede manejar su condición con un nuevo producto que han desarrollado. Para esto, se llevó a cabo un estudio con 1200

personas diagnosticadas con hipertensión, y los resultados mostraron que el 80% de estos individuos logró controlar su hipertensión con el medicamento. Se asume que estos 1200 participantes son representativos del grupo general de personas con hipertensión

a. Objetivo del estudio: Conocer la proporción de personas con hipertensión arterial
b. Población: 1200 participantes
c. Muestra: 80% que pudieron controlar su hipertensión con el medicamento
d. Variables: producto, hipertensión, medicamento.
e. Estadístico: 80%

8.2 CAPÍTULO 2

1. En la tabla estadística se puede identificar tres elementos fundamentales:
   a. Unidad de análisis – Categoría – Variables
   b. Categoría – Variables – Información
   c. Unidad de análisis – Variables – Datos
   d. Encabezado – Cuerpo – Nota de pie
   e. Número – Titulo – Cuerpo

   Rpta: c
2. Los dos tipos de tablas que existen son: tabla de frecuencias y cuadros de análisis
   a. Verdadero
   b. Falso

   Rpta: a
3. En el cuadro de análisis, los totales de cada columna se ubican en la………línea del cuerpo.
   a. Última
   b. Primera
   c. Tercera
   d. No se escribe los resultados
   e. Ninguna de las anteriores

   Rpta: a
4. La variable de un cuadro está precedida por la palabra……y está distribuida en……...
   a. Según / columna
   b. Por / filas
   c. Por / columnas
   d. Según / filas
   e. Ninguno de las anteriores es correcto

   Rpta: c
5. Una tabla de frecuencias se elabora con la finalidad de:
   a. Establecer las frecuencias absolutas
   b. Realizar cálculos a partir de ella
   c. Analizar una muestra

d. Establecer las frecuencias relativas
e. Incluirla en un informe
Rpta: b

6. En toda tabla de frecuencias se cumple que:
   a. $\sum f_i = n$
   b. $\sum h_i = n$
   c. $\sum H_i = n$
   d. $\sum h_i = 1$
   e. N.A
   Rpta: a y d
7. Para elaborar una tabla de frecuencias con 85 datos ¿cuántos intervalos se necesitan?
   a. 6
   b. 7
   c. 8
   d. 9
   e. N. A
   Rpta: b
8. Al elaborar una tabla de frecuencias se tiene R = 80 y m=8. la amplitud de los intervalos será:
   a. 10
   b. 11
   c. 12
   d. 13
   e. N. A
   Rpta: a
9. No es un elemento fundamental de una gráfica estadística:
   a. Número
   b. Título
   c. Diagrama
   d. Encabezado
   e. Fuente Rpta: d
10. El gráfico estadístico en el que usa iconos de diferentes tamaños para representar los valores de una variable, se denomina:

a. Gráfico lineal
b. Gráfico de superficie
c. Pictograma
d. Gráfico de dimensiones
e. Cartógrafa

Rpta: c

11. En un gráfico de frecuencias absolutas, los valores de la variable se representan a través de:
    a. Rectángulos verticales
    b. Segmento de recta, verticales
    c. Rectángulos horizontales
    d. Líneas diagonales
    e. N.A.

    Rpta: b

12. En un histograma que representa una distribución con intervalos de amplitudes diferentes, el eje vertical muestra.
    a. Las frecuencias relativas
    b. Las frecuencias absolutas
    c. Las amplitudes
    d. La densidad de frecuencia
    e. N.A.

    Rpta: d

13. En un diagrama escalonado:
    a. El polígono de frecuencias
    b. Histograma
    c. Diagrama acumulativo de frecuencias
    d. Ojiva
    e. Pictograma

    Rpta: c

14. Se refiere “Ojiva” al:
    a. Polígono de frecuencias
    b. Polígono acumulativo de frecuencias
    c. Pictograma
    d. Diagrama de sectores circulares

e. N.A.

Rpta: b

15. El diagrama de barras, para variable cualitativa:
   a. Debe haber la misma separación entre una barra y la siguiente
   b. Las barras deben estar unidas, sin separación
   c. a y b
   d. N. A

   Rpta: d

16. En un diagrama de sectores circulares el ángulo correspondiente a cada sector, se calcula mediante la siguiente relación:
   a. $f_i \times 360^0$
   b. $\dfrac{360^0}{f_i}$
   c. $360^0 - f_i$
   d. $h_i \times 360^0$
   e. $H_i \times 360^0$

   Rpta: d

17. Si en la Universidad Nacional de Jaén se selecciona una muestra de 620 estudiantes de diversas facultades y 130 de ellos pertenecen a la facultad de Ingeniería, el ángulo que representaría en un diagrama de sectores circulares sería:
   a. 37^0
   b. 53^0
   c. 65^0
   d. 68^0
   e. 71^0

   Rpta: c

18. Los ingresos mensuales de un grupo de pequeños comerciantes se organizaron en una distribución de frecuencias simétrica de 5 intervalos de igual amplitud resultando: ingreso mínimo 125 nuevos soles; marca de clase del cuarto intervalo es 300 nuevos soles. Si el 8% de los ingresos son menores que 165 nuevos soles y el 70% de los ingresos son menores a 275 nuevos soles ¿Qué porcentaje de los ingresos son superiores a 285 nuevos soles?

Rpta. C =50 Intervalos: 125, 175, 225, 275, 325, 375 Frecuencias relativas: 0.10, 0.20, 0.40, 0.20, 0.10 % de casos superiores a 285 es: 26%

19. La distribución del tiempo en minutos que 100 obreros tardaron en completar cierta tarea, se ha presentado en una tabla de frecuencias con cuatro intervalos de igual amplitud cuyo histograma correspondiente es simétrico si el intervalo $I = [6, ?)_i$, la frecuencia absoluta $f_2 = 2f_1 + 5$, y si se sabe que el 85% de los obreros demoran menos de 12 minutos. Completar la distribución de frecuencia. Rpta. Y_o = 6 ; C = 2 $f_i = 15{,}35{,}35{,}15$

20. Las calificaciones obtenidas en un examen se organizaron en una tabla de distribución que muestra las frecuencias relativas, cada uno una amplitud de 5 puntos. Si la calificación mínima es igual a 5, el 48% de las notas son menores que 12 y el 80% de las calificaciones son inferiores a 16, reconstruir la distribución de frecuencias. Rpta. Frecuencias relativas: 0.30, 0.45, 0.25

21. La puntuación obtenida en una prueba de aptitud se ha organizado en una tabla de distribución con 6 intervalos de igual tamaño. Dada la información proporcionada, que incluye las marcas de clase $Y_2 = 40 y Y_4 = 80$, frecuencias$h_1 = h_6$, $h_3 = h_5$, $h_4 = 0.25$, $h_2 = h_4 - h_1$, $h_3 = h_1 + 0.10$ y $F_6 = 60$, se requiere completar la distribución de frecuencias absolutas y crear un polígono que represente estos datos gráficamente.

    Rpta. Y_o = 10, C = 20 Frecuencias absolutas: 6, 9, 12, 15, 12, 6 n = 60

22. Se han recopilado datos sobre el tiempo (horas) que 120 familias utilizan su computadora, se tabularon en una distribución de frecuencias de 5 intervalos de amplitud iguales a 4, siendo; el tiempo mínimo de uso 2 horas, la primera y segunda frecuencias iguales al 10 y 15% del total de casos respectivamente. Si el 73.75% de las familias usaron menos de 17 horas y el 85% menos de 19 horas, determine las frecuencias.

    Rpta. $(\ ]$: n = 120 f_i: 12, 18, 36, 30, 24

23. Los salarios que una empresa paga a sus practicantes oscilan entre S/.150 y S/.270. Si los salarios se agrupan en cuatro intervalos de clase de longitudes iguales, se observa que el 40% de los practicantes ganan S/. 195 o menos, el 80% ganan S/.225 o menos, y el 15% ganan más de S/.232.50.

    a. Calcular el porcentaje de practicantes en cada intervalo.

b. Si la empresa establece un ingreso mínimo de S/.240 y aumenta a todos los practicantes en una cantidad uniforme para que el 20% supere el ingreso mínimo ¿Cuál sería el monto del aumento?

Rpta. Frecuencias relativas: 0.10, 0.60, 0.20, 0.10 b) $15.

24. Se llevó a cabo un estudio acerca del tiempo de espera de los clientes (en minutos) en el banco de la "Nación", utilizando una muestra de 50 clientes y se obtuvieron los resultados:

73	65	82	70	45	50	70	54	32	75
75	67	65	60	75	87	83	40	72	64
58	75	89	70	73	55	61	78	89	93
43	51	59	38	65	71	75	85	65	85
49	47	55	60	76	75	69	35	45	63

Elabore una tabla de frecuencias utilizando un número apropiado de clases para responder las preguntas siguientes:

a. ¿Cuántos clientes esperaron entre 35 y 52 minutos?
b. ¿Cuántos clientes esperaron más de una hora para ser atendidos?
c. ¿Qué porcentaje de clientes esperaron menos de 92 minutos?

Rpta. Intervalos: 31,40,49,58,67,76,85,95 $f_i = 3,5,6,11,15,4,6$

a)14 b) 25 c) 98%

25. La información sobre los egresos (en miles de soles) de 20 fábricas de la ciudad de San Ignacio se encontraba registrada en un archivo, pero lamentablemente se perdió, y ahora solo se cuenta con algunos datos:

$Y_1 = 22 Y_3 = 30 F_3 = 7 f_1 = 22 h_2 = 0.15 H_4 = 0.85$

a. Elaborar una tala que muestre la distribución de frecuencias
b. Interpretar $F_3, H_1, h_2, H_4^*, F_4^*$
c. ¿Cuántas fabricas tienen un egreso de 22 a menos de 30 mil nuevos soles?
d. ¿Cuál es el porcentaje de empresas que tienen un egreso de 34 a 38 mil nuevos soles?

Rpta. Intervalos: 20 ,24, 49, 28, 32, 36, 40 f_i : 2, 3, 6, 6, 3

26. Los capitales de inversión (en miles de soles) de 200 empresas constructoras están distribuidos de manera simétrica en siete intervalos de clase de igual amplitud, asimismo se conocen los siguientes datos:

$C = 10 f_1 = 8 Y_3 f_3 = 1260 f_2 + f_3 = 62 H_2 = 0.14 h_3 = 0.21 H_6 = 0.96$

Crear nuevamente la tabla de distribución de frecuencias.

a. ¿Cuántas fabricas tienen una inversión de 22 a menos de 42 mil nuevos soles?

b. ¿Cuál es el porcentaje de empresas que tienen una inversión de 34 a 58 mil nuevos soles?

c. ¿Cuál es la proporción de empresas que tienen una inversión de 50 a 70 mil nuevos soles?

Rpta. Inter: 5, 15, 25, 35, 45, 55, 65, 75. f_i : 8, 20, 42, 60, 42, 20, 8.

27. Completar la siguiente tabla: Indicar el límite inferior y superior de cada clase.

Intervalos	Marca de clase	Frec. Relativa	Frec. Relativa acumulada
[,]		0.10	
[,]	6		0.25
[,]		0.55	
[,]	14		0.9
[, >		0.10	

28. Se eligieron de manera aleatoria 20 baterías de un proceso de fabricación, y los datos siguientes muestran el tiempo de duración en horas para cada una de las 20 baterías.

52.4	62.6	58.8	65.6	49.2
58.8	57.2	60.3	59.5	58.0
62.3	64.3	52.6	54.8	48.7
56.7	53.0	58.6	61.5	63.2

a. Crear una tabla de distribución que muestre las frecuencias de los datos.

b. Generar un gráfico que represente la variable de estudio de acuerdo a los datos recopilados.

c. Interpretar f_2, f_3, h_2%, H_1

29. En el siguiente cuadro muestra la distribución de 40 familias según el número de hijos.

N° de Hijos Xi	N° de familias Fi
0	4
1	5
2	13
3	6
4	7
5	5

a. Construir una tabla que refleje la distribución de frecuencias de acuerdo al estudio de la variable
b. Graficar de acuerdo a la variable de estudio según los datos obtenidos
c. Interpretar f_1, f_3, $h_3\%$, H_1

30. En una encuesta de opinión que aborda las preferencias de bebidas gaseosas según su color, con las opciones de Blanco (B), Negro (N), y Rojo (R); se obtuvieron de una muestra de 20 consumidores que indicaron sus preferencias de la siguiente manera:

B	N	B	B	B	N	R	R	N	R
N	N	B	B	B	N	R	N	N	R

a. Elabore una tabla que muestre la distribución de frecuencias e interprete f_3, h_2, $h_2\%$
b. Construya un gráfico apropiado que represente la variable.

31. Se proporciona la distribución de frecuencias que describe el peso en gramos de las dosis de un antibiótico administrado al ganado para combatir una enfermedad. La precisión en la medición del peso de las dosis es crucial, ya que una sobredosis podría tener efectos perjudiciales en los animales.

Peso(gr)	fi
[16 - 21 >	8
[21 - 26 >	26
[26 - 31 >	32
[31 - 36 >	21
[36 - 41]	12

a. Identificar la variable en estudio
b. Construir una tabla de frecuencias simples y acumuladas.
c. Graficar de acuerdo a la variable en estudio

8.3 CAPÍTULO 3

1. Los estadísticos son medidas de resumen de una muestra que se utilizan con fines:
   a. De ilustración
   b. De resumen
   c. Descriptivos o analíticos
   d. Sustitución a un gráfico
   e. N. A

   Rpta. c
2. La media es un estadígrafo
   a. Posición
   b. Dispersión
   c. Asimetría
   d. Apuntamiento
   e. Kurtosis

   Rpta. a
3. Para determinar la mediana, es necesario organizar los datos previamente……………
   a. Simplificados
   b. Redondeados
   c. Sin redondear
   d. Ordenados
   e. N. A

   Rpta. d
4. La mediana divide la distribución en dos partes iguales, mientras que la moda corresponde al valor que tiene la frecuencia más alta
   a. Verdadero
   b. Falso

   Rpta. a
5. En los cuantiles no se encuentran:
   a. El decil
   b. El cuartil
   c. La moda
   d. La mediana

e. Percentil

Rpta. c y d

6. La información se refiere a la inversión anual en miles de dólares de una muestra compuesta por 40 pequeñas empresas.

11	27	28	34	04	15	18	41	12	18
36	29	33	27	25	33	22	31	35	17
20	10	25	24	39	30	26	46	23	19
37	24	31	28	28	23	29	21	45	23

Encontrar:

a. La media aritmética

b. La mediana

c. El cuantil uno

d. Decil tres

e. El percentil cincuenta

f. Moda

Rpta C = 6. m = 7 $y'_0 = 4$ $y'_7 = 46$

a. $\bar{x} = 26.8 \approx 27$ b. $M_e = 27$ c. $Q_1 = 22$ d. $D_3 = 23$

e. $P_{50} = 27$ f. $M_O = 27.143$

7. Los datos representan las edades de 60 turistas que se hospedaron en el hotel "La Joya" durante el mes de febrero del presente año.

54	23	23	27	22	20	29	43	33	28
23	33	25	58	32	41	41	50	39	50
26	35	26	36	49	61	57	25	31	52
27	63	46	24	44	31	37	30	25	26
45	42	34	30	63	30	63	69	35	38
35	38	28	68	48	35	67	35	34	35

Encontrar:

a. La media aritmética

b. La mediana

c. El cuantil uno

d. Decil tres

e. El percentil cincuenta

f.Moda

Rpta C = 7 m = 7 $y'_0 = 20$ $y'_7 = 69$

a. $\bar{x} = 39.25$ b. $M_e = 37$ c. $Q_1 = 28.6$ d. $D_3 = 30.2$

e. $P_{40} = 33.5$ f. $M_o = 34$

8. En una evaluación, cinco alumnos obtuvieron cada uno una calificación de doce, y un alumno obtuvo una calificación de dieciocho. Si se indica que la calificación promedio es de trece, se plantea las preguntas ¿Qué nota promedio es? ¿Es el promedio adecuado? ¿Cuánto es el promedio adecuado? Rpta. a) 15.6 b) La media aritmética no es adecuada porque es afectado por el alumno que tiene 18 c) Me =12
9. Para calcular el suministro de energía eléctrica de una ciudad, se utilizan los datos proporcionados mensuales por la empresa eléctrica. Se seleccionan al azar 15 familias de la ciudad y se obtienen los resultados en vatios (wats):

11.2	21.5	16.4	19.7	14.6	16.9	32.2	18.2
13.1	23.8	18.3	15.5	18.8	22.7	14.0	

10. Si en la ciudad hay 5000 familias ¿Cuántos wats se requiere mensualmente si el consumo promedio por familia permanece igual?

   Rpta. $\bar{x} = 18.46$ si N = 5 000 el consumo en toda la población es 92300
11. Dado un ahorro inicial de \$100 que acumula intereses variables del 3, 5 y 8% durante tres años, se busca calcular:
   a. La tasa de crecimiento anual del ahorro
   b. La tasa promedio de crecimiento del ahorro en un periodo de tres años

   Rpta a. 103 108.5 116.802

   b. $\bar{x}_G = 1.5313$ tasa 5.313%

12. Tres mecanógrafas tienen velocidades de escritura de 40, 30 y 80 palabras por minuto respectivamente. Si cada una de ellas escribe un texto, calcular la velocidad media de escritura.

   Rpta. $\bar{x}_A = 52.2$
13. El rendimiento de tres automóviles que recorrieron una distancia de 200 Km cada uno. Los rendimientos de estos automóviles fueron 50, 45 y 60, Km. por galón respectivamente. Calcular el rendimiento promedio de los tres automóviles. Rpta $\bar{x}_A = 50.94$

14. Las llamadas telefónicas registradas en el departamento de Ciencias Básicas y Aplicadas de la Universidad Nacional de Jaén, durante un día específico tienen una duración en segundos y los datos son:

$[L_i - L_s)$	y_i	n_i	N_i	h_i	H_i
70 –			10		
–				0.20	
–					0.70
–		22			
– 95					
		2			
Total		120			

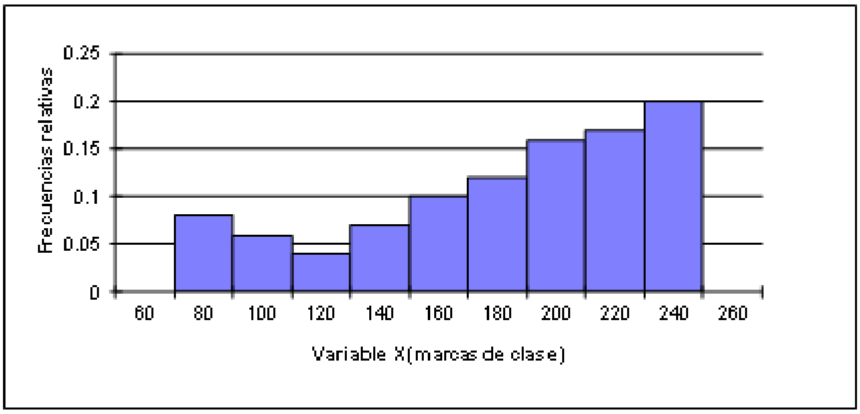
a. Completa la tabla de frecuencias.
b. Calcula la moda e interpreta los resultados.
c. ¿Cuántas llamadas telefónicas duraron menos de 95 segundos?

15. La tabla proporciona la distribución de los importes de las facturas por reparación de carrocería (en miles de soles) de una muestra de 80 vehículos en un taller.

Importe	y_i	Número de vehículos n_i	N_i	$y_i n_i$
0 – 60		10		
60 – 80		20		
80 – 120		40		
120 – 180		10		

a. Calcular el importe medio y la mediana
b. ¿Cuál es la interpretación de los deciles en esta situación? Calcular el tercer decil.

16. En histograma muestra la distribución de los salarios (variable X), en miles de soles de una industria del sector cerámico:



a. De acuerdo con estos datos, determinar el salario mediano y moda

17. El número de aplicativos en Android desarrollados por 20 programadores en una empresa de desarrollo de software en el año 2018 se presenta en la tabla de frecuencias:

Número de aplicativos	Número de programadores
1	4
2	5
3	8
4	6
5	2

a. Calcular e interpretar el ingreso medio

18. Los salarios proporcionados por una empresa hotelera a sus practicantes se muestran en la tabla.

Salario	Número de practicantes
[650 – 700 >	6
[700 – 750 >	8
[750 – 800 >	12
[850 – 900 >	7
[900 – 950]	5

a. Calcular el importe medio e interpretar el resultado
b. La mediana
c. La moda
d. Calcular e interpretar el percentil 75%
e. Calcular el cuartil 3
f. Calcular el decil 4

8.4 CAPÍTULO 4

1. Los datos representan las calificaciones obtenidas por 20 estudiantes sometidos a un test. Los estudiantes fueron divididos en dos grupos

A	86	81	79	73	95	86	94	90	86	88
b	84	96	70	100	79	84	96	96	99	84

   a. Calcular el promedio y la dispersión de cada grupo, ¿Cuál de los grupos es más homogéneo?

   b. ¿Cuál de los estudiantes que tienen 79 puntos en cada uno de los grupos, están mejor en su grupo?

   Rpta a. $CV_1 = 0.074$ $CV_2 = 0.107$

   b. $Z_1 = 1.066$ $Z_2 = 1.03$

   el segundo grupo es mejor

2. Dado que la media es de 7.9 calcular la varianza y la desviación estándar de esta distribución

Notas de estudiantes	h_i
[0.5 - 2.5 ›	0.02
[2.5 - 4.5 ›	0.10
[4.5 - 6.5 ›	h_3
[6.5 - 8.5 ›	016
[8.5 - 10.5 ›	h_5
[10.5 – 12.5 ›	0.10
[12.5 – 14.5 ›	0.02
[14.5 - más ›	0

3. Se presenta una distribución que representa a 80 profesores según su tiempo de servicio en la docencia.

Tiempo de servicio (años)	Número de profesores
[0 - 4 ›	10
[4 - 8 ›	13
[8 - 12 ›	20
[12 - 16 ›	17
[16 - 20 ›	12
[20 - 24 ›	3
[24 - 28 ›	3
[28 - 32 ›	1

Calcular e interpretar:

a. La desviación de medias
b. La desviación cuartílica
c. La varianza utilizando por el método abreviado
d. La desviación estándar
e. Coeficiente de variación

4. La tabla muestra las calificaciones de 40 alumnos en el curso de Estadística y Probabilidades.

Calificaciones Li - Ls	N´° de alumnos f_i
[03 -06 >	2
[06 -09 >	5
[09 -12 >	11
[12 -15 >	5
[15 -18 >	2
[18 -21 >	6
[21 -24 >	4
[24 -27 >	5

Se pide:

a. Calcular e interpretar los siguientes estadísticos: varianza, desviación estándar y coeficiente de variación

5. Un fabricante de componentes electrónicos desea conocer la duración el tiempo de vida (en horas) de sus dispositivos, para lo que ha obtenido una muestra de 14 observaciones:

124	117	121	131	123	111	176
127	126	111	120	122	124	132

Calcule e interprete el coeficiente de variación para estos datos.

6. Se tiene dos muestras con las calificaciones de 11 estudiantes en sus exámenes parciales

grupo 1	61	75	67	35	61	67	58	72	40	58
grupo 2	73	83	76	47	74	75	71	83	61	62

Calcular e interpretar el coeficiente de variación, para determinar qué grupo de datos exhibe mayor homogeneidad o heterogeneidad.

8.5 CAPÍTULO 5

1. Se establece una relación entre el peso de los niños (variable X) desde la infancia a la adolescencia y su estatura (variable Y) en cm, los datos se presentan en la tabla:

X:	40	60	85	43	72	94	70	42	96	100
Y:	140	147	143	160	166	169	172	180	191	200

X:	52	57	65	80	95	76	91	70	55	100
Y:	145	158	163	172	176	180	192	195	197	200

a. Calcular la tabla de distribución de frecuencias absolutas simples y acumuladas.
b. Determinar la tabla de distribución de frecuencias relativas simples y acumuladas
c. Realizar las tablas de distribución de los porcentajes simples y acumulados.
d. Calcular las tablas de las distribuciones de (X e Y).
e. Elabora el gráfico con los porcentajes simples:
f. La distribución de y condicionada a que x = [40 – 52) x = [88 – 100].
g. La distribución de x condicionada por y = [152 – 164) y = [176 – 188)
h. Promedio de x, promedio de y, varianza muestral de x, varianza muestral de y, desviación estándar de x y desviación estándar de y.
i. La Cov (x,y)
j. V (x + y), V (x – y)

2. El investigador está examinando posible relación entre la inteligencia de los niños, medida mediante el coeficiente intelectual (CI en puntos) y la cantidad de hermanos que tienen. Para su estudio, ha seleccionado al azar a un grupo de 40 niños y ha recopilado los datos.

CI	110	115	120	118	110	108	105	104	98	99	98	100	90	93	90
Hermanos	0	1	1	1	2	2	2	3	3	4	4	5	5	5	6
fi	2	5	3	2	4	2	3	2	1	4	3	2	3	1.	2.

a. Construya la tabla de distribución de frecuencias absolutas simples y acumuladas e interprete.
b. Realizar la tabla de distribución de frecuencias relativas simples y acumuladas. Interprete.

c. Determinar la tabla de distribución de los porcentajes simples y acumuladas. Interprete.
d. Elabore las tablas de distribuciones marginales (x e y). Interprete.
e. Obtener el gráfico con los porcentajes simples e interpretarlo
f. La distribución de y condicionada a que x= [110 115) x= [95 100)
g. La distribución de x condicionada por y=1 y=5
h. Promedio de x, Promedio de y, varianza muestral de x, varianza muestral de y, desviación estándar de x y desviación estándar de y. Interprete.
i. La $Cov\ (x, y)$
j. $V\ (x + y)$, $V\ (x - y)$

3. Se ha administrado un cuestionario a un grupo de 40 alumnos con el propósito de evaluar su vocabulario X, y su capacidad de razonamiento Y. a partir de los resultados obtenidos, se han calculado los estadísticos:
   $\Sigma xi = 1800$, $\Sigma xi^2{}_i = 86000$ $\Sigma yi = 1390$, $\Sigma yi^2{}_i = 53100$, $\Sigma xiyi = 66500$

a. Promedio de x, Promedio de y, varianza muestral de x, varianza muestral de y, desviación estándar de x y desviación estándar de y. Interprete.
b. La $Cov\ (x, y)$
c. $V\ (x + y)$, $V\ (x - y)$

4. En una clínica se obtiene la edad de 10 pacientes con su debida estatura, así como indica la tabla.

X:	18	20	22	23	19	20	18	23	21	22
Y:	160	165	165	155	170	168	150	164	156	156

a. Construir la tabla de distribución de frecuencias absolutas simples y acumuladas.
b. Elabore la tabla de distribución de frecuencias relativas simples y acumuladas.
c. Presentar la tabla de distribución de los porcentajes simples y acumuladas.
d. Indicar las tablas de distribuciones marginales (x e y).
e. Obtener el gráfico con los porcentajes simples e interpretarlo
f. La distribución de y condicionada a que x=18
g. La distribución de x condicionada por y =[155 – 160>

h. Promedio de x, Promedio de y, varianza muestral de x, varianza muestral de y, desviación estándar de x y desviación estándar de y.

5. En el colegio se han seleccionado muestras a 20 niños que tiene problemas de obesidad (x), y se ha medido su presión sanguínea, los datos obtenidos se presentan en la tabla:

X:	10	12	13	11	14	10	13	14	11	12
Y:	89	96	105	110	116	103	99	110	107	113
fi	3	2	2	2	1	2	2	3	1	2

a. Presentar la tabla de distribución de frecuencias absolutas simples y acumuladas
b. Elabore la tabla de distribución de frecuencias relativas simples y acumuladas.
c. Construir la tabla de distribución de los porcentajes simples y acumuladas.
d. Analizar las tablas de distribuciones marginales (x e y).
e. Obtener el gráfico con los porcentajes simples e interpretarlo.
f. La distribución de y condicionada a que x=12.
g. Promedio de x, Promedio de y, varianza muestral de x, varianza muestral de y, desviación estándar de x y desviación estándar de y

8.6 CAPÍTULO 6

1. Se presentan datos relacionados con los ingresos (X) y el consumo (Y) de seis familias que se dan:

Consumos $ (Y)	Ingresos $ (X)
30	35
35	40
30	38
50	55
35	42
50	60

Se solicita:

a. Mostrar los datos en un gráfico de dispersión.
b. Determinar la ecuación de la recta de regresión y representar en un gráfico.
c. Calcular el coeficiente de determinación (r^2) e interpretar.
d. Determinar el consumo esperado para una familia si su ingreso es de $ 45.00.

2. El propietario de una empresa agrícola en Perú buscar establecer si existe relación del costo por accidente (dólares) de sus operarios con el número de accidentes ocurridos en la jornada de trabajo. El dueño registró información de ambas variables de 7 jornadas de trabajo, los resultados se dan a continuación en la tabla:

N° de accidentes(X)	5	7	3	8	4	2	6
Costo por accidente(dólares) Y	330	210	124	270	160	80	190

a. Presentar los datos en un diagrama de dispersión
b. Calcular la ecuación estimada de regresión lineal simple
c. Calcular e interpretar el coeficiente de determinación

3. En la tabla, se tiene las puntuaciones: X (inteligencia) e Y (rendimiento académico)

X	104	106	115	102	125	138	125	113	130	119	106
Y	3	5	7	1	5	9	8	4	11	7	6

a. Calcular la ecuación estimada de regresión lineal simple

b. Calcular e interpretar el coeficiente de determinación

4. Una empresa que se dedica a la venta de Arándanos busca realizar proyecciones sobre sus ventas totales en un país. Su objetivo es investigar la posible relación entre las ventas y el ingreso nacional. Para investigar la relación cuenta con los siguientes datos que se presentan a continuación:

X	188	191	207	228	240	253	258	276	294	309	319
Y	401	405	411	426	431	437	443	449	459	471	471

Dónde: X representa la renta nacional en millones de soles e Y representa las ventas de la empresa en miles de soles en el periodo que va desde 2013 hasta 2023

La recta de regresión de Y sobre X.

a. Calcular la recta de la regresión de Y sobre X

b. Calcular e interpretar el coeficiente de correlación lineal

c. Si en el 2023 la renta nacional del país fue de 337 millones de soles. ¿Cuál será la predicción para las ventas de la empresa en el año 2024?

REFERENCIAS BIBLIOGRÁFICAS

Ávila Acosta, Roberto1997 “Estadística elemental”. Editorial Estudios y Ediciones. Lima.

Moya Calderón, Rufino1991.” Estadística Descriptiva Conceptos y aplicaciones”. Primera Edición. Edit, San Marcos lima, Perú.

García Ore, Celestino 2011,” Estadística Descriptiva y Probabilidades para ingenieros”. Edit. MACRO EIRL. Lima-Perú.

Córdova Zamora, Manuel 2009.” Estadística Descriptiva e inferencial” Tercera edic. Edit. Moshera R.L. Lima -Perú

Martínez Bencardino, Ciro 2012 “Estadística y Muestreo, Cuarta Edición. México

Douglas C. Montgomery y George C. Runger 2002 “Probabilidad y Estadística” aplicados a la ingeniería, Edición 2, Editor Limusa Wiley

Moya Calderón, Rufino 2009 “Probabilidad e Inferencia Estadística, Tercera Edición. Lima –Perú, San Marcos.

Irwin Miller, Jhon ,E. Freund 2008,”Probabilidad y Estadística” para ingenieros. México. Reverte S.A

Mendenhall William, T y Sincich 2005 “Probabilidad y Estadística”, para ingenieros y ciencias. México Pretice –Hall Hispano Americana S.A (5ta Edición)

E.Walpole Ronald, H.MYERS Raymond , L.MYERS Sharon Y KEYING YE 2012, “Probabilidad y estadística para ingeniería y ciencias” (9na edición) México

Willian Navidi. (2006). Estadística Para Ingenieros. México: McGraw Hill/Interamericana Editores, S.A. DE C.V.

Delgado de la Torre, R (2007) Probabilidad y estadística para ciencias e ingenierías. Madrid. Delta

Martínez, C. (2002). Estadística y Muestreo. México Ecoe Ediciones.

Montoro Lorenzo J.M. (2007) Estadística Descriptiva. Madrid. Thomson Paraninfo.

Ronald E. Walpole y Raymond H. Myers. (1989). Probabilidad y estadística. México Cuarta Edición. Mc Graw Hill Interamericana.

Martínez, C. (2002). Estadística y Muestreo. México. Ecoe

Murray R. Spiegel. (2009). Estadística teoría y problemas resueltos. Quinta Edición. México. McGraw.

Navidi W. (2006). Estadística para ingenieros y científicos. México. Capítulo I y VII. Editorial McGraw Gill