



INSTITUTO SUPERIOR
UNIVERSITARIO

SUCE

**GUÍA GENERAL DE ESTUDIO
DE ESTADÍSTICA**



Guía General de Estudio de Estadística

Francisco Felipe Soria Suárez

Paola Estefanía Gonzáles Mendoza

Patricio Gonzalo Endara Calahorrano

2026

Esta publicación ha sido sometida a revisión por pares académicos específicos por:

Edwin Manuel Castelo Vinueza

Universidad Central del Ecuador

Corrección de estilo:

- Teresa Del Carmen Ramírez - Docente - Sucre

Diseño y diagramación:

- Freddy Javier Centeno Martínez - Docente - Sucre

Editorial RIMANA

Primera Edición
Quito – Ecuador

INSTITUTO SUPERIOR UNIVERSITARIO SUCR

ISBN: 978-9942-590-04-6

Esta publicación está bajo una licencia de Creative Commons Reconocimiento-No Comercial-Compartir Igual 4.0 Internacional.



MISIÓN

Ser una Institución Superior Universitaria con estándares de calidad académica e innovación, reconocida a nivel nacional con proyección internacional.

VISIÓN

Formamos profesionales competentes con espíritu emprendedor, capaces de contribuir al desarrollo integral del país.

Los contenidos de este trabajo están sujetos a una licencia internacional Creative Commons Reconocimiento-No Comercial-Compartir Igual 4.0 (CC BY-NC-SA 4.0). Usted es libre de Compartir — copiar y redistribuir el material en cualquier medio o formato. Adaptar — remezclar, transformar y construir a partir del material citando la fuente, bajo los siguientes términos: Reconocimiento- debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante. No Comercial-no puede hacer uso del material con propósitos comerciales. Compartir igual-Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la misma licencia del original. No puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.
<https://creativecommons.org/licenses/by-nc-sa/4.0/>



**Reconocimiento-NoComercial-CompartirIgual
4.0 Internacional (CC BY-NC-SA 4.0)**

Usted acepta y acuerda estar obligado por los términos y condiciones de esta Licencia, por lo que, si existe el incumplimiento de algunas de estas condiciones, no se autoriza el uso de ningún contenido.

ÍNDICE DE CONTENIDO:

<i>Presentación de la asignatura</i>	7
<i>Resultados del aprendizaje</i>	8
UNIDAD 1: FUNDAMENTOS DE LA ESTADÍSTICA	9
Definiciones básicas.....	9
Importancia de la estadística	9
Investigación científica	9
Conocimiento científico.....	10
Variables	10
Clasificación de variables	10
Variable Independiente	11
Variable dependiente	11
Variable de control.....	12
Descripción de datos	13
Frecuencias	13
Frecuencias absolutas, relativas y acumuladas	14
UNIDAD 2: MEDIDAS ESTADÍSTICAS	16
Medidas de tendencia central.....	16
Medidas de posición.	19
Medidas de dispersión:	21
Medidas de forma:	23
UNIDAD 3 ESTADISTICA INFERENCIAL	26
Modelos poblacionales.....	26
Muestreo.	27
Correlación.....	27
Coefficiente de correlación.	27
Regresión lineal.	28
Test chi cuadrado.	28
UNIDAD 4 ESTADISTICA APLICADA A MERCADOS	30
GEOGEBRA:.....	30
SPSS:	32
Power BI:	34
Análisis de datos categóricos.	35
Análisis de dispersión:	36

INDICE DE TABLAS:

Tabla 1: <i>Tabla de datos para ejercicio de análisis descriptivo</i>	14
Tabla 2: <i>Tabla de datos para ejercicio de determinación de medidas de tendencia central</i>	16
Tabla 3: <i>Tabla de datos para ejercicio de cálculo de MTC</i>	17
Tabla 4: <i>Tabla de datos para tabulación de datos agrupados</i>	18
Tabla 5: <i>Tabla de datos desagrupados para encontrar medias estadísticas</i>	19
Tabla 6: <i>Tabla de datos agrupados para realizar tabla de frecuencias.</i>	19
Tabla 7: <i>Tabla de datos agrupados para realizar tabla de frecuencias.</i>	20
Tabla 8: <i>Tabla de datos para ubicación de cuartiles</i>	21
Tabla 9: <i>Tabla de datos para medidas de orden, datos agrupados.</i>	21
Tabla 10: <i>Tabla de datos para obtención de varianza</i>	22

INDICE DE FIGURAS:

Figura 1: <i>Gráfico comparativo de asimetría</i>	24
Figura 2: <i>Gráfico comparativo de curtosis</i>	25
Figura 3: <i>Ejemplo de test chi cuadrado</i>	29
Figura 4: <i>Pantalla con funciones de Geogebra</i>	31
Figura 5: <i>Gráfico de distribución normal-estadística con Geogebra</i>	31
Figura 6: <i>Pantalla ingreso de datos SPSS</i>	33
Figura 7: <i>Pantalla salida de informes SPSS</i>	33
Figura 8: <i>Menú principal superior Power BI.</i>	34
Figura 9: <i>Menú opciones de informe de resultados power BI</i>	35
Figura 10: <i>Análisis inferencial modelo lineal en Geogebra</i>	37

Presentación de la asignatura.

La estadística es un pilar fundamental para el análisis del comportamiento de los consumidores y para la toma de decisiones estratégicas en las empresas a todo nivel desde los pequeños emprendimientos hasta las empresas posicionadas por muchos años. En el campo de la investigación de mercados, su aplicación convierte grandes volúmenes de información en conocimiento ideal para la administración de las empresas, útil y aplicable. La presente guía de estudio, basada en el P.E.A. y el syllabus de la materia de estadística de la carrera de marketing, ha sido construida con el fin de proporcionar un enfoque práctico y accesible a los métodos estadísticos más relevantes utilizados en el análisis de mercados. En lo posterior, se desarrollará actividades teóricas acompañadas con ejemplos reales del mundo empresarial y ejercicios que le permitirán asimilar cada concepto. Se consideran instrumentos descriptivos para el resumen de datos y procedimientos inferenciales para generalizar resultados y evaluar hipótesis de negocios. También se incluyen algunas pautas metodológicas para diseñar estudios, establecer objetivos y elegir muestras. La finalidad de la presente guía es netamente para insertar en el mercado laboral a los estudiantes del Instituto Superior Tecnológico Sucre con una robusta capacidad analítica, habilidad para interpretar información con precisión y tomar decisiones informadas. La presente guía está dirigida para estudiantes, académicos y profesionales consistiendo en consejos prácticos para facilitar un mejor uso de la estadística en la investigación. En un entorno competitivo donde el manejo masivo de los datos cada vez tiene más importancia, aprender estas herramientas será fundamental. Y por ello, aquí no sólo se expone teoría, sino que también plantea una perspectiva práctica en la que los números se relacionan con las realidades de mercado. Con esta herramienta, el estudiante tecnológico tendrá los conocimientos necesarios para abordar los desafíos del análisis de datos de una perspectiva descriptiva e inferencial.

Resultados del aprendizaje.

- El estudiante comprende la terminología básica de estadística.
- El estudiante será capaz de reconocer las medidas estadísticas y las emplea en el desarrollo de ejercicios.
- El estudiante conoce y reconoce correctamente cada uno de los métodos de regresión y correlación y los emplea en ejercicios prácticos.
- El estudiante será capaz de comprender y realizar una prueba de hipótesis
- Realizar análisis de datos estadísticos en programas informáticos para sacar conclusiones y tomar decisiones.

UNIDAD 1: FUNDAMENTOS DE LA ESTADÍSTICA

Definiciones básicas

“La estadística es la ciencia que se encarga de recoger, organizar e interpretar los datos, es la ciencia de los datos” (Gorgas García, Cardiel López , & Zamorano Calvo, 2023)

La Estadística (del latín, Status o ciencia del estado) se ocupaba sobre todo de la descripción de los datos fundamentalmente sociológicos: datos demográficos y económicos (censos de población, producciones agrícolas, riquezas, etc.), principalmente por razones fiscales.

Además, es una disciplina fundamental para la investigación científica y la toma de decisiones, ya que permite recolectar, organizar, analizar e interpretar datos con el fin de describir fenómenos o apoyar conclusiones objetivas.

El objetivo fundamental de la estadística es obtener conclusiones de la investigación empírica usando modelos matemáticos.

Importancia de la estadística

La estadística es una herramienta importante para la investigación científica, facilita la toma de decisiones informadas, permite explicar y describir fenómenos de índole social, económica y natural presentando las bases para áreas como marketing, economía, ingeniería y ciencias sociales.

Investigación científica

La investigación científica es una búsqueda sistemática, controlada, empírica y crítica de fenómenos naturales y que se propone aportar nuevos hechos, datos, relaciones o leyes en cualquier ámbito del conocimiento científico. (Gorgas García, Cardiel López , & Zamorano Calvo, 2023)

La investigación científica “se plantea como finalidad la descripción, explicación y predicción de los fenómenos” (Rojas Soriano, 1996). Es reflexiva, sistemática y en algunos casos, controlada. Se erige como una forma de plantear problemas y soluciones de manera. Es sistemática, al encajar los problemas o las hipótesis en un marco teórico que funciona como encuadre referencial. Es controlada, en tanto que exige la comprobación o verificación del hecho o fenómeno, que se estudia mediante la confrontación empírica. (Rojas Soriano, 1996)

Si hablamos de investigación científica, necesariamente se ha de referir a la utilización de la Estadística, como herramienta de recolección, análisis e interpretación de datos del problema que es objeto de estudio.

La Investigación Científica está encaminada a profundizar el conocimiento de un proceso ya sea teórico, práctico o teórico-práctico, parte del conocimiento científico y lo lleva a la solución de problemas de la sociedad que de una forma u otra no han sido investigados o su investigación se ha conducido en otra dirección.

Conocimiento científico

El conocimiento se refiere al conjunto de experiencias, sensaciones y reflexiones que conllevan al razonamiento y el aprendizaje, es a través de este que podemos reconocer el contexto donde nos desenvolvemos. (Hernández-Sampieri et al., 2014)

El conocimiento científico se caracteriza por presentar de manera lógica y organizada la información acerca de fenómenos demostrables. El conocimiento científico se apoya en teorías, leyes y fundamentos a fin de comprobar el análisis y la validez de la información.

Variables

. En un conjunto de observaciones hay diferentes valores para diferentes rasgos o atributos. Esto es lo que constituye una variable (Garza Olvera, 2014).

De acuerdo con (Anderson, Sweeney, & Williams, 2020), una variable es “cualquier característica medible de una población o muestra”.

Clasificación de variables

- a) Variables cualitativas: Describen cualidades o categorías y no se expresan numéricamente.
 - Ejemplos: género, estado civil, tipo de carrera.
- b) Variables cuantitativas: Se expresan mediante valores numéricos.
 - Discretas: toman valores enteros (número de estudiantes).
 - Continuas: pueden tomar cualquier valor dentro de un intervalo (peso, estatura).
 - Ejemplo

En un estudio sobre estudiantes universitarios:

- Carrera → variable cualitativa
- Edad → variable cuantitativa discreta
- Promedio académico → variable cuantitativa continua

Ejercicio propuesto

1. Identifique el tipo de variable:

- a) Número de materias aprobadas
- b) Color de ojos
- c) Tiempo de estudio diario (horas)

Variable Independiente

La variable independiente se considera la causa o el predictor en la relación que se está estudiando.

Para entender mejor la variable independiente, veamos un ejemplo. Supongamos que un investigador quiere estudiar cómo la cantidad de horas de estudio (variable independiente) afecta el rendimiento académico de los estudiantes (variable dependiente). En este caso, la cantidad de horas de estudio es la variable independiente porque el investigador puede cambiarla al variar el tiempo que los estudiantes dedican al estudio. Luego, el investigador observará si estos cambios en la variable independiente afectan la variable dependiente, es decir, el rendimiento académico de los estudiantes.

Es importante que la variable independiente sea manipulable o controlable en un estudio de investigación. Esto significa que los investigadores deben poder cambiar o ajustar los valores de la variable independiente para observar cómo estos cambios afectan a la variable dependiente. Esto se hace para establecer relaciones de causa y efecto entre las variables y comprender mejor cómo ciertos factores influyen en otros.

Variable dependiente

La variable dependiente es la característica o el resultado que se mide y observa en respuesta a la manipulación o variación de la variable independiente. En la mayoría de los casos, la variable

dependiente se considera el efecto o el resultado que el investigador está interesado en comprender o explicar.

Siguiendo con el ejemplo anterior, si la cantidad de horas de estudio es la variable independiente, el rendimiento académico de los estudiantes (por ejemplo, calificaciones obtenidas) sería la variable dependiente. El investigador observará cómo cambia el rendimiento académico (variable dependiente) a medida que se modifica la cantidad de horas de estudio (variable independiente). En otras palabras, la variable dependiente depende de la variable independiente. Es importante medir y registrar adecuadamente la variable dependiente para obtener resultados confiables en una investigación. Esto implica utilizar métodos de medición válidos y fiables, como pruebas estandarizadas, cuestionarios, observaciones o análisis de datos, según corresponda al tipo de variable dependiente que se esté estudiando. (Hernández-Sampieri et al., 2014)

Variable de control

Una variable de control en el contexto de la investigación es una variable que se mantiene constante o se manipula de manera deliberada para asegurar que no tenga un efecto significativo en la relación entre la variable independiente y la variable dependiente.

En otras palabras, una variable de control es una variable que se mantiene constante o se ajusta para eliminar su influencia potencial sobre los resultados de un estudio. Esto permite a los investigadores aislar y examinar más precisamente el efecto de la variable independiente sobre la variable dependiente. (Hernández-Sampieri et al., 2014)

Por ejemplo, en un estudio sobre el efecto del tiempo de estudio en el rendimiento académico de los estudiantes, la edad y el nivel socioeconómico podrían ser variables de control. Esto significa que se mantienen constantes o se controlan para asegurar que cualquier diferencia en el rendimiento académico pueda ser atribuida principalmente al tiempo de estudio y no a otras variables como la edad o el nivel socioeconómico

Descripción de datos

La descripción de datos consiste en organizar y resumir la información recolectada para facilitar su comprensión.

Levin y Rubin indican que la estadística descriptiva permite “*presentar los datos de forma significativa mediante tablas, gráficos y medidas numéricas*”.

Clasificación de datos

Los datos pueden clasificarse según su origen y naturaleza:

a) Según su origen

- Datos primarios: recolectados directamente.
- Datos secundarios: obtenidos de fuentes existentes (informes, bases de datos).

b) Según su naturaleza

- Cualitativos
- Cuantitativos

Ejemplo

Una encuesta aplicada a estudiantes sobre hábitos de estudio genera datos primarios y puede contener variables cualitativas y cuantitativas.

Ejercicio

Clasifique los siguientes datos:

- Fuente de información: encuesta institucional
- Tipo de dato: promedio de calificaciones

Frecuencias

Las frecuencias permiten resumir cuántas veces aparece un valor o categoría dentro de un conjunto de datos.

Según Triola, las distribuciones de frecuencia facilitan la visualización y análisis de grandes volúmenes de datos.

Frecuencias absolutas, relativas y acumuladas

- Frecuencia absoluta (fa): número de veces que aparece un valor.
- Frecuencia relativa (fr): proporción respecto al total.
- Frecuencia acumulada (Fa): suma progresiva de frecuencias.

Ejemplo

Calificaciones de 10 estudiantes:

6, 7, 7, 8, 8, 8, 9, 9, 10, 10

Tabla 1:

Tabla de datos para ejercicio de análisis descriptivo

Nota	fa	fr	Fa
6	1	0,10	1
7	2	0,20	3
8	3	0,30	6
9	2	0,20	8
10	2	0,20	10

Nota: tabla de datos de calificaciones de 10 estudiantes para tabulación en tabla de frecuencias

Ejercicio práctico

Construya una tabla de frecuencias con los datos:

5, 6, 6, 7, 7, 7, 8, 9

Distribución Normal

La distribución normal es un modelo probabilístico continuo con forma de campana, ampliamente utilizado en estadística.

Según Walpole et al., la distribución normal describe fenómenos naturales y sociales donde los datos se concentran alrededor de un valor central.

Distribución continua, media, desviación y probabilidad

- Media (μ): valor central de la distribución.
- Desviación estándar (σ): mide la dispersión de los datos.

- Probabilidad: área bajo la curva normal.

Regla empírica

- 68% de los datos están dentro de $\pm 1\sigma$
- 95% dentro de $\pm 2\sigma$
- 99,7% dentro de $\pm 3\sigma$

Ejercicio

Si el promedio de una prueba es 70 con $\sigma = 5$:

- ¿Entre qué valores se encuentra el 68% de los estudiantes?

Práctica 1: Elaboración de tablas de frecuencias

Actividad guiada

1. Recolecte las edades de 20 estudiantes.
2. Organice los datos en una tabla.
3. Calcule frecuencias absolutas y relativas.
4. Analice la concentración de los datos.

UNIDAD 2: MEDIDAS ESTADÍSTICAS

Medidas de tendencia central

Media aritmética:

Es un promedio básico de un conjunto de datos tabulados de manera ordenada o desordenada; y básicamente consiste en sumar todos los valores de las observaciones expresadas como datos cuantitativos y dividirlos entre el número de estas.

La media se obtiene de aplicar la fórmula:

$$\bar{x} = \frac{\sum_{i=1}^N x_i n_i}{N}$$

Ejemplo en marketing: Una empresa va a estudiar el gasto mensual que 10 consumidores hacen de snacks. Se tiene los siguientes datos en dólares:

Tabla 2:

Tabla de datos para ejercicio de determinación de medidas de tendencia central

16	7,5	18	14,6	7,4	9,5	10	11	11	15
----	-----	----	------	-----	-----	----	----	----	----

Nota: datos para determinar y calcular medidas de tendencia central de datos desagrupados

Si la suma es \$120, el consumidor promedio gasta \$12 por mes, según la media aritmética, estimación del tamaño del mercado.

Media geométrica

Definición: Es la raíz n-ésima del producto de n números. Se aplica cuando los datos son tasas de crecimiento o tienen cambios porcentuales.

Ejemplo 1: en marketing: Una firma experimenta crecimientos en sus ventas anuales: 10%, 20% y 30%. La media geométrica permite aproximar el promedio de crecimiento real sin que el mercado se vea distorsionado por cambios extremos .

Ejemplo 2: para calcular la media geométrica en el ejemplo de la media aritmética, sería:

$$\sqrt[10]{16 \times 7,5 \times 18 \times 14,6 \times 7,4 \times 9,5 \times 10 \times 11 \times 11 \times 15} = 11,4938$$

Media armónica

Definición: Consiste en dividir el número de observaciones entre la suma de los inversos de los valores. Es útil cuando se tiene datos que son tasas o proporciones, en particular promedios de velocidad o rendimiento.

Ejemplo en marketing: Si se desea saber el precio promedio al que pagan los clientes por unidad y algunos compran paquetes con diferentes precios unitarios, la media armónica ofrece un promedio más cercano del “precio por bolígrafo” real.

Ejemplo 2: si queremos averiguar la media armónica de los 10 datos de la media aritmética, la media armónica será:

$$\frac{12}{\frac{1}{16} + \frac{1}{7,5} + \frac{1}{18} + \frac{1}{14,6} + \frac{1}{7,4} + \frac{1}{9,5} + \frac{1}{10} + \frac{1}{11} + \frac{1}{11} + \frac{1}{15}} = 13,2047$$

Mediana:

Definición La mediana es el número o cantidad del medio de un conjunto de datos ordenados de menor a mayor, en caso de tener dos cantidades en la mitad, se requiere sacar el promedio de dichos 2 datos.

Ejemplo en marketing: En una encuesta de periodicidad de compra, si los datos ordenados son 1, 2, 2, 3, 5, 7, 20 compras/mes, la mediana es 3, que refleja el comportamiento típico del consumidor cuando hay alguien que compra demasiado (que podría distorsionar la media).

Ejemplo 2: tomando en cuenta el ejemplo inicial, se ordena y se distingue la mediana como promedio de los 2 datos de la mitad de la siguiente manera:

Tabla 3:

Tabla de datos para ejercicio de cálculo de MTC

7,4	7,5	9,5	10	11		11	14,6	15	16	18
-----	-----	-----	----	----	--	----	------	----	----	----

Nota: Para determinar medidas de orden hay que tener ordenados los datos ascendente o descendentemente.

El promedio entre los dos 11, es 11, por tanto, la mediana es 11.

5. Moda Definición: Es el valor que más se repite en un conjunto de datos. Ejemplo en marketing: En una encuesta de sabor del yogur, si la opción “fresa” obtiene más votaciones, se trata de la moda, con

lo que indica la opción más popular entre los consumidores, ya sea para decisiones de producción o de publicidad.

Ejemplo 2: en los datos iniciales, la moda es 11, debido a que es el dato que más se repite, o sea tiene la frecuencia absoluta más alta.

Ejercicio de aplicación:

Determinar los valores de a, b, c, d, e.

Tabla 4:

Tabla de datos para tabulación de datos agrupados

Datos	Mediana	Moda	Media
Altura (m): 1,52; 1,74; 1,83; 1,52; 1,67; 1,91	a	b	1,70
Edad (años): 21, 34, 17,22, 56, 38	28	No hay	c
Peso (kg): 54,7; 48,6; 63,2; 55,1; 77,9; 48,6	d	48,6	e

Nota: Mediante esta tabla podemos calcular medidas de tendencia central de variables cualitativas.

Ejercicio para resolver:

Los pesos de ocho calabazas son:

26,3kg; 12,6kg; 33,5kg; 8,9kg; 18,7kg; 22,6kg; 31,8kg y 45,3kg

a. Halle la mediana de los pesos.

b. Calcule la media de los pesos.

Para datos agrupados, una vez determinados, los límites inferiores, superior, la marca de clase y la amplitud de cada una de las clases o intervalos de datos, podemos calcular las medidas de tendencia central:

Media: $\bar{x} = \frac{\sum_{i=1}^N x_i f_i}{N}$ siendo f_i la frecuencia absoluta de cada intervalo

Moda: $Mo = a_j + \frac{\delta_1}{\delta_1 + \delta_2} \cdot (a_{j+1} - a_j)$

Determinar el valor de la media aritmética, geométrica, armónica y la moda de los siguientes datos desagrupados:

Tabla 5:

Tabla de datos desagrupados para encontrar medias estadísticas

11	5	5	9	7	12	4	6	10	8
----	---	---	---	---	----	---	---	----	---

Nota: Mediante el uso de fórmulas podemos determinar los distintos tipos de medias de datos estadísticos

Determinar el valor de la media aritmética, y la moda de los siguientes datos agrupados:

Tabla 6:

Tabla de datos agrupados para realizar tabla de frecuencias.

Intervalo	Límite Inferior	Límite Superior	Frecuencia absoluta
Clase A	10	12,4	2
Clase B	12,4	14,8	4
Clase C	14,8	17,2	7
Clase D	17,2	19,6	3

Nota: Tabla de variable cualitativa de datos agrupados según conglomerados.

Medidas de posición.

Las medidas de posición son medidas estadísticas que sirven para situar o localizar un elemento dentro de un conjunto de datos ordenado. Sirven para encontrar valores que dividen a la distribución en partes iguales y nos indican cómo están distribuidos los datos alrededor de ciertos puntos de interés. Algunas de las medidas más importantes son la mediana, los cuartiles, deciles y percentiles, que son útiles para describir el grado de dispersión y concentración de estos. Estas métricas hacen que sea más sencillo comparar individuos o grupos dentro de un mercado dada su prevalencia en el análisis de consumidores. También pueden ser utilizadas para segmentar, encontrando patrones y diferencias entre niveles o dentro de las categorías. En investigación de mercados, orientan las decisiones relacionadas con la conducta, consumo y perfil de los clientes.

Cuartiles

Definición: Los cuartiles dividen una muestra ordenada en cuatro partes iguales y se obtienen por los valores en los que se separa el 25%, 50% y 75% de la distribución (Q1, Q2, Q3). Aplicación en la investigación de mercados: Una empresa está estudiando la cantidad que gastan al mes sus 100

clientes. Si $Q_1 = \$ 20$, significa que el 25% de los clientes gasta \$ 20 o menos, lo que ayuda a planificar promociones dirigidas a consumidores de bajo gasto.

Ejemplo de obtención de cuartiles: Se tiene los siguientes datos ordenados de menor a mayor:

Tabla 7:

Tabla de datos agrupados para realizar tabla de frecuencias.

11,5	13	14,5	16	16,7	18,2	19,6	20,5	21,6	24,7	26
------	----	------	----	------	------	------	------	------	------	----

Nota: Cuando se levanta datos como en la presente tabla hay que ordenar los datos para analizar su tendencia central.

Se tendrá que el cuartil 1 es: 15,25 ; el cuartil 2 también conocido como mediana es: 18,2; y el cuartil 3 es: 21,05

Los deciles:

Consiste en tener un conjunto ordenado de datos en diez partes iguales diferenciándose así por ciertos valores que marcan el 10%, 20%, 30% ... 90% de la distribución (D_1, D_2, \dots, D_9). Ejemplo en marketing podemos nombrar al estudio de frecuencia de compra, si $D_7 = 8$ compras cada mes, indica que el 70% de los compradores realiza 8 o menos compras durante el mes, entonces con ello podemos segmentar mercados de compradores frecuentes y poco frecuentes.

Percentiles

Definición: Son valores que dividen a los datos ordenados en 100 partes iguales y representa el valor por debajo del cual se sitúa un cierto porcentaje de las observaciones (P_1, P_2, \dots, P_{99}). Para dar un ejemplo en el campo del marketing, al realizar una investigación se puede tener en un sondeo de satisfacción al cliente, si $P_{90} = 4.8$ de 5, indica que el 90% de los clientes declaró tener una satisfacción igual o menor a 4.8, esta información nos sirve para establecer el nivel de calidad en el campo del benchmarking.

Para datos agrupados, una vez determinados, los límites inferiores, superior, la marca de clase y la amplitud de cada una de las clases o intervalos de datos, podemos calcular las medidas de posición:

$$\text{Cuartil 1-}Q_1: Q_{1/4} = a_j + \frac{\frac{N}{4} - N_{j-1}}{n_j} \cdot (a_{j+1} - a_j)$$

$$\text{Cuartil 2-}Q_2\text{- (mediana): } Q_{1/4} = a_j + \frac{\frac{N}{2} - N_{j-1}}{n_j} \cdot (a_{j+1} - a_j)$$

$$\text{Cuartil 3 - Q3: } Q_{1/4} = a_j + \frac{\frac{3N}{4} - N_{j-1}}{n_j} \cdot (a_{j+1} - a_j)$$

Ejercicios de aplicación de conocimientos:

Determinar el valor de los cuartiles de los siguientes datos desagrupados:

Tabla 8:

Tabla de datos para ubicación de cuartiles

11	3	5	9	7	12	4	6	10	8
----	---	---	---	---	----	---	---	----	---

Nota: La presente tabla me daría un indicativo al determinar la caja y los bigotes, la concentración de datos en la cuarta parte del conjunto total de los mismos.

Determinar el valor de los cuartiles de los siguientes datos agrupados:

Tabla 9:

Tabla de datos para medidas de orden, datos agrupados.

Intervalo	Límite Inferior	Límite Superior	Frecuencia absoluta
Clase A	10	12,4	2
Clase B	12,4	14,8	4
Clase C	14,8	17,2	7
Clase D	17,2	19,6	3

Nota: la ubicación de cuartiles de datos agrupados requiere el uso de fórmulas donde el límite inferior de la clase tiene mucha importancia.

Medidas de dispersión:

Varianza

Concepto: La varianza nos sirve para saber cuánto se alejan o dispersan los datos con respecto a la media aritmética. Es la media de las diferencias al cuadrado entre cada punto y la media. Ejemplo investigación declaraciones de mercado: Una empresa estudia la satisfacción de 10 clientes (escala 1–10). Una varianza elevada implica que las opiniones están muy diseminadas y se traduce en un mercado muy heterogéneo en el que algunos clientes están muy satisfechos y otros muy insatisfechos. Menor varianza también significa más uniformidad en la experiencia del cliente.

Ejemplo para calcular la desviación típica o estándar:

Tabla 10:

Tabla de datos para obtención de varianza

Datos (x)	Frecuencia (f)	X por f	X - media	(x- media) ²	(x- media) ² por f
5	3	15	-3,46	11,9716	35,9148
6,2	2	12,4	-2,26	5,1076	10,2152
7	5	35	-1,46	2,1316	10,658
7,5	6	45	-0,96	0,9216	5,5296
8	4	32	-0,46	0,2116	0,8464
9	7	63	0,54	0,2916	2,0412
9,5	5	47,5	1,04	1,0816	5,408
10	3	30	1,54	2,3716	7,1148
11,5	3	34,5	3,04	9,2416	27,7248
12	2	24	3,54	12,5316	25,0632
	40	338,4			Var= 3,629
		Media= 8,46			Desv.E.= 1,806

Nota: La obtención de la varianza requiere el cálculo exacto de la media, le diferencia y su cuadrado nos llevan a la varianza y posteriormente la desviación estándar.

Covarianza

Concepto: La covarianza es una medida de la relación lineal entre dos variables. • Si está positiva las dos variables tienden a incrementarse juntas. • Negativo es cuando una sube la otra baja, es decir, es cuando hay una covarianza negativa. • Cuando se aproxima a cero, hay una nula o débil relación lineal. Ejemplo investigación declaraciones de mercado: Una compañía estudia el vínculo entre el ingreso del consumidor y la compra de productos premium. Una covarianza positiva representa eso para mayores ingresos y mayores gastos en parte para la segmentación y la estrategia de precios.

Desviación estándar

Definición: Es la raíz cuadrada de la varianza. Nos dice lo alejados que están en promedio los datos de la media, pero está en las mismas unidades de la variable, con lo cual interpretarlo es más fácil.

Ejemplo investigación en declaraciones de markers: Si el gasto por cliente al mes en café es 15 y la

desviación estándar es 5, eso quiere decir que la mayoría de los clientes se encuentra ± 5 del promedio.

Esto ayuda a estimar la variabilidad del consumo y planificar la demanda

Medidas de forma:

Definición:

Las medidas de forma nos ayudan a entender cómo se configura o se ve una distribución de datos, revelando si es más “alta”, “plana”, simétrica o si tiene asimetrías. Las más importantes son la asimetría (skewness) y la curtosis (kurtosis).

Ejemplo aplicado a marketing:

Al observar la frecuencia de compra mensual, las medidas de forma nos permiten identificar si la mayoría de los clientes compra de manera similar o si hay grupos extremos, como aquellos que compran con mucha frecuencia o muy de vez en cuando.

Distribución unimodal

Definición:

Una distribución unimodal es aquella que tiene una sola moda, es decir, un único valor que se repite con mayor frecuencia.

Ejemplo en investigación de mercados:

En una encuesta sobre preferencias de sabores de helados, si “chocolate” es la opción más popular y ninguna otra se le acerca, estamos ante una distribución unimodal, centrada en un único sabor predominante.

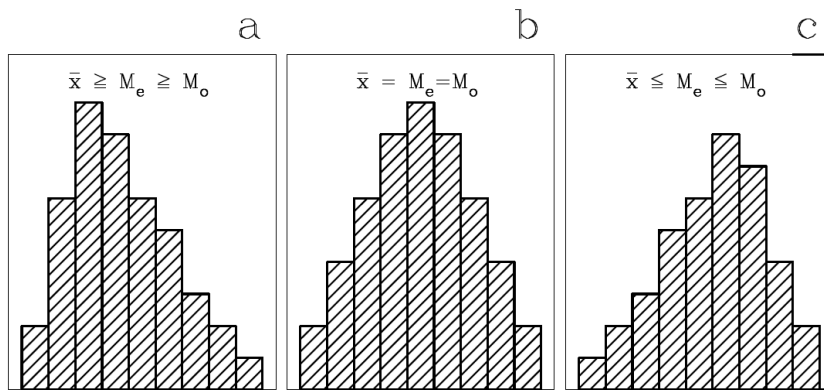
Asimetría (skewness)

Definición:

La asimetría nos indica si la distribución de los datos está inclinada hacia la izquierda o hacia la derecha.

- Asimetría positiva: cola hacia la derecha (pocos valores muy altos).
- Asimetría negativa: cola hacia la izquierda (pocos valores muy bajos).

Figura 1:
Gráfico comparativo de asimetría



Nota: Distribución con asimetría hacia la derecha, positiva, (panel a), simétrica (panel b) y con asimetría hacia la izquierda, negativa (panel c).

Ejemplo aplicado a marketing:

En un estudio sobre el gasto mensual en restaurantes, si la mayoría gasta entre 10 y 30 dólares, pero unos pocos gastan más de 150, la distribución muestra asimetría positiva, evidenciando un pequeño segmento de alto gasto.

Curtosis (kurtosis)

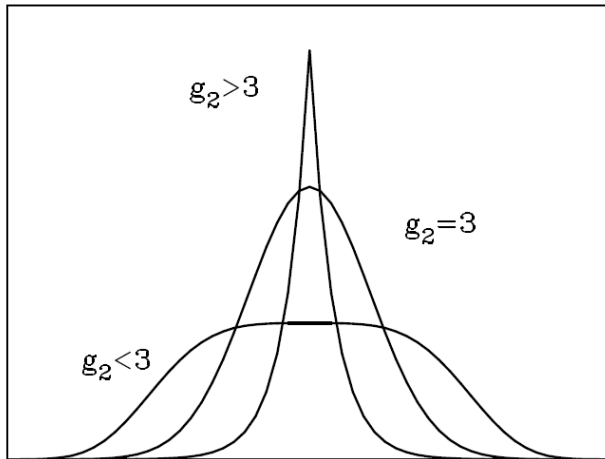
Definición:

La curtosis mide cuán concentrados están los datos alrededor de la media y qué tan “altos” o “planos” son los picos de una distribución.

- Leptocúrtica: pico muy alto (datos muy concentrados).
- Platicúrtica: pico bajo (datos muy dispersos)

Mesocúrtica: distribución normal.'

Figura 2:
Gráfico comparativo de curtosis



Nota: Distribuciones con diferente grado de apuntamiento: leptocúrtica ($g_2 > 3$), mesocúrtica ($g_2 = 3$) y platicúrtica ($g_2 < 3$).

Para calcular el coeficiente de curtosis tenemos las siguientes fórmulas:

$$g_2 = \frac{m_4}{s^4} \quad \text{donde} \quad m_4 = \frac{\sum_{i=1}^k (x_i - \bar{x})^4 n_i}{N}$$

Ejemplo aplicado a marketing:

En un estudio de satisfacción del cliente, si la mayoría responde entre 4.5 y 5 en una escala de 1 a 5, la distribución es leptocúrtica, mostrando mucha concentración en valores positivos y una experiencia homogénea.

UNIDAD 3 ESTADISTICA INFERENCIAL

Es la rama de la estadística que utiliza los datos obtenidos de muestra para obtener conclusiones, hacer estimaciones y tomar decisiones sobre una población, apoyándose en la probabilidad para medir la incertidumbre de los resultados obtenidos.

Sus principales objetivos son:

Estimar parámetros poblacionales

Probar hipótesis

Tomar decisiones bajo incertidumbre

La estadística inferencial se apoya en conceptos como el error muestral, la distribución de probabilidad y el nivel de confianza.

Modelos poblacionales.

Son representaciones matemáticas o estadísticas que describen y analizan el comportamiento de una población, la cual permite estudiar su tamaño, crecimiento, distribución y cambios a lo largo del tiempo, con el fin de hacer predicciones y apoyar la toma de decisiones.

Ya sea X una variable aleatoria que representa una característica de la población. Su comportamiento se describe mediante una función de probabilidad o función de densidad, según el tipo de variable.

Ejemplo de distribución normal:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Donde:

μ es la media poblacional

σ es la desviación estándar poblacional

Muestreo.

Es el procedimiento mediante el cual se va seleccionando una parte que representa una población, llamada muestra, con el propósito de analizarla y obtener conclusiones validas sobre el total de la población sin estudiar a todos sus integrantes.

Media muestral

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_{i1}$$

Varianza muestral

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{i1} - \bar{x})^2$$

Error estándar de la media

$$SE = \frac{s}{\sqrt{n}}$$

El error estándar mide la variabilidad de la media muestral como estimador de la media poblacional.

Correlación.

La correlación es una forma sencilla de explicar que tan relacionados están dos puntos, indica si cuando uno cambia, el otro también lo hace, ya que sea en el mismo sentido, en sentido contrario o sin una relación clara. No dice que uno afecte al otro, solamente que tan juntos se mueven.

Coefficiente de correlación.

Es un número que indica que tan fuerte y cuál es la dirección de dos variables: si el número es positivo ambas aumentan juntas, si el número es negativo, una sube mientras la otra baja y si es cercano a 0, la relación es débil o inexistente.

$$r = \frac{\Sigma(x_{i1} - \bar{x})(y_{i1} - \bar{y})}{\sqrt{\Sigma(x_{i1} - \bar{x})^2 \Sigma(y_{i1} - \bar{y})^2}}$$

Donde:

\bar{x} y \bar{y} son las medias muestrales

x_{i1} y y_{i1} son los valores observados

Los valores de r cumplen:

$$-1 \leq r \leq 1$$

Regresión lineal.

Es un método estadístico que permite analizar y predecir la relación entre dos variables mediante una recta, mostrando como cambia una variable dependiente Y con respecto a la variable independiente X .

Modelo matemático

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

En la estimación muestral se expresa como:

$$y = mx + b$$

Donde:

m es la pendiente

b es la intersección con el eje y

Test chi cuadrado.

Podríamos estar interesados en averiguar si ciertos conjuntos de datos son independientes.

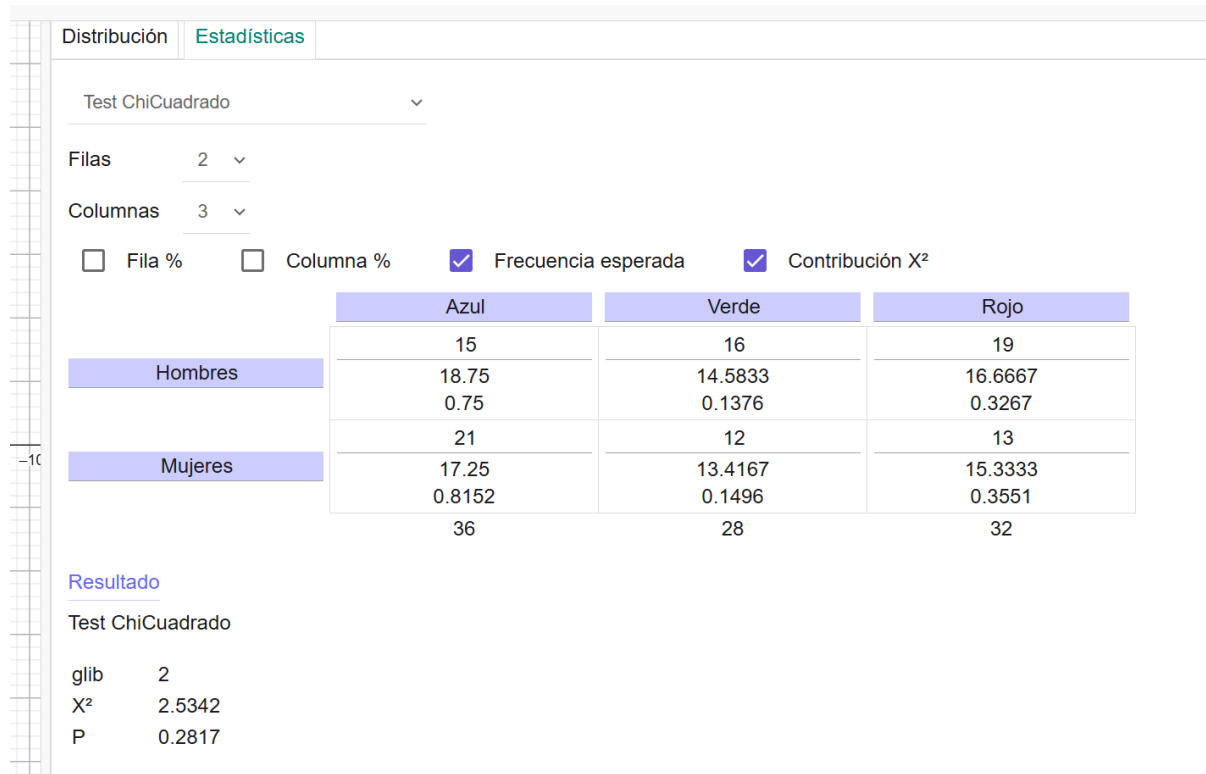
Supongamos que recopilamos datos acerca del color favorito de camiseta para hombres y mujeres.

Podríamos querer averiguar si el color y el sexo son independientes o no. Una forma de hacerlo es

llevando a cabo una prueba de chi cuadrado χ^2

Figura 3:

Ejemplo de test chi cuadrado



Nota: Test chi cuadrado de el nivel de independencia entre a variable sexo del comprador y color de camiseta ofrecido por un local comercial.

UNIDAD 4 ESTADISTICA APLICADA A MERCADOS

GEOGEBRA:

GeoGebra es un software libre de matemática que puede aplicarse en geometría, álgebra y cálculo.

Como sistema de geometría dinámica, permite construir figuras con puntos, vectores, segmentos, rectas y cónicas, entre otras, y también gráficas de funciones. Mediante la explicitación de fórmulas y coordenadas, ofrece un repertorio de comandos propios del análisis matemático para hallar derivadas e integrales de funciones, e identificar puntos singulares como raíces o extremos, por ejemplo. Este software se puede descargar en forma gratuita de www.geogebra.org.

En la pantalla inicial de GeoGebra encontramos una vista gráfica, una numérica, llamada vista algebraica, y, además, una vista de hoja de cálculo.

A continuación, algunos elementos de su interfaz básica:

Barra de menú: Contiene diferentes menús desplegables que facilitan el trabajo con archivos y determinan la configuración del programa. Los menús corresponden a: Archivo, Edita, Vista, Opciones, Herramientas, Ventana y Ayuda.

Barra de herramientas: Contiene distintas opciones para realizar construcciones geométricas, información de la herramienta seleccionada, y los botones para deshacer y rehacer las acciones realizadas.

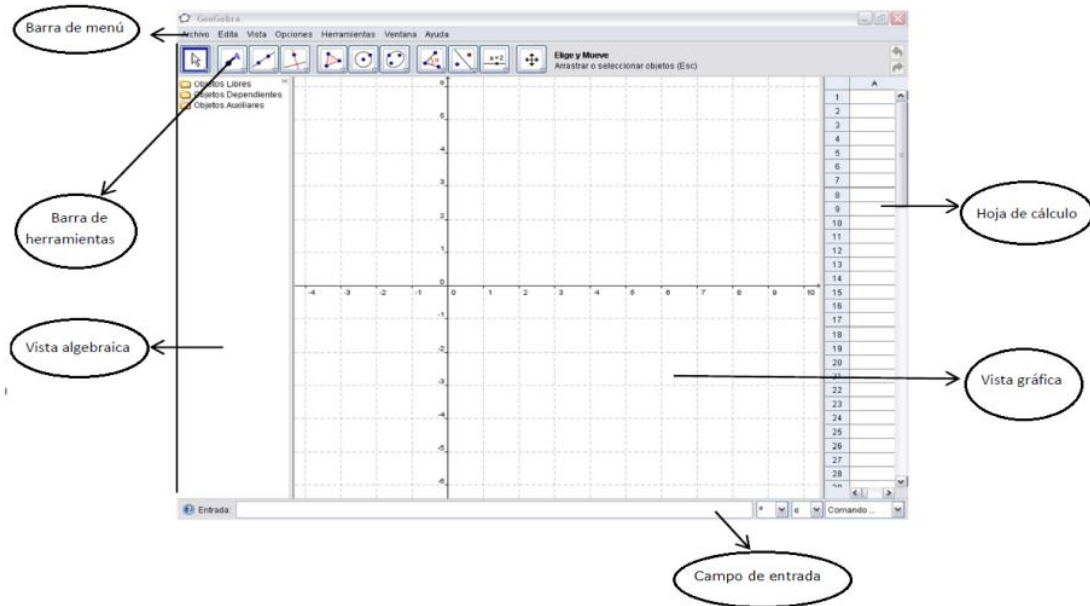
Ventana algebraica: Ofrece la información del proceso realizado, indicando los objetos libres, dependientes y los auxiliares que también se podrán mostrar.

Vista gráfica: Es la zona principal de GeoGebra, donde se ven y manipulan los gráficos.

Hoja de cálculo: Ofrece funciones similares a las de Microsoft Excel.

Campo de entrada: Permite introducir expresiones, además de las opciones para seleccionar distintas funciones, caracteres o comandos

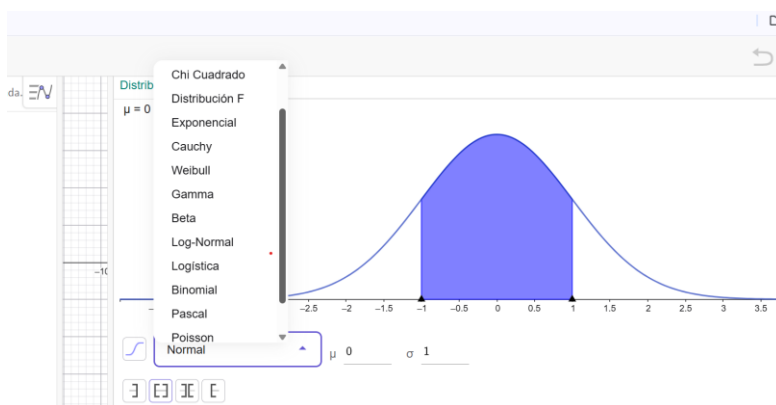
Figura 4:
Pantalla con funciones de Geogebra



Nota: la figura indica la pantalla primaria donde se puede de terminar valores de modelos lineales y polinómicos de una nube de puntos, e inclusive calcular la correlación.

Una de las herramientas más usadas es el de estadística y probabilidad, como un ejemplo el análisis de números triangulares es de la forma $(n(n+1))/2$, para $n = 1,2...$ Estudiando los errores que se producen al medir reiteradas veces una misma magnitud, Gauss probó que éstos se distribuyen según una ecuación exponencial cuya gráfica es:

Figura 5:
Gráfico de distribución normal-estadística con Geogebra



Nota: Ejemplo de cálculo de probabilidad en una distribución normal, de manera visual en la campana de Gauss.

O en test de hipótesis como el test chi cuadrado: Véase ilustración 3.

Geogebra puede usarse en estadística y probabilidad en línea, o descargarse la aplicación de manera gratuita, para cualquier dispositivo audiovisual.

SPSS:

SPSS es uno de los Software más conocidos y utilizados para la realización de investigaciones de mercado e investigaciones relacionadas al área de la sociología y psicología. Una de las características fundamentales de SPSS es su facilidad de uso, junto a la potencia e integridad del software, convirtiéndolo en una de las herramientas más potentes para este tipo de trabajos.

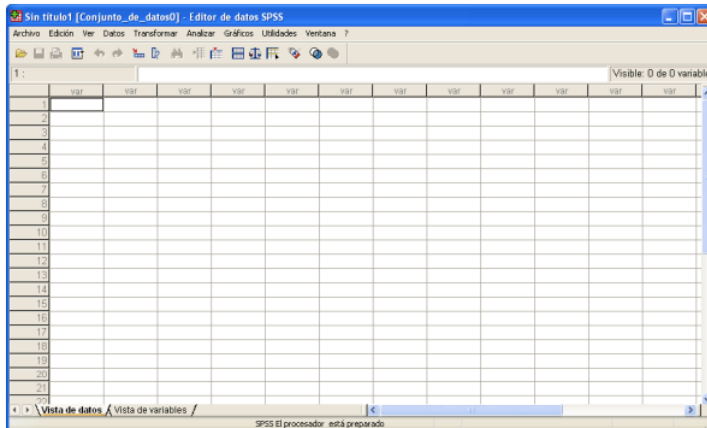
Al iniciar el Software, lo primero que se despliega es una ventana de información (que se observa a la izquierda) la que indica la versión del software, y además muestra la información referente al propietario de la licencia.

Vista de editor de datos:

SPSS se compone de dos ventanas de trabajo. La primera recibe el nombre de editor de datos (arriba) y es la ventana donde se ingresan los datos con los que se trabajará. Para los familiarizados con excel, esta ventana les resultará familiar y su forma de trabajo será bastante similar. En las columnas (hacia la izquierda) se introducirán las distintas variables que compongan la base de datos. En términos prácticos, cada pregunta de un cuestionario realizado corresponde a una variable. En las filas (hacia abajo) son introducidos los datos, o las respuestas entregadas por los encuestados, o donde sea que se hayan extraído los datos para la realización de la base de datos. Así entonces, podemos observar si se entrevistó a 182 personas, hacia abajo deberá haber 182 registros. Este Editor posee dos tipos de vista: “Vista de datos” y “Vista de variables”. En la primera, el investigador observará los datos ingresados, mientras que en la segunda se puede acceder a los atributos de las distintas variables para poder editarlos

Figura 6:

Pantalla ingreso de datos SPSS



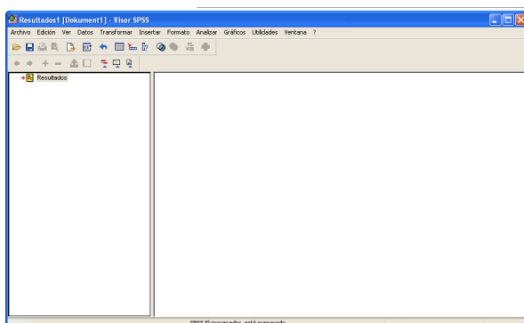
Nota: es una imagen de la pantalla de ingreso de valores y variables que serán analizadas por SPSS.

Visor de resultados:

La segunda ventana correspondiente al área de trabajo de SPSS se conoce como “Visor de Resultados” y, tal como su nombre lo indica, corresponde a una ventana donde el software despliega los resultados que se obtienen de los distintos análisis realizados. En la sección izquierda, el software mostrará los títulos de los distintos análisis y los desplegará como un diagrama de árbol. Por otra parte, en la sección derecha se muestran los resultados obtenidos, los cuales pueden ser copiados y pegados a software como Microsoft Word ® o Excel ® para realizar informes u otras acciones que el investigador considere necesarias.

Figura 7:

Pantalla salida de informes SPSS



Nota: esta es la pantalla donde se expondrán los resultados de cruce entre variables y datos para determinar medidas estadísticas.

Power BI:

Power BI es una herramienta de análisis de negocios que convierte datos de diferentes fuentes en informes interactivos y atractivos. Con ella, puedes visualizar tus datos de múltiples maneras, y lo mejor.

Una vez instalado Power BI Desktop en tu computadora, es fundamental familiarizarte con su interfaz. Power BI presenta una interfaz intuitiva y amigable, pero comprender sus componentes clave desde el principio te ayudará a navegar y utilizar la herramienta de manera más efectiva. Vamos a desglosar los elementos principales de la interfaz.

Si observamos el Ribbon o cinta de opciones: detalla el menú principal y las herramientas con las cuales se pueden emitir informes de acuerdo a los datos ingresados como matriz.

Figura 8:

Menú principal superior Power BI.

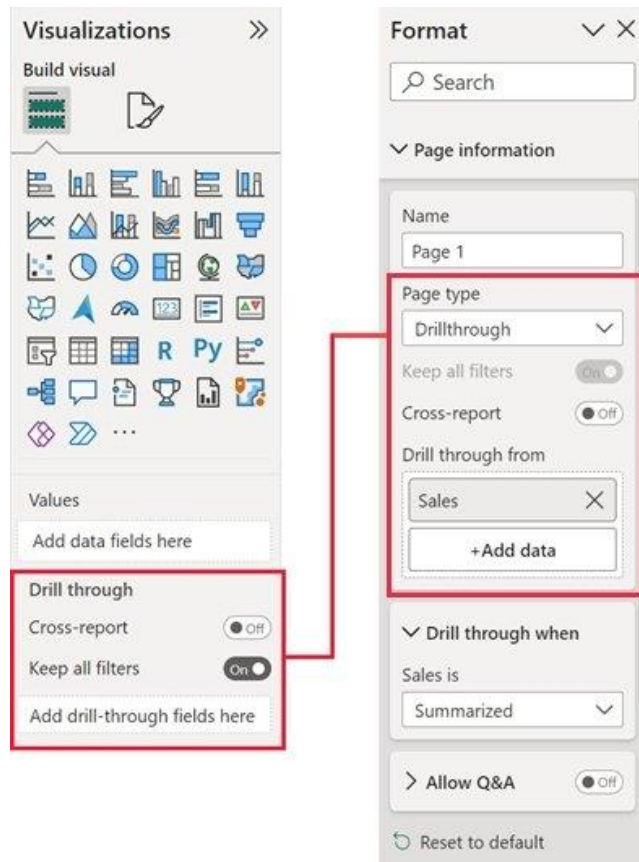


Nota: este es el menú de ingreso de datos, formatos, texto, y variables de análisis en PBI.

El panel de visualizaciones nos indica las opciones para emitir informes solamente con la información más relevante.

Figura 9:

Menú opciones de informe de resultados power BI



Nota: Utilizando el modelo canvas, podemos decir que el ‘Canvas de Reporte’ es el área central de la interfaz, donde se creará y organizará las visualizaciones.

Debemos imaginarnos como un lienzo en blanco donde se puede arrastrar gráficos, tablas y otros elementos visuales para construir el reporte. Aquí es donde el análisis de datos cobra vida visualmente.

Crear el primer reporte en Power BI es muy interesante. Este proceso se compone de varias etapas clave, desde la importación de datos hasta la creación de visualizaciones dinámicas.

Análisis de datos categóricos.

Analizar datos categóricos en Power BI implica transformar textos no estructurados en insights visuales mediante técnicas como gráficos de barras, columnas de datos, treemaps y segmentadores.

Para mejores resultados se puede incluir un proceso de revisión de la calidad de los datos en Query

Editor, agrupar categorías similares, y usar herramientas de análisis automático como "explicar el aumento" para encontrar factores influyentes.

El proceso de análisis conlleva primero, por tanto, asegurar la consistencia. Activa en la vista "Ver" las opciones de "Calidad de columnas" y "Distribución de columnas" para identificar valores nulos, vacíos o errores en tus categorías, previamente en Power Query.

La presentación de informes con datos categóricos, en power BI, incluyen mapas de árbol, gráficos de cinta, informes cronológicos, etc.

Para análisis avanzado se puede realizar el siguiente proceso:

Funcionalidad "Analizar": hacer click derecho en un gráfico de barras o líneas y selecciona "Analizar > Explicar el aumento/disminución". Power BI usará algoritmos para mostrar qué categorías influyen más en el cambio.

Influenciadores Clave: Utiliza el objeto visual de "Influenciadores clave" para identificar qué categorías tienen la mayor probabilidad de influir en una métrica específica, cambiando el tipo de análisis a categórico en el panel de formato.

Interactividad: se puede usar segmentadores (Slicers) para filtrar el informe por categorías específicas (ej. filtrar todo el reporte por "Categoría de Producto").

Análisis de dispersión:

La dispersión de datos es mucho más fácil y visual en GeoGebra, se visualiza principalmente mediante el diagrama de dispersión (nube de puntos) usando la herramienta "Análisis de regresión de dos variables". Permite ingresar datos en una hoja de cálculo, graficarlos, y calcular estadísticos como la correlación, desviación estándar y la línea de mejor ajuste.

Pasos para realizar la dispersión de datos en GeoGebra

Activar la Hoja de Cálculo: En el menú, seleccione Vista y luego Hoja de cálculo.

Ingresar los datos: introduzca los valores de la variable independiente en la columna A y los de la dependiente en la columna B.

Seleccionar los datos: seleccione con el mouse las celdas que contienen los pares de datos.

Crear el diagrama: haga clic en el botón análisis de regresión de dos variables (ícono de histograma sobre puntos) y luego en analiza.

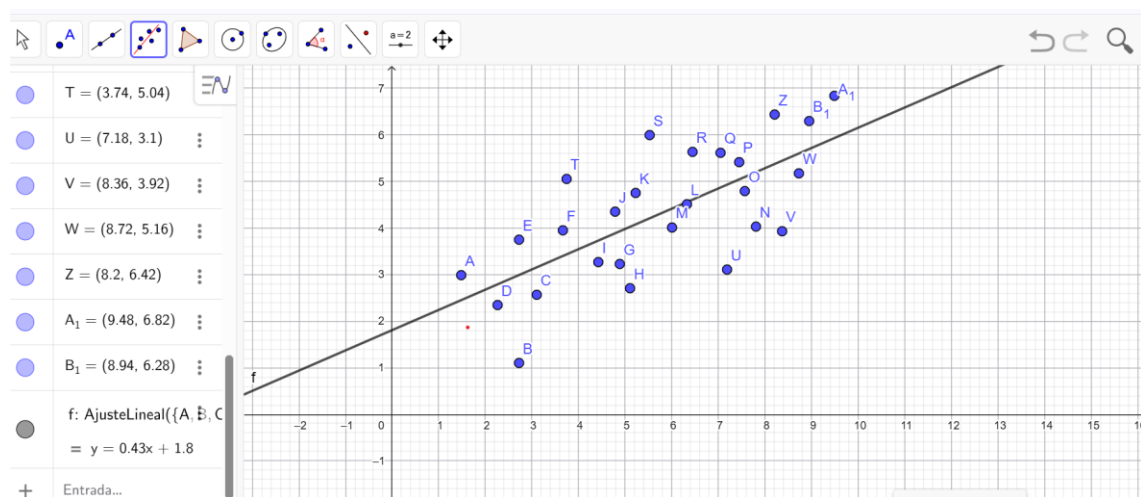
Visualizar y ajustar: GeoGebra mostrará automáticamente el diagrama de dispersión, en el menú desplegable de "modelo de regresión", puede elegir lineal, polinómico, etc.

La herramienta muestra estadísticos como el coeficiente de correlación r.

Opciones de visualización: Puede activar la casilla "mostrar línea" para ver la línea de mejor ajuste o visualizar el diagrama de residuos.

Figura 10:

Análisis inferencial modelo lineal en Geogebra



Nota: Gráfico de modelo lineal generado con datos categóricos expresados como nube de puntos y determinado el modelo lineal.

BIBLIOGRAFÍA

- Anderson, D., Sweeney, D., & Williams, T. (2020). *Statistics for Business and Economics*. Cengage Learning.
- Garza Olvera, B. (2014). *Estadística y probabilidad*. México: Pearson.
- Gorgas García, J., Cardiel López, N., & Zamorano Calvo, J. (2023). *Estadística Básica para estudiantes de Ciencias*. Madrid: Universidad Complutense de Madrid.
- Rojas Soriano, R. (1996). *El proceso de la investigación científica*. México: Trillas.
- Malhotra, N. K. (2008). *Investigación de mercados* (Ed.). Pearson Educación.
- Spiegel, M. R., Schiller, J. J., & Srinivasan, R. A. (2011). *Probabilidad y estadística* (4.^a ed.). McGraw-Hill Educación.
- Triola, M. F. (2018). *Estadística* (12.^a ed.). Pearson Educación.

SUCRE



ISBN: 978-9942-590-04-6



SUCREInstitutooficial



@SUCREInstituto



@SUCREInstituto