

CLASE

1

ESTADÍSTICA II

Repaso aplicado de Estadística I

Técnicas de muestreo

Educación de calidad

INTRO DUCCIÓN DE LA CLASE

GRADO / POSGRADO / TECNOLOGÍAS

Estudia a tu tiempo
sin interrupciones

EDUCACIÓN EN LÍNEA DE CALIDAD



En esta primera semana se realizará un repaso aplicado de los fundamentos de Estadística I, abordando los conceptos esenciales que permitirán a los estudiantes interpretar y analizar información de manera rigurosa dentro de contextos empresariales y administrativos, profundizando en las medidas descriptivas y de dispersión como herramientas para organizar, resumir y comprender grandes volúmenes de datos, así como en las distribuciones de probabilidad, especialmente la normal y el cálculo del valor Z, enfatizando la interpretación de los resultados en situaciones prácticas y la aplicación de estos conocimientos para la toma de decisiones fundamentadas en evidencia, desarrollando competencias analíticas que son indispensables para la gestión eficiente de información en entornos profesionales.

De manera complementaria, se explorarán las técnicas de muestreo aplicadas a la investigación administrativa, comprendiendo los métodos probabilísticos y no probabilísticos y su aplicación según los objetivos y características de cada estudio, analizando detalladamente la determinación del tamaño de muestra, la relevancia del error de muestreo y la importancia del nivel de confianza como indicadores de la validez y fiabilidad de los resultados obtenidos, de modo que los estudiantes puedan integrar estos conceptos y herramientas de manera práctica y estratégica, fortaleciendo su capacidad de sustentar decisiones y de diseñar estudios que generen información confiable y pertinente para el análisis y la gestión de datos en organizaciones.

Educación de calidad



INTRO DUC CIÓN

DE LA CLASE

GRADO / POSGRADO / TECNOLOGÍAS

Estudia a tu tiempo
son interrupciones

Resultado o resultados de aprendizaje
que será abordado con el contenido
de la clase.

- 1) Identificar los conceptos, definiciones, propiedades y teoremas de la Estadística Inferencial como base del análisis de datos en el contexto de su área de conocimiento.

1. Repaso aplicado de Estadística I

Este apartado ofrece una revisión exhaustiva de los principios fundamentales de la estadística, destacando su relevancia para la organización, análisis e interpretación de datos en distintos contextos. Se hace especial énfasis en cómo el dominio de estos conceptos básicos constituye la base para aplicar técnicas estadísticas avanzadas y tomar decisiones fundamentadas en información objetiva.

Además, se busca que el estudiante desarrolle habilidades analíticas sólidas que le permitan identificar patrones, tendencias y relaciones en conjuntos de datos, fortaleciendo su capacidad crítica y su preparación para abordar problemas prácticos en ámbitos académicos y profesionales.

1.1 Conceptos básicos

Esta sección introduce los principios fundamentales de la estadística que permiten comprender y organizar datos de manera coherente, incluyendo la identificación de patrones, tendencias y relaciones significativas dentro de conjuntos de información. Se enfatiza la importancia de dominar estas nociones como base para aplicar técnicas estadísticas más avanzadas y realizar análisis precisos que respalden la toma de decisiones.

Además, se estudia la interpretación de datos desde una perspectiva crítica, enseñando a reconocer la relevancia de cada medida y su aplicación en distintos contextos, lo que permite evaluar la información de manera objetiva y preparar al estudiante para enfrentar problemas estadísticos reales de manera efectiva y fundamentada.

1.1.1 Medidas descriptivas y de dispersión

Empecemos por lo fundamental: los datos por sí solos son solo números; la estadística descriptiva los organiza para que cuenten una historia clara. Las medidas descriptivas (o de tendencia central) nos indican dónde se concentra la información, mientras que las medidas de dispersión nos muestran qué tan extendidos o variables son esos datos.

Las principales medidas de tendencia central son la media aritmética, la mediana y la moda. La media se calcula sumando todos los valores y dividiendo por el número de observaciones; es muy útil pero sensible a valores extremos. La mediana es el valor central cuando los datos están ordenados, ideal para distribuciones asimétricas. La moda es el valor que más se repite y ayuda a identificar patrones frecuentes.

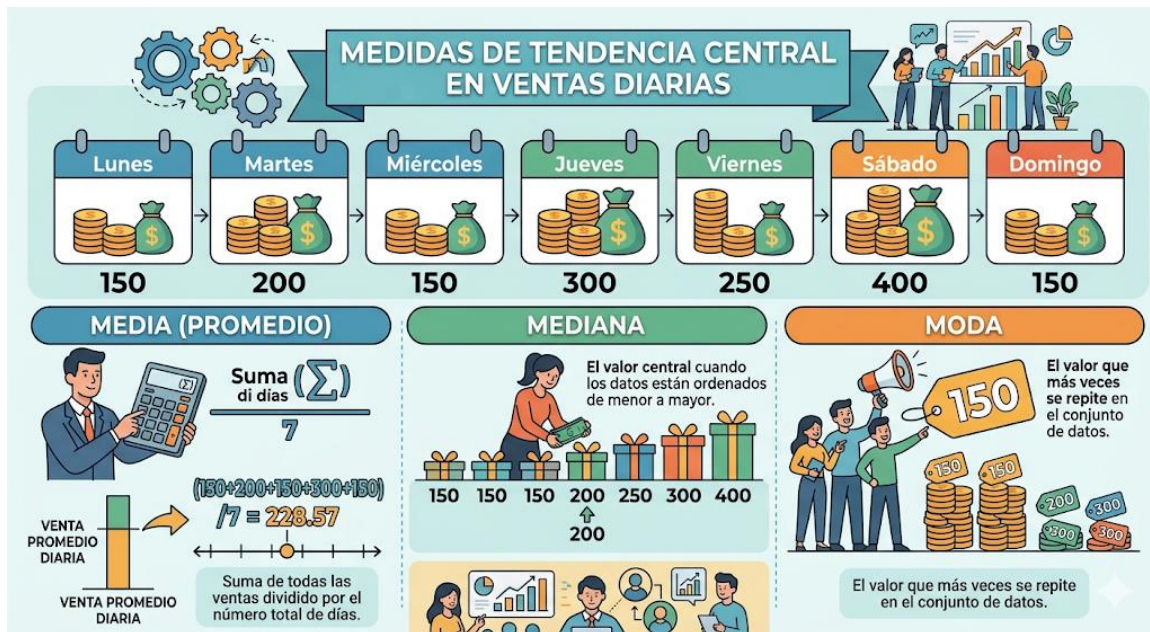


Imagen 1. Ejemplo de medidas de tendencia central aplicadas a datos de ventas diarias (creación propia del autor).

Ahora pasemos a las medidas de dispersión, que responden a la pregunta: ¿los datos están muy concentrados alrededor de la media o muy dispersos? El rango es la diferencia entre el valor máximo y mínimo; es simple pero ignora los valores intermedios. La varianza mide el promedio de las desviaciones al cuadrado respecto a la media, y la **desviación estándar** (raíz cuadrada de la varianza) expresa esa dispersión en las mismas unidades de los datos, siendo la más interpretada en la práctica empresarial.

Estas medidas nos ayudan a resumir grandes conjuntos de datos de manera eficiente. Por ejemplo, en una empresa, una baja desviación estándar en los tiempos de entrega indica procesos estables y confiables para los clientes.

1.1.2 Distribución normal y distribución Z

Muchos fenómenos naturales y empresariales siguen una distribución normal, también conocida como curva de campana: simétrica, con la media, mediana y moda coincidiendo en el centro, y donde la mayoría de los datos se concentran cerca de la media. Aproximadamente el 68% de los datos caen dentro de una desviación estándar, el 95% dentro de dos y el 99.7% dentro de tres.

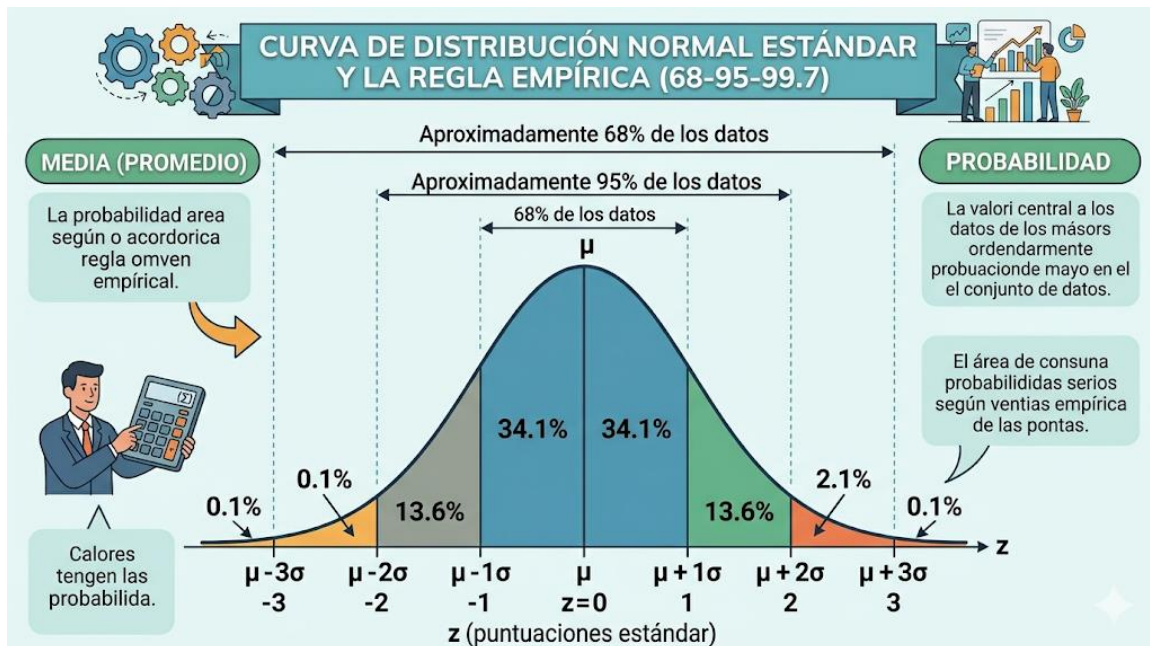


Imagen 2. Curva de distribución normal estándar con áreas de probabilidad según la regla empírica (creación propia del autor).

Para estandarizar cualquier distribución normal y compararla, usamos las puntuaciones Z (o scores Z). La fórmula es:

$$z = \frac{x - \mu}{\sigma}$$

donde x es el valor observado, μ la media y σ la desviación estándar. Una puntuación Z de +1.5 significa que el valor está 1.5 desviaciones estándar por encima de la media. Esto permite calcular probabilidades usando tablas estándar y comparar datos de diferentes escalas.

Además de su utilidad para comparar datos de distintas escalas, las puntuaciones Z permiten identificar valores atípicos o extremos dentro de un conjunto de datos. Valores Z mayores a +3 o menores a -3 suelen considerarse inusuales y pueden señalar errores de medición, eventos raros o casos excepcionales que requieren análisis adicional. Esta capacidad de detectar anomalías es especialmente valiosa en áreas como finanzas, control de calidad y medicina, donde reconocer desviaciones significativas puede prevenir pérdidas económicas, fallas de producción o diagnósticos incorrectos.

Por otro lado, la distribución normal sirve como base para muchos métodos estadísticos inferenciales, incluyendo pruebas de hipótesis y la construcción de intervalos de confianza. Al asumir normalidad en los datos, es posible estimar probabilidades de ocurrencia, comparar medias entre grupos y tomar decisiones fundamentadas con un margen de error conocido. En la práctica, esto convierte a la distribución normal y a las puntuaciones Z en herramientas esenciales para investigadores y profesionales que buscan interpretar datos de manera rigurosa y objetiva.

1.1.3 Interpretación de datos

La interpretación va más allá de calcular números: implica contextualizarlos. Una media alta de ventas no es útil si la dispersión es enorme (muchos días con ventas bajas). En la distribución normal, los valores atípicos ($Z > |3|$) pueden indicar problemas o oportunidades. Siempre pregunta: ¿este resultado es representativo? ¿Hay sesgos en los datos? Una buena interpretación combina medidas descriptivas con visualizaciones para contar la historia completa de los datos.

1.2 Aplicación en contexto empresarial

Se aborda la aplicación práctica de la estadística en entornos empresariales, mostrando cómo el análisis de datos reales permite sustentar decisiones estratégicas y optimizar procesos organizacionales. La correcta interpretación de la información posibilita identificar oportunidades de mejora, anticipar riesgos y establecer estrategias basadas en evidencia concreta.

Además, el uso de técnicas estadísticas en contextos empresariales refuerza la capacidad de evaluar resultados, generar reportes claros y comunicar hallazgos de manera efectiva a distintos niveles de la organización, fortaleciendo así la planificación y el control de actividades dentro de la empresa.

1.2.1 Análisis de datos reales

En el mundo empresarial, estos conceptos se aplican diariamente. Supongamos una cadena de retail que analiza las ventas diarias de un producto. Calculando media y desviación estándar, detecta que la media es alta pero la dispersión grande debido a picos estacionales. Usando un boxplot identifican outliers (días de promoción) y ajustan inventarios. En otro caso, una empresa de servicios evalúa tiempos de respuesta al cliente: una puntuación Z alta en retrasos alerta sobre procesos ineficientes.

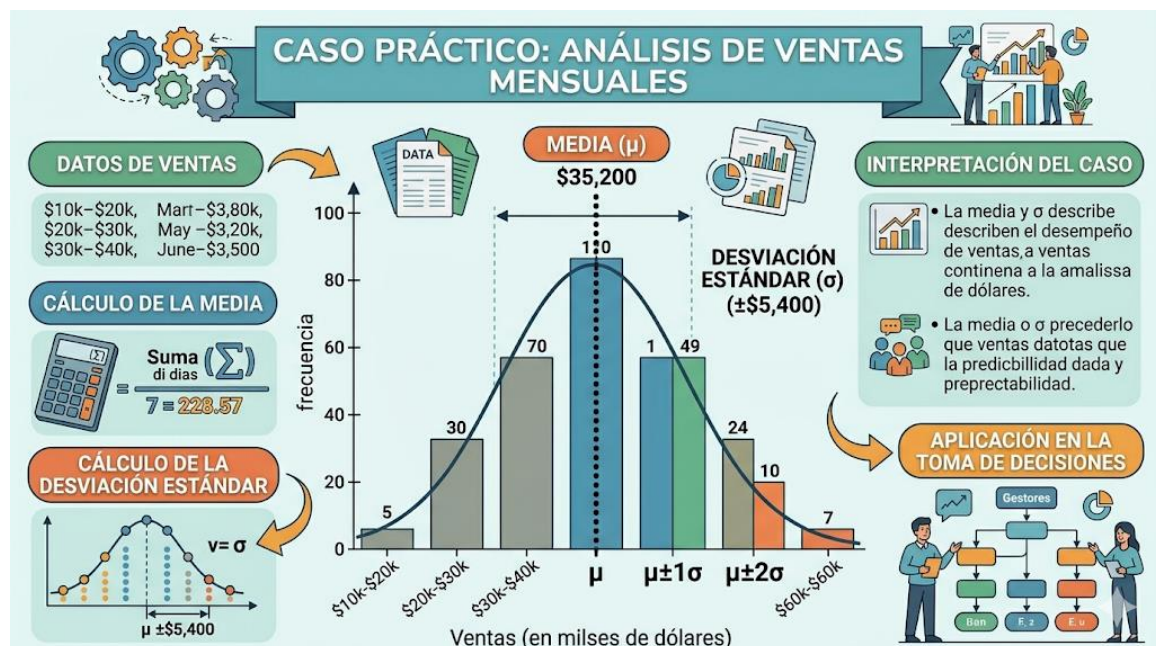


Imagen 3. Caso práctico: Histograma de ventas mensuales con media y desviación estándar marcadas (creación propia del autor).

De esta manera, la combinación de la distribución normal, los boxplots y las puntuaciones Z permite a las empresas tomar decisiones más precisas y estratégicas. Por ejemplo, al identificar los días con ventas significativamente fuera del promedio, un retailer puede ajustar promociones, planificar inventario y optimizar la logística. Asimismo, en el sector servicios, detectar desviaciones importantes en tiempos de respuesta facilita implementar mejoras operativas, reasignar recursos y elevar la satisfacción del cliente. Así, estos conceptos estadísticos no solo explican patrones históricos, sino que se convierten en herramientas predictivas que apoyan la eficiencia y competitividad empresarial.

1.2.2 Interpretación de resultados

Interpretar significa convertir números en insights accionables. Si la desviación estándar de los costos de producción es baja, el proceso es predecible y se pueden negociar mejores contratos con proveedores. Una puntuación Z de -2.3 en la satisfacción del cliente indica que ese segmento está muy por debajo del promedio, justificando una intervención inmediata. La clave es combinar la estadística con el conocimiento del negocio: ¿el outlier es un error o una oportunidad de innovación?

1.2.3 Toma de decisiones

La estadística descriptiva respalda decisiones informadas. Un gerente de marketing puede decidir aumentar el presupuesto publicitario si el análisis de dispersión muestra correlación positiva entre inversión y ventas. En recursos humanos, las puntuaciones Z ayudan a identificar empleados con desempeño excepcional ($Z > +2$) para promociones. Empresas líderes usan estas herramientas para reducir incertidumbre, optimizar recursos y ganar ventaja competitiva. Recuerda: una decisión sin datos es solo una suposición.



Imagen 4. Infografía: Flujo de toma de decisiones basado en estadística descriptiva en una empresa (creación propia del autor).

Enlace 1: Este video explica de forma clara las medidas de tendencia central y dispersión con ejemplos prácticos y uso de software. [Ver en YouTube: Estadística - Medidas de tendencia central y dispersión](#) Descripción del enlace: Este recurso visual te permitirá consolidar cómo calcular e interpretar media, mediana, moda, varianza y desviación estándar en contextos reales.

2. Técnicas de muestreo

Este apartado estudia los métodos para seleccionar muestras representativas de una población, considerando tanto muestreos probabilísticos como no probabilísticos y el impacto de posibles errores de muestreo en la confiabilidad de los resultados. La comprensión de estas técnicas es crucial para garantizar que los estudios estadísticos reflejen con precisión la realidad de la población analizada.

Además, se profundiza en el diseño muestral, incluyendo la determinación del tamaño adecuado de la muestra y el nivel de confianza necesario para obtener resultados válidos, asegurando que los datos obtenidos sean precisos, consistentes y aplicables a la resolución de problemas reales dentro de contextos académicos y profesionales.

2.1. Tipos de muestreo

El muestreo es el proceso de seleccionar una porción de la población para estudiarla y luego inferir características sobre el todo. La elección del tipo de muestreo influye directamente en la validez y generalización de los resultados.

2.1.1 Muestreo probabilístico

En el muestreo probabilístico cada elemento de la población tiene una probabilidad conocida y distinta de cero de ser seleccionado. Esto permite calcular el error de muestreo y generalizar los resultados a la población con un nivel de confianza determinado.

Los principales tipos son:

- **Aleatorio simple:** Cada elemento tiene la misma probabilidad de ser elegido (por ejemplo, sorteo de números).
- **Sistemático:** Se selecciona cada k-ésimo elemento después de un punto de partida aleatorio.
- **Estratificado:** Se divide la población en estratos homogéneos y se muestrea proporcionalmente de cada uno.
- **Por conglomerados (clusters):** Se dividen en grupos naturales y se seleccionan conglomerados completos al azar.



Imagen 5. Ventajas del muestreo probabilístico: representatividad y posibilidad de inferencia estadística (creación propia del autor).

2.1.2 Muestreo no probabilístico

En el muestreo no probabilístico la selección depende del criterio del investigador o de la disponibilidad de los elementos; no todos tienen la misma probabilidad de ser incluidos. Es más rápido y económico, pero los resultados no se pueden generalizar con precisión estadística.

Tipos principales:

- **Por conveniencia:** Se eligen los elementos más accesibles.

- **Por cuotas:** Se establecen cuotas según características y se completan con sujetos disponibles.
- **Por juicio o intencional:** El investigador selecciona según su expertise.
- **Bola de nieve:** Útil en poblaciones difíciles de alcanzar (se parte de unos pocos y se pide referencias).

A pesar de sus limitaciones en cuanto a representatividad, el muestreo no probabilístico es muy útil en investigaciones exploratorias o en estudios piloto donde se busca obtener información preliminar rápidamente. Por ejemplo, puede emplearse para identificar tendencias, generar hipótesis o evaluar la viabilidad de un estudio más amplio. Sin embargo, al depender del criterio del investigador o de la accesibilidad de los sujetos, los resultados pueden estar sesgados y no reflejar con precisión la población total, por lo que cualquier conclusión debe interpretarse con cautela y, cuando sea posible, complementarse con muestreos probabilísticos para validar los hallazgos.

2.1.3 Error de muestreo

El **error de muestreo** es la diferencia inevitable entre el valor obtenido en la muestra y el verdadero valor en la población. Se divide en error aleatorio (debido al azar) y error sistemático (sesgos en el diseño).

El margen de error (E) se relaciona directamente con el tamaño de muestra y el nivel de confianza. A mayor tamaño de muestra, menor error. Un error típico aceptable en estudios administrativos es del 3% al 5%.

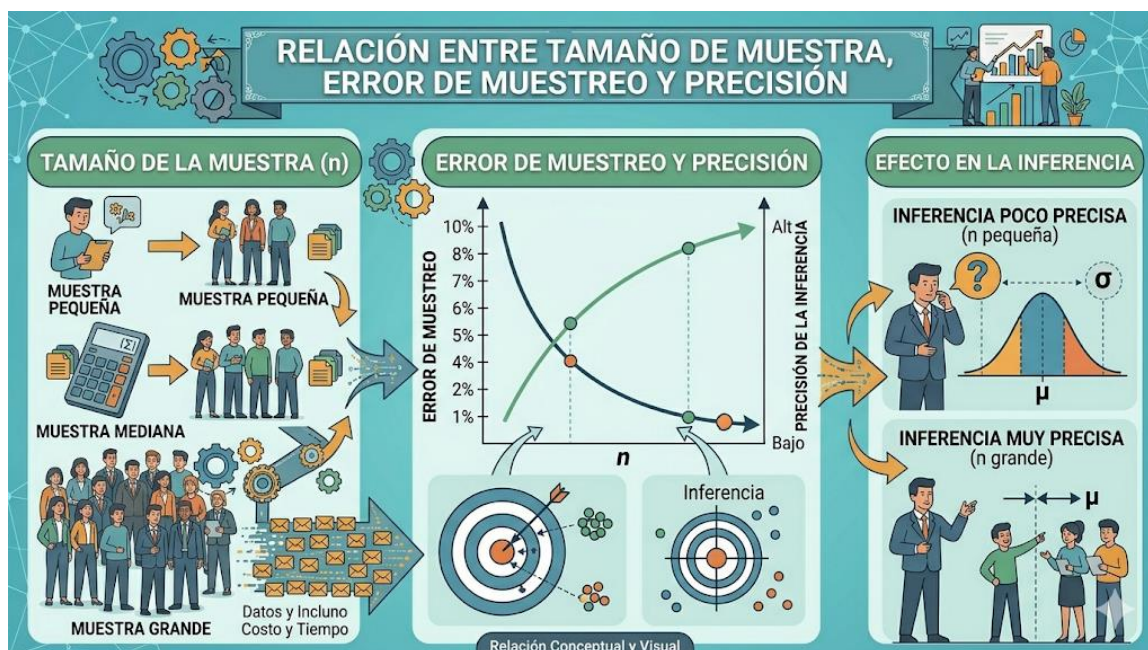


Imagen 6. Relación entre tamaño de muestra, error de muestreo y precisión (creación propia del autor).

2.2 Diseño muestral

Un buen diseño muestral combina objetivos de investigación, recursos disponibles y precisión deseada.

2.2.1 Tamaño de muestra

El tamaño de muestra (n) se calcula para garantizar representatividad. Para poblaciones finitas se usa:

$$n = \frac{N \cdot Z^2 \cdot p \cdot q}{E^2(N - 1) + Z^2 \cdot p \cdot q}$$

Para poblaciones grandes o infinitas:

$$n = \frac{Z^2 \cdot p \cdot q}{E^2}$$

donde p = 0.5 (máxima variabilidad), q = 1-p, Z depende del nivel de confianza y E es el margen de error.

2.2.2 Nivel de confianza

El nivel de confianza indica la probabilidad de que el intervalo construido contenga el parámetro poblacional real. Los más usados son 90% (Z=1.645), 95% (Z=1.96) y 99% (Z=2.576). Un 95% es el estándar en investigación administrativa.

El nivel de confianza no solo indica cuán seguro podemos estar de que el parámetro poblacional se encuentra dentro del intervalo, sino que también ayuda a gestionar riesgos en la toma de decisiones. Por ejemplo, en una empresa que analiza la satisfacción de clientes, un intervalo de confianza del 95% sobre la proporción de clientes satisfechos permite estimar con bastante certeza el comportamiento general de la población, aun si se encuestó solo una muestra. Elevar el nivel de confianza reduce la probabilidad de error, pero a su vez genera intervalos más amplios, lo que implica un balance entre precisión y certeza que los analistas deben considerar según el contexto y la criticidad de la decisión.

2.2.3 Aplicaciones

En el contexto empresarial, un buen diseño muestral se aplica en:

- Estudios de mercado para conocer preferencias de clientes.
- Encuestas de clima laboral.
- Análisis de calidad de proveedores.
- Evaluación de campañas publicitarias.

Un diseño deficiente puede llevar a decisiones erróneas y pérdidas económicas.



Imagen 7. Infografía: Flujo del diseño muestral en un proyecto de investigación administrativa (creación propia del autor).

Enlace 2: Tutorial paso a paso sobre cómo calcular el tamaño de muestra con ejemplos resueltos y explicación del rol del nivel de confianza y margen de error. [Ver en YouTube: Cálculo del Tamaño de la Muestra / Ejercicio Resuelto](#) Descripción del enlace: Ideal para practicar los cálculos que vimos y aplicarlos directamente en proyectos de investigación administrativa.

RE FE REN CIAS

- Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). John Wiley & Sons.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.
- Levine, D. M., Stephan, D. F., Krehbiel, T. C., & Berenson, M. L. (2021). *Statistics for managers using Microsoft Excel* (9th ed.). Pearson.
- Lohr, S. L. (2019). *Sampling: Design and analysis* (2nd ed.). CRC Press.
- Montgomery, D. C., & Runger, G. C. (2018). *Applied statistics and probability for engineers* (7th ed.). Wiley.

Definición de los términos citados en la semana 1

- **Desviación estándar:** Medida de dispersión que indica cuánto se alejan, en promedio, los valores individuales de la media de un conjunto de datos, expresada en las mismas unidades que los datos originales.
- **Error de muestreo:** Diferencia entre el valor estimado a partir de la muestra y el valor real del parámetro en la población; su magnitud depende del tamaño de la muestra y del nivel de confianza.

Recursos de Profundización

1. Simulación de Distribución Normal y Z

Descripción del enlace: Simulador interactivo que permite comprender la distribución normal y la estandarización mediante la distribución Z. Permite modificar media y desviación estándar para observar cambios en la curva.

Autor: Carlos Julio Mayorga Arias

Enlace: [RECURSO 1](#)

2. Calculadora de Tamaño de Muestra



Descripción del enlace: Recurso que permite calcular el tamaño de muestra según nivel de confianza, error permitido y proporción esperada, aplicado a estudios reales.

Autor: Carlos Julio Mayorga Arias

Enlace: [RECURSO 2](#)



síguenos en:



www.pucevirtual.puce.edu.ec