SESIÓN INFERENCIA E INTERVALOS DE CONFIANZA PARA LA MEDIA

CONTENIDOS:

- Inferencia estadística.
- Distribuciones muestrales:
  - Medias muestrales.
  - Proporciones muestrales.
- Cálculo de intervalos de confianza para la media con desviación estándar de la población conocida.
- Cálculo de intervalos de confianza para la media con desviación estándar de la población desconocida.
- Cálculo de intervalos de confianza para una proporción.
- Cálculo de niveles de confianza.
- Cálculo del tamaño muestral.

INFERENCIA ESTADÍSTICA

La inferencia estadística es una herramienta fundamental en la ciencia de datos, ya que permite hacer generalizaciones sobre una población a partir de muestras. Es el proceso mediante el cual se utilizan datos muestrales para extraer conclusiones válidas y confiables sobre las características de una población más amplia.

Definición y Alcance

- Definición: La inferencia estadística es el conjunto de técnicas y métodos que permiten tomar decisiones o sacar conclusiones sobre una población utilizando información obtenida a partir de una muestra.
- Propósito: Su objetivo principal es generalizar resultados observados en una muestra hacia toda la población, considerando un nivel de incertidumbre asociado al proceso de muestreo.

Componentes Clave de la Inferencia Estadística

1. Estimación de Parámetros:

   Consiste en usar los datos muestrales para estimar los valores de parámetros poblacionales desconocidos.

   Ejemplos:

   - Estimar la media poblacional (μ) a partir de la media muestral ($\bar{X}$).
   - Estimar la proporción poblacional (p) a partir de la proporción muestral ($\hat{p}$).

   Tipos de estimación:

   - Estimación puntual: Proporciona un único valor como estimación del parámetro (por ejemplo, $\overline{\mathrm{X}}$ como estimador de μ).
   - Estimación por intervalo: Proporciona un rango de valores dentro del cual se espera que esté el parámetro con cierto nivel de confianza (por ejemplo, intervalo de confianza para μ).

2. Pruebas de Hipótesis:

   - Evalúan afirmaciones sobre los parámetros poblacionales utilizando datos muestrales.
   - Ejemplo: Determinar si el ingreso promedio de una ciudad es significativamente diferente de un valor específico (hipótesis nula vs. hipótesis alternativa).
   - Elementos clave:

     Hipótesis nula H0: Afirmación inicial que se desea probar (por ejemplo, "el ingreso promedio es igual a 50,000").

     Hipótesis alternativa (Ha): Afirmación contraria a la hipótesis nula (por ejemplo, "el ingreso promedio no es igual a 50,000").

     Valor-p: Probabilidad de observar los datos muestrales bajo la hipótesis nula. Si el valor-p es menor que un nivel de significancia (α), se rechaza H0

.

Objetivo Principal

El objetivo central de la inferencia estadística es generalizar resultados de una muestra a una población, asegurando que las conclusiones sean válidas y confiables. Esto implica:

- Minimizar el error asociado al muestreo.
- Cuantificar la incertidumbre inherente al proceso de inferencia (por ejemplo, mediante intervalos de confianza o errores estándar).

Ejemplo Práctico: Estimar el ingreso promedio de una ciudad

Supongamos que queremos estimar el ingreso promedio de todos los hogares en una ciudad grande. Dado que es impráctico encuestar a todos los hogares, tomamos una muestra representativa de 500 hogares.

Estimación puntual:

- Calculamos la media muestral ($\bar{X}$) de los ingresos de los 500 hogares. Supongamos que $\bar{X}$ =48,000.
- Este valor se usa como estimación puntual del ingreso promedio poblacional (μ).

Estimación por intervalo:

- Construimos un intervalo de confianza del 95% para μ. Supongamos que el error estándar es SE=1,000. El intervalo sería:

$$\bar{X} \pm Z \cdot SE = 48,000 \pm 1.96 \cdot 1,000 = [46,040, 49,960]$$

Ilustración 1 Estimación por intervalo

Concluimos que estamos 95% seguros de que el verdadero ingreso promedio poblacional está entre 46,040 y 49,960.

Prueba de hipótesis:

- Supongamos que queremos probar si el ingreso promedio es igual a 50,000.
- Calculamos el estadístico Z:

$$Z = \frac{\bar{X} - \mu_0}{SE} = \frac{48,000 - 50,000}{1,000} = -2.0$$

Ilustración 2 Calculo estadístico Z

- Usamos la tabla de distribución normal para encontrar el valor-p. Si el valor-p es menor que α=0.05, rechazamos la hipótesis nula.

En Python

Ejemplo: Estimar el ingreso promedio de una ciudad

Supongamos que tienes una muestra de ingresos de 500 hogares y quieres calcular la media muestral ($\bar{X}$) y construir un intervalo de confianza del 95%.

```python
import numpy as np
from scipy.stats import norm

# Datos de la muestra
ingresos = np.random.normal(loc=48000, scale=10000, size=500)  # Muestra simulada (media=48000, desv=10000)

# Paso 1: Calcular la media muestral
media_muestral = np.mean(ingresos)
print(f"Media muestral: {media_muestral:.2f}")

# Paso 2: Calcular el error estándar
n = len(ingresos)
error_estandar = np.std(ingresos, ddof=1) / np.sqrt(n)  # Error estándar (ddof=1 para usar desviación estándar muestral)
print(f"Error estándar: {error_estandar:.2f}")

# Paso 3: Construir el intervalo de confianza del 95%
z_critico = norm.ppf(0.975)  # Valor crítico Z para 95% de confianza (dos colas)
limite_inferior = media_muestral - z_critico * error_estandar
limite_superior = media_muestral + z_critico * error_estandar

print(f"Intervalo de confianza (95%): [{limite_inferior:.2f}, {limite_superior:.2f}]")
```

Ilustración 3 Cálculos en Python

Salida esperada

```
Media muestral: 48123.45
Error estándar: 447.21
Intervalo de confianza (95%): [47246.78, 48999.12]
```

Ilustración 4 Salida

Prueba de Hipótesis

- Ejemplo: Probar si el ingreso promedio es igual a 50,000
- Queremos probar la siguiente hipótesis:

$$H_0 : \mu = 50,000 \text{ (el ingreso promedio es igual a } 50,000\text{).}$$
$$H_a : \mu \neq 50,000 \text{ (el ingreso promedio no es igual a } 50,000\text{).}$$

Ilustración 5 Hipótesis

```python
from scipy.stats import ttest_1samp

# Hipótesis nula
mu_0 = 50000

# Prueba de hipótesis
t_stat, p_valor = ttest_1samp(ingresos, mu_0)

# Resultados
print(f"Estadístico t: {t_stat:.2f}")
print(f"Valor-p: {p_valor:.4f}")

if p_valor < 0.05:
    print("Rechazamos la hipótesis nula (el ingreso promedio no es igual a 50,000).")
else:
    print("No rechazamos la hipótesis nula (el ingreso promedio podría ser igual a 50,000).")
```

Salida esperada:

```
Estadístico t: -4.13
Valor-p: 0.0001
Rechazamos la hipótesis nula (el ingreso promedio no es igual a 50,000).
```

Ilustración 6 Salida

Estimación de Proporciones

Ejemplo: Estimar la proporción de personas que prefieren un producto

Supongamos que encuestas a 300 personas y encuentras que 120 prefieren tu producto ($\hat{p}$ =0.40). Quieres calcular un intervalo de confianza del 95% para la proporción poblacional (p).

```
from scipy.stats import norm

# Datos de la muestra
n = 300  # Tamaño de la muestra
p_hat = 120 / n  # Proporción muestral

# Paso 1: Calcular el error estándar
error_estandar = np.sqrt(p_hat * (1 - p_hat) / n)

# Paso 2: Construir el intervalo de confianza del 95%
z_critico = norm.ppf(0.975)  # Valor crítico Z para 95% de confianza (dos colas)
limite_inferior = p_hat - z_critico * error_estandar
limite_superior = p_hat + z_critico * error_estandar

print(f"Proporción muestral: {p_hat:.2f}")
print(f"Intervalo de confianza (95%): [{limite_inferior:.2f}, {limite_superior:.2f}]")
```

Ilustración 7 Estimación de proporciones en Python

Salida esperada:

```
Proporción muestral: 0.40
Intervalo de confianza (95%): [0.34, 0.46]
```

Ilustración 8 Salida

Simulación de Muestras para Inferencia

Si deseas simular múltiples muestras y observar la distribución muestral de un estadístico (por ejemplo, la media o la proporción), puedes hacerlo de la siguiente manera:

```
import matplotlib.pyplot as plt

# Simulación de múltiples muestras
num_simulaciones = 1000  # Número de simulaciones
medias_muestrales = []

for _ in range(num_simulaciones):
    muestra = np.random.normal(loc=48000, scale=10000, size=500)  # Muestra simulada
    medias_muestrales.append(np.mean(muestra))

# Gráfico de la distribución muestral
plt.figure(figsize=(10, 6))
plt.hist(medias_muestrales, bins=30, color="skyblue", edgecolor="black", density=True, label="Distribución muestral")
x = np.linspace(min(medias_muestrales), max(medias_muestrales), 100)
plt.plot(x, norm.pdf(x, loc=48000, scale=10000 / np.sqrt(500)), color="red", label="Distribución normal aproximada")
plt.title("Distribución Muestral de Medias")
plt.xlabel("Media muestral")
plt.ylabel("Densidad")
plt.legend()
plt.grid()
plt.show()
```

Ilustración 9 Simulación de muestras para inferencia

Interpretación:

- El histograma muestra la distribución empírica de las medias muestrales obtenidas en las simulaciones.
- Se podrá ver una curva roja que representa la aproximación normal basada en los parámetros teóricos.
- Observarás que el histograma se aproxima a la curva normal, validando el Teorema del Límite Central.

DISTRIBUCIONES MUESTRALES

Las distribuciones muestrales son herramientas fundamentales en estadística inferencial, ya que describen cómo se comportan los estadísticos (como la media o la proporción) cuando se toman múltiples muestras de una población. A continuación, se detallan las distribuciones muestrales de medias y proporciones, junto con ejemplos prácticos y su implementación en Python.

Medias muestrales

La distribución muestral de medias describe cómo se distribuyen las medias muestrales ($\bar{X}$) en múltiples muestras extraídas de una población.

Teorema del Límite Central (TLC):

Para muestras grandes (n≥30), la distribución de $\bar{X}$ es aproximadamente normal:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right),$$

Ilustración 10 TLC

Donde

- μ es la media poblacional
- σ es la desviación estándar poblacional.
- n: tamaño de la muestra.

Ejemplo:

Si μ=50 y σ=10, la distribución de $\bar{X}$ para n=36 será

$$\bar{X} \sim N\left(50, \frac{10}{\sqrt{36}}\right) = N(50, 1.67)$$

Ilustración 11 Fórmula TLC

Implementación en Python:

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Parámetros de la población
mu = 50       # Media poblacional
sigma = 10    # Desviación estándar poblacional
n = 36        # Tamaño de la muestra
num_simulaciones = 1000  # Número de simulaciones

# Simular múltiples muestras y calcular sus medias
medias_muestrales = [np.mean(np.random.normal(loc=mu, scale=sigma, size=n)) for _ in range(num_simulaciones)]

# Gráfico del histograma
plt.figure(figsize=(10, 6))
plt.hist(medias_muestrales, bins=30, color="skyblue", edgecolor="black", density=True, label="Distribución muestral")

# Añadir la curva normal aproximada
x = np.linspace(min(medias_muestrales), max(medias_muestrales), 100)
error_estandar = sigma / np.sqrt(n)  # Error estándar de la media
plt.plot(x, norm.pdf(x, loc=mu, scale=error_estandar), color="red", label="Distribución normal aproximada")

# Personalización del gráfico
plt.title("Distribución Muestral de Medias", fontsize=16)
plt.xlabel("Media muestral", fontsize=14)
plt.ylabel("Densidad", fontsize=14)
plt.legend(fontsize=12)
plt.grid(alpha=0.3)

# Mostrar el gráfico
plt.show()
```

Ilustración 12 Implementación en Python

Interpretación:

- El histograma representa la distribución empírica de las medias muestrales obtenidas en las simulaciones.
- La curva roja mostrará la distribución normal teórica N (50,1.67), que se ajusta bien al histograma debido al Teorema del Límite Central.

Proporciones muestrales

La distribución muestral de proporciones describe cómo se distribuyen las proporciones muestrales ($\hat{p}$) en múltiples muestras.

Fórmula:

Para muestras grandes, $\hat{p}$ sigue una distribución normal:

$$\hat{p} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$$

Ilustración 13 Fórmula

Donde:

p: Proporción poblacional.

N: Tamaño de la muestra

Ejemplo:

Si p=0.4 y n=100, la distribución $\hat{p}$ será:

$$\hat{p} \sim N\left(0.4, \sqrt{\frac{0.4(1-0.4)}{100}}\right) = N(0.4, 0.049)$$

Implementación en Python:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Parámetros de la población
p = 0.4         # Proporción poblacional
n = 100         # Tamaño de la muestra
num_simulaciones = 1000  # Número de simulaciones

# Simular múltiples muestras y calcular sus proporciones
proporciones_muestrales = [np.random.binomial(n, p) / n for _ in range(num_simulaciones)]

# Gráfico del histograma
plt.figure(figsize=(10, 6))
plt.hist(proporciones_muestrales, bins=30, color="lightgreen", edgecolor="black", density=True, label="Distribución muestral")

# Añadir la curva normal aproximada
x = np.linspace(min(proporciones_muestrales), max(proporciones_muestrales), 100)
error_estandar = np.sqrt(p * (1 - p) / n)  # Error estándar de la proporción
plt.plot(x, norm.pdf(x, loc=p, scale=error_estandar), color="red", label="Distribución normal aproximada")

# Personalización del gráfico
plt.title("Distribución Muestral de Proporciones", fontsize=16)
plt.xlabel("Proporción muestral", fontsize=14)
plt.ylabel("Densidad", fontsize=14)
plt.legend(fontsize=12)
plt.grid(alpha=0.3)

# Mostrar el gráfico
plt.show()
```

Ilustración 14 Implementación en Python

Interpretación:

- El histograma representa la distribución empírica de las proporciones muestrales obtenidas en las simulaciones.
- La curva roja mostrará la distribución normal teórica N(0.4,0.049), que se ajusta bien al histograma debido al Teorema del Límite Central.

CÁLCULO DE INTERVALOS DE CONFIANZA PARA LA MEDIA CON DESVIACIÓN ESTÁNDAR DE LA POBLACIÓN CONOCIDA

Un intervalo de confianza es un rango de valores que probablemente contenga el parámetro poblacional con un cierto nivel de confianza. Este concepto es fundamental en estadística inferencial, ya que permite cuantificar la incertidumbre asociada a una estimación puntual.

Fórmula:

El intervalo de confianza para la media (μ) cuando la desviación estándar poblacional (σ) es conocida se calcula como:

$$\mathrm{IC} = \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}},$$

Ilustración 15 Fórmula intervalo de confianza

- donde:
    - $\bar{X}$: Media muestral.
    - $Z\alpha/2$: Valor crítico de la distribución normal estándar.
    - σ: Desviación estándar poblacional.
    - n: Tamaño de la muestra.

Ejemplo:

- $\bar{X}$ = 50 (media muestral)
- σ=10 (desviación estándar poblacional)
- n=36 (tamaño de la muestra)
- $Z\alpha/2$=1.96 (valor crítico para un nivel de confianza del 95%).

El cálculo del intervalo de confianza sería:

$$\text{Límite inferior} = \bar{X} - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 50 - 1.96 \cdot \frac{10}{\sqrt{36}} = 50 - 1.96 \cdot 1.67 = 50 - 3.27 = 46.73$$

$$\text{Límite superior} = \bar{X} + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 50 + 1.96 \cdot \frac{10}{\sqrt{36}} = 50 + 1.96 \cdot 1.67 = 50 + 3.27 = 53.27$$

Ilustración 16 Intervalo de confianza

Por lo tanto, el intervalo de confianza es:

$$\boxed{(46.73, 53.27)}$$

Ilustración 17 Resultado intervalo de confianza

Esto significa que estamos 95% seguros de que la media poblacional (μ) está dentro del rango de 46.73 a 53.27.

Implementación en Python

A continuación, se muestra cómo calcular este intervalo de confianza en Python utilizando las bibliotecas Numpy y scipy.stats.

```python
import numpy as np
from scipy.stats import norm

# Parámetros
X_bar = 50        # Media muestral
sigma = 10        # Desviación estándar poblacional
n = 36            # Tamaño de la muestra
confianza = 0.95 # Nivel de confianza (95%)

# Paso 1: Calcular el valor crítico Z_alpha/2
Z_critico = norm.ppf(1 - (1 - confianza) / 2)

# Paso 2: Calcular el margen de error
margen_error = Z_critico * (sigma / np.sqrt(n))

# Paso 3: Calcular los límites del intervalo de confianza
limite_inferior = X_bar - margen_error
limite_superior = X_bar + margen_error

# Resultados
print(f"Valor crítico Z: {Z_critico:.2f}")
print(f"Margen de error: {margen_error:.2f}")
print(f"Intervalo de confianza ({confianza * 100}%): [{limite_inferior:.2f}, {limite_superior:.2f}]")
```

Salida esperada

```
Valor crítico Z: 1.96
Margen de error: 3.27
Intervalo de confianza (95.0%): [46.73, 53.27]
```

Ilustración 18 Salida

Interpretación del Resultado

- Valor crítico $Z\alpha/2$: Para un nivel de confianza del 95%, el valor crítico es 1.96, que corresponde al percentil 97.5 de la distribución normal estándar.
- Margen de error: Representa la distancia desde la media muestral hasta los límites del intervalo de confianza. En este caso, es aproximadamente 3.27.
- Intervalo de confianza: El rango (46.73,53.27) indica que estamos 95% seguros de que la media poblacional (μ) está dentro de este intervalo.

CÁLCULO DE INTERVALOS DE CONFIANZA PARA LA MEDIA CON DESVIACIÓN ESTÁNDAR DE LA POBLACIÓN DESCONOCIDA

Cuando la desviación estándar de la población (σ) es desconocida, se utiliza la distribución t de Student en lugar de la distribución normal estándar. La distribución t es más adecuada para muestras pequeñas o cuando no se conoce σ, ya que tiene en cuenta la incertidumbre adicional asociada a la estimación de la desviación estándar a partir de la muestra.

Fórmula:

El intervalo de confianza para la media (μ) cuando la desviación estándar poblacional (σ) es desconocida se calcula como:

$$\mathrm{IC} = \bar{X} \pm t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}},$$

Ilustración 19 Fórmula intervalo de confianza

Donde:

- s: Desviación estándar muestral.
- $t_{\alpha/2,n-1}$: Valor crítico de la distribución t de Student con n−1 grados de libertad.

Ejemplo

Supongamos qué:

- $\bar{X}$ = 50 (media muestral)
- s=10 (desviación estándar muestral)
- n=20 (tamaño de la muestra)
- $t\alpha/2,19$=2.093 (valor crítico para un nivel de confianza del 95% y n−1=19 grados de libertad).

El cálculo del intervalo de confianza sería:

$$\text{Límite inferior} = \bar{X} - t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} = 50 - 2.093 \cdot \frac{10}{\sqrt{20}} = 50 - 2.093 \cdot 2.236 = 50 - 4.68 = 45.32$$

$$\text{Límite superior} = \bar{X} + t_{\alpha/2,n-1} \cdot \frac{s}{\sqrt{n}} = 50 + 2.093 \cdot \frac{10}{\sqrt{20}} = 50 + 2.093 \cdot 2.236 = 50 + 4.68 = 54.68$$

Ilustración 20 Calculo intervalo de confianza

Por lo tanto, el intervalo de confianza es:

$$\boxed{(45.32, 54.68)}$$

Ilustración 21 Resultado

Esto significa que estamos 95% seguros de que la media poblacional (μ) está dentro del rango de 45.32 a 54.68.

Implementación en Python

A continuación, te muestro cómo calcular este intervalo de confianza en Python utilizando las bibliotecas numpy y scipy.stats.

```python
import numpy as np
from scipy.stats import t

# Parámetros
X_bar = 50       # Media muestral
s = 10           # Desviación estándar muestral
n = 20           # Tamaño de la muestra
confianza = 0.95 # Nivel de confianza (95%)

# Paso 1: Calcular los grados de libertad
grados_libertad = n - 1

# Paso 2: Calcular el valor crítico t_alpha/2
t_critico = t.ppf(1 - (1 - confianza) / 2, df=grados_libertad)

# Paso 3: Calcular el margen de error
margen_error = t_critico * (s / np.sqrt(n))

# Paso 4: Calcular los límites del intervalo de confianza
limite_inferior = X_bar - margen_error
limite_superior = X_bar + margen_error

# Resultados
print(f"Valor crítico t: {t_critico:.3f}")
print(f"Margen de error: {margen_error:.2f}")
print(f"Intervalo de confianza ({confianza * 100}%): [{limite_inferior:.2f}, {limite_superior:.2f}]")
```

Ilustración 22 Implememntación en Python

Salida esperada:

```
Valor crítico t: 2.093
Margen de error: 4.68
Intervalo de confianza (95.0%): [45.32, 54.68]
```

Interpretación del Resultado

- Valor crítico $t_{\alpha/2,n-1}$: Para un nivel de confianza del 95% y 19 grados de libertad, el valor crítico es aproximadamente 2.093.
- Margen de error: Representa la distancia desde la media muestral hasta los límites del intervalo de confianza. En este caso, es aproximadamente 4.68.
- Intervalo de confianza: El rango (45.32,54.68) indica que estamos 95% seguros de que la media poblacional (μ) está dentro de este intervalo.

CÁLCULO DE INTERVALOS DE CONFIANZA PARA UNA PROPORCIÓN.

El intervalo de confianza para una proporción se utiliza para estimar el rango dentro del cual es probable que se encuentre la proporción poblacional (p) con un cierto nivel de confianza. Este cálculo es especialmente útil en estudios de opinión, encuestas y análisis de proporciones en datos categóricos.

Fórmula:

$$\mathrm{IC} = \hat{p} \pm Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Ilustración 23 Fórmula intervalo de confianza

Donde:

$\hat{p}$: Proporción muestral (estimación de la proporción poblacional).

$Z_{\alpha/2}$: Valor crítico de la distribución normal estándar correspondiente al nivel de confianza deseado.

Ejemplo:

$\hat{p}$ = 0.4 (proporción muestral)

n=100 (tamaño de la muestra)

Zα/2=1.96 (valor crítico para un nivel de confianza del 95%).

El cálculo del intervalo de confianza sería:

$$\text{Margen de error} = Z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96 \cdot \sqrt{\frac{0.4(1-0.4)}{100}} = 1.96 \cdot \sqrt{\frac{0.24}{100}} = 1.96 \cdot 0.049 = 0.096$$

$$\text{Límite inferior} = \hat{p} - \text{Margen de error} = 0.4 - 0.096 = 0.304$$

$$\text{Límite superior} = \hat{p} + \text{Margen de error} = 0.4 + 0.096 = 0.496$$

Ilustración 24 Calculo intervalo de confianza

Por lo tanto, el intervalo de confianza es: 0.304,0.496.

Esto significa que estamos 95% seguros de que la proporción poblacional (p) está dentro del rango de 0.304 a 0.496.

Implementación en Python

A continuación, se muestra cómo calcular este intervalo de confianza en Python utilizando las bibliotecas numpy y scipy.stats.

```python
import numpy as np
from scipy.stats import norm

# Parámetros
p_hat = 0.4       # Proporción muestral
n = 100           # Tamaño de la muestra
confianza = 0.95  # Nivel de confianza (95%)

# Paso 1: Calcular el valor crítico Z_alpha/2
Z_critico = norm.ppf(1 - (1 - confianza) / 2)

# Paso 2: Calcular el margen de error
margen_error = Z_critico * np.sqrt((p_hat * (1 - p_hat)) / n)

# Paso 3: Calcular los límites del intervalo de confianza
limite_inferior = p_hat - margen_error
limite_superior = p_hat + margen_error

# Resultados
print(f"Valor crítico Z: {Z_critico:.2f}")
print(f"Margen de error: {margen_error:.3f}")
print(f"Intervalo de confianza ({confianza * 100}%): [{limite_inferior:.3f}, {limite_superior:.3f}]")
```

Ilustración 25 Implementación en Python

Salida esperada:

```
Valor crítico Z: 1.96
Margen de error: 0.096
Intervalo de confianza (95.0%): [0.304, 0.496]
```

Ilustración 26 Salida esperada

Interpretación del Resultado

- Valor crítico Zα/2: Para un nivel de confianza del 95%, el valor crítico es 1.96.
- Margen de error: Representa la distancia desde la proporción muestral hasta los límites del intervalo de confianza. En este caso, es aproximadamente 0.096.
- Intervalo de confianza: El rango (0.304,0.496) indica que estamos 95% seguros de que la proporción poblacional (p) está dentro de este intervalo.

CÁLCULO DE NIVELES DE CONFIANZA

El nivel de confianza es la probabilidad de que el intervalo de confianza contenga el parámetro poblacional. Es una medida de la fiabilidad del intervalo y se expresa como un porcentaje (por ejemplo, 90%, 95%, 99%). Cuanto mayor sea el nivel de confianza, más amplio será el intervalo, ya que estamos intentando capturar el parámetro con mayor certeza.

Conceptos Clave:

- Nivel de confianza (1−α): Probabilidad de que el intervalo contenga el parámetro poblacional.
- Significancia (α): Probabilidad de error al no capturar el parámetro en el intervalo.
- Valor crítico (Zα/2): Puntuación Z correspondiente al nivel de confianza deseado.

Ejemplo:

Para un nivel de confianza del 95%:

- α=0.05,

- $\alpha/2=0.025$,
- $Z\alpha/2 = 1.96$ (valor crítico de la distribución normal estándar).

Esto significa que estamos 95% seguros de que el intervalo contiene el parámetro poblacional.

Implementación en Python

A continuación, veamos cómo calcular los valores críticos $Z\alpha/2$ para diferentes niveles de confianza utilizando Python. También incluiremos un ejemplo práctico para calcular un intervalo de confianza.

```python
from scipy.stats import norm

# Función para calcular el valor crítico Z_alpha/2
def calcular_valor_critico(nivel_confianza):
    alpha = 1 - nivel_confianza
    Z_critico = norm.ppf(1 - alpha / 2)
    return Z_critico

# Ejemplo: Calcular Z_alpha/2 para niveles de confianza comunes
niveles_confianza = [0.90, 0.95, 0.99]
valores_criticos = {f"{int(nc * 100)}%": calcular_valor_critico(nc) for nc in niveles_confianza}

print("Valores críticos Z_alpha/2 para diferentes niveles de confianza:")
for nivel, valor in valores_criticos.items():
    print(f"Nivel de confianza: {nivel}, Z_alpha/2: {valor:.3f}")
```

Ilustración 27 Implementación en Python

Salida esperada:

```
Valores críticos Z_alpha/2 para diferentes niveles de confianza:
Nivel de confianza: 90%, Z_alpha/2: 1.645
Nivel de confianza: 95%, Z_alpha/2: 1.960
Nivel de confianza: 99%, Z_alpha/2: 2.576
```

Ilustración 28 Salida esperada

Ejemplo Práctico: Cálculo de un Intervalo de Confianza

Supongamos que queremos calcular un intervalo de confianza para la media poblacional (μ) con los siguientes datos:

- Media muestral ($\bar{X}$): 50
- Desviación estándar poblacional (σ): 10,
- Tamaño de la muestra (n): 36,
- Nivel de confianza: 95%.

Fórmula

$$\text{Intervalo de confianza} = \bar{X} \pm Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

Ilustración 29 Fórmula intervalo de confianza

Implementación en Python:

```
import numpy as np
from scipy.stats import norm

# Parámetros
X_bar = 50       # Media muestral
sigma = 10       # Desviación estándar poblacional
n = 36           # Tamaño de la muestra
nivel_confianza = 0.95  # Nivel de confianza (95%)

# Paso 1: Calcular el valor crítico Z_alpha/2
alpha = 1 - nivel_confianza
Z_critico = norm.ppf(1 - alpha / 2)

# Paso 2: Calcular el margen de error
margen_error = Z_critico * (sigma / np.sqrt(n))

# Paso 3: Calcular los límites del intervalo de confianza
limite_inferior = X_bar - margen_error
limite_superior = X_bar + margen_error

# Resultados
print(f"Valor crítico Z: {Z_critico:.2f}")
print(f"Margen de error: {margen_error:.2f}")
print(f"Intervalo de confianza ({nivel_confianza * 100}%): [{limite_inferior:.2f}, {limite_superior:.2f}]")
```

Ilustración 30 Implementación en Python

Salida esperada

```
Valor crítico Z: 1.96
Margen de error: 3.27
Intervalo de confianza (95.0%): [46.73, 53.27]
```

Ilustración 31 Salida esperada

Interpretación del Resultado

- Valor crítico $Z\alpha/2$: Para un nivel de confianza del 95%, el valor crítico es 1.96.
- Margen de error: Representa la distancia desde la media muestral hasta los límites del intervalo de confianza. En este caso, es aproximadamente 3.27.
- Intervalo de confianza: El rango (46.73,53.27) indica que estamos 95% seguros de que la media poblacional (μ) está dentro de este intervalo.

CÁLCULO DEL TAMAÑO MUESTRAL

El cálculo del tamaño muestral es esencial para garantizar que una muestra sea lo suficientemente grande como para obtener resultados confiables y precisos. Este cálculo depende del nivel de confianza deseado, el margen de error permitido y la variabilidad de la población.

Fórmula

El tamaño muestral necesario para un intervalo de confianza específico se calcula como:

$$n = \left(\frac{Z_{\alpha/2} \cdot \sigma}{E} \right)^2,$$

Ilustración 32 Fórmula tamaño muestral

Donde

- n: Tamaño de la muestra.
- Zα/2: Valor crítico de la distribución normal estándar correspondiente al nivel de confianza deseado.
- σ: Desviación estándar poblacional (si es desconocida, se puede estimar con datos previos o una muestra piloto).
- E: Margen de error deseado.

Ejemplo:

Supongamos que:

- $\sigma = 10$ (desviación estándar poblacional)
- $E = 2$ (margen de error deseado)
- $Z\alpha/2 = 1.96$ (valor crítico para un nivel de confianza del 95%).

El cálculo del tamaño muestral sería:

$$n = \left(\frac{1.96 \cdot 10}{2}\right)^2 = 96.04.$$

Ilustración 33 Calculo tamaño muestral

Dado que el tamaño muestral debe ser un número entero, redondeamos hacia arriba:

$$n = 97$$

Ilustración 34 Numero entero

Por lo tanto, se requiere una muestra de al menos 97 observaciones para lograr un margen de error de 2 con un nivel de confianza del 95%.

Implementación en Python

A continuación, te muestro cómo calcular el tamaño muestral en Python utilizando las bibliotecas numpy y math.

```python
import math
from scipy.stats import norm

# Parámetros
sigma = 10       # Desviación estándar poblacional
E = 2            # Margen de error deseado
confianza = 0.95 # Nivel de confianza (95%)

# Paso 1: Calcular el valor crítico Z_alpha/2
Z_critico = norm.ppf(1 - (1 - confianza) / 2)

# Paso 2: Calcular el tamaño muestral
n = ((Z_critico * sigma) / E) ** 2
n_redondeado = math.ceil(n)  # Redondear hacia arriba

# Resultados
print(f"Valor crítico Z: {Z_critico:.2f}")
print(f"Tamaño muestral requerido: {n_redondeado}")
```

Ilustración 35 Implementación en Python

Salida esperada

```
Valor crítico Z: 1.96
Tamaño muestral requerido: 97
```

Ilustración 36 Salida esperada

Interpretación del Resultado

- Valor crítico $Z_{\alpha/2}$: Para un nivel de confianza del 95%, el valor crítico es 1.96.
- Tamaño muestral (n): Se requieren al menos 97 observaciones para lograr un margen de error de 2 con un nivel de confianza del 95%.
- Redondeo: El tamaño muestral siempre se redondea hacia arriba, ya que no es posible tener una fracción de una observación.

Todos estos conceptos son fundamentales para comprender la inferencia estadística. Proporcionan una base sólida para realizar análisis rigurosos y tomar decisiones informadas basadas en datos. Con herramientas como Python, podemos implementar estos conceptos de manera eficiente y visualizar su impacto en estudios reales.