

Guía para el Parcial de Estadística

Desarrollado a partir de los Apuntes de Cátedra y Exámenes Anteriores

Índice

1. Estimación Puntual	2
1.1. Muestra Aleatoria	2
1.2. Concepto de Estimador y Estimación	2
1.3. Propiedades Deseables de un Estimador	2
1.4. Ejemplos de Evaluación	3
2. Métodos para Encontrar Estimadores	4
2.1. Método de los Momentos (MM)	4
2.2. Método de Máxima Verosimilitud (EMV)	4
3. Intervalos de Confianza	5
3.1. Resumen de Casos Comunes	5
3.2. Introducción	6
3.3. IC para la Media: Varianza Conocida	6
3.4. IC para la Media: Varianza Desconocida	7
3.5. Consideraciones sobre el Tamaño de Muestra	7
3.6. IC para Proporciones	7
3.7. IC para la Varianza	7
4. Pruebas de Hipótesis	8
4.1. Estructura General	8
4.2. Errores en una Prueba	8
4.3. Prueba para la Media con Varianza Desconocida (población normal)	8
4.4. Prueba para Proporciones	9
4.5. Prueba para Igualdad de Medias (Muestras Independientes)	9
4.6. Relación entre IC y Pruebas	9
4.7. Ejemplo Aplicado	9

1. Estimación Puntual

La estimación puntual es un procedimiento de la inferencia estadística en el cual se intenta inferir el valor de un parámetro desconocido de una población a partir de una muestra. En lugar de estimar un rango, se propone un único valor que, bajo ciertas condiciones, es el más razonable para representar al parámetro.

1.1. Muestra Aleatoria

Una muestra aleatoria es un subconjunto de la población que se selecciona mediante un mecanismo aleatorio. Esto garantiza que cada observación sea independiente y provenga de la misma distribución que la población.

Definición 1 (Muestra Aleatoria). Sea X una variable aleatoria con función de densidad $f(x)$. Una muestra aleatoria de tamaño n está compuesta por X_1, X_2, \dots, X_n , que son variables aleatorias independientes e idénticamente distribuidas con la misma función $f(x)$.

1.2. Concepto de Estimador y Estimación

Definición 2 (Parámetro, Estimador y Estimación). Un **parámetro** θ es una característica fija pero desconocida de la población. Un **estimador** $\hat{\theta}$ es una función de los datos de la muestra que sirve para inferir el valor de θ . Una **estimación puntual** es el valor numérico obtenido al evaluar el estimador con una muestra particular.

Ejemplo práctico:

- Dada una muestra $x_1 = 24, x_2 = 30, x_3 = 27, x_4 = 32$, el estimador de la media poblacional μ es la media muestral:

$$\hat{\mu} = \bar{x} = \frac{24 + 30 + 27 + 32}{4} = 28,25$$

- Si se desconoce la varianza poblacional σ^2 , un estimador insesgado es la varianza muestral:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

1.3. Propiedades Deseables de un Estimador

- **Inssegadez:** $\hat{\theta}$ es insesgado si $E(\hat{\theta}) = \theta$.
- **Sesgo:** $b(\hat{\theta}) = E(\hat{\theta}) - \theta$. Un estimador insesgado tiene sesgo cero.
- **Varianza:** Cuanto menor es la varianza de $\hat{\theta}$, más concentrados están sus valores en torno a θ .
- **Error Cuadrático Medio:** $ECM(\hat{\theta}) = V(\hat{\theta}) + b(\hat{\theta})^2$.
- **Consistencia:** $\hat{\theta}_n \rightarrow \theta$ cuando $n \rightarrow \infty$. Esto implica que:

$$\lim_{n \rightarrow \infty} E(\hat{\theta}_n) = \theta \quad \text{y} \quad \lim_{n \rightarrow \infty} V(\hat{\theta}_n) = 0$$

1.4. Ejemplos de Evaluación

- La media muestral es un estimador insesgado de μ : $E(\bar{X}) = \mu$
- La varianza muestral con divisor $n - 1$ es insesgada:

$$E\left(\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2\right) = \sigma^2$$

- Si en cambio se usa el divisor n , el estimador resulta sesgado, con un sesgo negativo.

2. Métodos para Encontrar Estimadores

2.1. Método de los Momentos (MM)

Se basa en igualar momentos muestrales con los momentos teóricos poblacionales.

1. Calcular $E(X^k)$ como función del parámetro θ
2. Calcular $\frac{1}{n} \sum_{i=1}^n X_i^k$
3. Igualar ambos y despejar θ

Ejemplo: Sea $f(x) = \frac{2(\theta-x)}{\theta^2}$ en $[0, \theta]$. Se obtiene:

$$E(X) = \int_0^\theta x \cdot \frac{2(\theta-x)}{\theta^2} dx = \frac{\theta}{3} \Rightarrow \hat{\theta}_{MM} = 3\bar{X}$$

2.2. Método de Máxima Verosimilitud (EMV)

Busca el valor de θ que maximiza la probabilidad de observar la muestra obtenida.

1. Se construye la función de verosimilitud:

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta)$$

2. Se aplica logaritmo: $\ln L(\theta) = \sum \ln f(x_i; \theta)$
3. Se deriva respecto a θ , se iguala a cero y se despeja:

$$\frac{d}{d\theta} \ln L(\theta) = 0$$

Invarianza del EMV: Si $\hat{\theta}$ es EMV de θ , entonces $g(\hat{\theta})$ es EMV de $g(\theta)$.

3. Intervalos de Confianza

3.1. Resumen de Casos Comunes

A continuación se detallan los intervalos de confianza más comunes que pueden encontrarse en exámenes, junto con sus fórmulas y supuestos:

1. μ, σ^2 conocida:

$$\left[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right]$$

2. μ, σ^2 desconocida, población normal:

$$\left[\bar{X} - t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \cdot \frac{S}{\sqrt{n}} \right]$$

3. μ , muestra grande, σ^2 desconocida:

$$\left[\bar{X} - z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{S}{\sqrt{n}} \right]$$

4. $\mu_1 - \mu_2$, varianzas conocidas:

$$\left[(\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \right]$$

5. $\mu_1 - \mu_2$, muestras grandes, varianzas desconocidas:

$$\left[(\bar{X}_1 - \bar{X}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]$$

6. $\mu_1 - \mu_2$, varianzas iguales y desconocidas (uso de varianza combinada):

$$\left[(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2, n_1+n_2-2} \cdot S_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right]$$

$$\text{con } S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}$$

7. $\mu_1 - \mu_2$, varianzas distintas y desconocidas:

$$\left[(\bar{X}_1 - \bar{X}_2) \pm t_{\alpha/2, \nu} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \right]$$

donde ν es el número de grados de libertad aproximado con la fórmula de Welch.

8. μ_D (diferencias pareadas):

$$\left[\bar{D} \pm t_{\alpha/2, n-1} \cdot \frac{S_D}{\sqrt{n}} \right]$$

9. σ^2 :

$$\left[\frac{(n-1)S^2}{\chi_{\alpha/2, n-1}^2}, \frac{(n-1)S^2}{\chi_{1-\alpha/2, n-1}^2} \right]$$

10. σ_1^2/σ_2^2 (cociente de varianzas):

$$\left[\frac{S_1^2}{S_2^2} \cdot f_{1-\alpha/2; n_2-1, n_1-1}, \frac{S_1^2}{S_2^2} \cdot f_{\alpha/2; n_2-1, n_1-1} \right]$$

11. p (proporción):

$$\left[\hat{p} \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

12. $p_1 - p_2$:

$$\left[(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \right]$$

3.2. Introducción

En lugar de estimar un único valor para un parámetro poblacional desconocido, muchas veces es más informativo proporcionar un rango de valores posibles: un **intervalo de confianza** (IC). Este intervalo se construye de forma tal que, con una cierta probabilidad conocida (nivel de confianza), contenga al verdadero valor del parámetro.

Definición 3 (Intervalo de Confianza). *Un intervalo de confianza para un parámetro θ es un intervalo aleatorio $[\hat{\theta}_1, \hat{\theta}_2]$ construido a partir de una muestra tal que:*

$$P(\hat{\theta}_1 \leq \theta \leq \hat{\theta}_2) = 1 - \alpha$$

3.3. IC para la Media: Varianza Conocida

Supongamos $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ con σ^2 conocida. Sabemos que la media muestral \bar{X} sigue una normal:

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

Entonces podemos definir el estadístico pivote:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Para un nivel de confianza $1 - \alpha$, se cumple que:

$$P\left(-z_{\alpha/2} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq z_{\alpha/2}\right) = 1 - \alpha \Rightarrow P\left(\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Conclusión: El IC para μ es:

$$\left[\bar{X} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

Ejemplo:

Sea $n = 12$, $\bar{X} = 3250$ psi, $\sigma = \sqrt{1000} \approx 31,62$. Para 95 % de confianza:

$$z_{\alpha/2} = 1,96 \Rightarrow IC = \left[3250 - 1,96 \cdot \frac{31,62}{\sqrt{12}}, 3250 + 1,96 \cdot \frac{31,62}{\sqrt{12}} \right]$$

3.4. IC para la Media: Varianza Desconocida

Cuando σ^2 no se conoce, se reemplaza por la varianza muestral S^2 y el pivote se transforma en una t de Student:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

IC:

$$\left[\bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \right]$$

3.5. Consideraciones sobre el Tamaño de Muestra

Dado un nivel de confianza y una precisión deseada L , el tamaño muestral necesario puede calcularse:

$$n \geq \left(\frac{z_{\alpha/2} \cdot \sigma}{L} \right)^2$$

Esto permite planificar experimentos para lograr estimaciones con error máximo acotado.

3.6. IC para Proporciones

Para estimar la proporción p de éxito:

- Supuestos: muestra grande, $np > 10$ y $n(1-p) > 10$.

IC:

$$\left[\hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right]$$

3.7. IC para la Varianza

Para población normal, con varianza desconocida:

$$\left[\frac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}, \frac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}} \right]$$

Donde χ^2 es la distribución chi-cuadrado con $n-1$ grados de libertad.

4. Pruebas de Hipótesis

4.1. Estructura General

Una prueba de hipótesis es un procedimiento estadístico para decidir, a partir de los datos muestrales, si se acepta o se rechaza una afirmación sobre un parámetro poblacional.

1. Formular las Hipótesis:

- Hipótesis nula: H_0 (por ejemplo, $\mu = \mu_0$)
- Hipótesis alternativa: H_1 (por ejemplo, $\mu > \mu_0$, $\mu < \mu_0$ o $\mu \neq \mu_0$)

2. Fijar el nivel de significación α (usualmente 0.05 o 0.01).

3. Elegir el estadístico de prueba y determinar su distribución bajo H_0 .

4. Obtener el valor observado del estadístico a partir de la muestra.

5. Calcular el p-valor o definir la región de rechazo.

6. Tomar la decisión: Rechazar H_0 si el p-valor $< \alpha$, o si el estadístico cae en la región de rechazo.

4.2. Errores en una Prueba

- **Error tipo I:** Rechazar H_0 siendo verdadera (probabilidad α).
- **Error tipo II:** No rechazar H_0 siendo falsa (probabilidad β).
- **Potencia de la prueba:** $1 - \beta$, probabilidad de detectar un efecto real.

4.3. Prueba para la Media con Varianza Desconocida (población normal)

- Estadístico:

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t_{n-1}$$

- Hipótesis:

- $H_0 : \mu = \mu_0$
- $H_1 : \mu \neq \mu_0$ (o unilaterales)

- Rechazo de H_0 si $|T| > t_{\alpha/2, n-1}$, o si el p-valor $< \alpha$.

4.4. Prueba para Proporciones

- Estadístico:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \sim N(0, 1)$$

- Hipótesis:

- $H_0 : p = p_0$
- $H_1 : p \neq p_0, p > p_0 \text{ o } p < p_0$

- Rechazo de H_0 si $|Z| > z_{\alpha/2}$ o según el p-valor.

4.5. Prueba para Igualdad de Medias (Muestras Independientes)

- Supuesto: poblaciones normales con varianzas iguales

- Estadístico:

$$T = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

con

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

4.6. Relación entre IC y Pruebas

Para pruebas bilaterales, si el valor hipotético cae fuera del IC de $1 - \alpha$, se rechaza H_0 al nivel α .

4.7. Ejemplo Aplicado

Se desea probar si la varianza poblacional es $\sigma^2 = 0,025$ con $\alpha = 0,01$. Se calcula el IC bilateral para la varianza con nivel de confianza 99 %. Si 0,025 cae fuera del IC, se rechaza H_0 .