Guía para el Parcial de Estadística

Desarrollado a partir de los Apuntes de Cátedra y Exámenes Anteriores

2 de julio de 2025

Índice

1. Estimación Puntual

En estadística, a menudo no tenemos acceso a toda la población, por lo que debemos inferir sus características (parámetros) a partir de una muestra. La estimación puntual es el proceso de encontrar un único valor, derivado de los datos de la muestra, que sirva como la "mejor conjetura" para un parámetro poblacional desconocido.

1.1. Conceptos Fundamentales

Definición 1 (Parámetro y Estimador)**.** *Un **parámetro** (θ) es una característica numérica de una población (ej. la media poblacional μ, la varianza σ^2). Un **estimador** ($\hat{\Theta}$) es una regla o función matemática que nos dice cómo calcular una estimación de un parámetro a partir de los datos de una muestra. Es una variable aleatoria, ya que su valor depende de la muestra seleccionada.*

1.2. Métodos para Encontrar Estimadores

Para proponer buenos estimadores, la teoría estadística nos provee de métodos formales. Los dos más importantes evaluados en los exámenes son:

1.2.1. Método de los Momentos (MM)

Este método se basa en una idea intuitiva: los momentos calculados a partir de la muestra deben ser buenas aproximaciones de los momentos teóricos de la población.

- **Procedimiento Detallado**:
  1. **Calcular Momentos Poblacionales**: Se expresa el momento poblacional de orden k, $\mu_k = E(X^k)$, como una función del parámetro o parámetros desconocidos θ. Para la mayoría de los problemas, basta con el primer momento ($k = 1$), que es la esperanza $E(X)$.
  2. **Calcular Momentos Muestrales**: Se calcula el momento muestral correspondiente, $M_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$. Para $k = 1$, este es simplemente la media muestral, $M_1 = \overline{X}$.
  3. **Igualar y Resolver**: Se igualan los momentos teóricos y muestrales ($\mu_k = M_k$) y se resuelve la ecuación para el parámetro desconocido θ, obteniendo así su estimador $\hat{\theta}_{MM}$.
- **Ejemplo de Aplicación**: Consideremos una variable aleatoria con función de densidad $f(x) = \frac{2(\theta - x)}{\theta^2}$ para $0 \leq x \leq \theta$. En un examen se pidió encontrar el estimador de θ. El primer paso es calcular la esperanza teórica:

$$E(X) = \int_0^\theta x \cdot \frac{2(\theta - x)}{\theta^2} dx = \frac{2}{\theta^2}\int_0^\theta (\theta x - x^2) dx = \frac{2}{\theta^2}\left[\frac{\theta x^2}{2} - \frac{x^3}{3}\right]_0^\theta = \frac{\theta}{3}$$

  Ahora, se iguala este resultado a la media muestral: $\frac{\theta}{3} = \overline{X}$. Al despejar θ, obtenemos el estimador por el método de los momentos: $\hat{\Theta}_{MM} = 3\overline{X}$.

1.2.2. Método de Máxima Verosimilitud (EMV)

Este es el método más importante y con mejores propiedades asintóticas. Su objetivo es encontrar el valor del parámetro que hace que la muestra que hemos observado sea lo más probable posible.

- **Procedimiento Detallado**:
  1. **Función de Verosimilitud** $L(\theta)$: Se construye la función de verosimilitud como el producto de la función de densidad (o probabilidad) de la población, evaluada en cada uno de los datos de la muestra. Asumiendo independencia, esto es: $L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$.
  2. **Función de Log-Verosimilitud** $\ln L(\theta)$: Para simplificar la maximización, se aplica el logaritmo natural. Esto transforma el producto en una suma, mucho más fácil de derivar: $\ln L(\theta) = \sum_{i=1}^{n} \ln f(x_i; \theta)$.
  3. **Maximización**: Se deriva la función de log-verosimilitud con respecto al parámetro θ y se iguala a cero: $\frac{d}{d\theta} \ln L(\theta) = 0$.
  4. **Resolución**: Se despeja θ de la ecuación anterior para hallar su estimador, $\hat{\theta}_{EMV}$.
- **Propiedad de Invarianza**: Si $\hat{\theta}$ es el EMV de θ, el EMV de una función $g(\theta)$ es $g(\hat{\theta})$. Esto es fundamental para problemas donde se pide estimar una probabilidad que depende del parámetro, como $P(X \leq 9)$ para una distribución exponencial.

1.3. Evaluación de la Calidad de los Estimadores

Un estimador es una variable aleatoria, y por lo tanto, puede tener "buenas.º "malas"propiedades.

- **Sesgo (Precisión)**: Un estimador es **insesgado** si, en promedio, su valor coincide con el del parámetro. Formalmente, $E(\hat{\Theta}) = \theta$. El **sesgo** es $b(\hat{\Theta}) = E(\hat{\Theta}) - \theta$. Un sesgo cero es una propiedad deseable.
  - La media muestral $\overline{X}$ es insesgada para μ, ya que $E(\overline{X}) = E(\frac{1}{n} \sum X_i) = \frac{1}{n} \sum E(X_i) = \frac{1}{n} \cdot n\mu = \mu$.
  - La varianza muestral $S^2 = \frac{1}{n-1} \sum (X_i - \overline{X})^2$ es insesgada para σ^2. Es crucial el divisor $n-1$.
- **Error Cuadrático Medio (Eficiencia)**: El ECM mide el error esperado al cuadrado y es la métrica principal para comparar estimadores: $ECM(\hat{\Theta}) = E[(\hat{\Theta} - \theta)^2]$. Se demuestra que es igual a la suma de la varianza del estimador y el cuadrado de su sesgo:

  $$ECM(\hat{\Theta}) = V(\hat{\Theta}) + [b(\hat{\Theta})]^2$$

  Para estimadores insesgados, el ECM es simplemente su varianza. Entre dos estimadores insesgados, el más **eficiente** es el que tiene menor varianza.
- **Consistencia**: Un estimador es consistente si su valor converge al verdadero valor del parámetro a medida que el tamaño de la muestra n tiende a infinito. Un teorema fundamental establece que $\hat{\Theta}_n$ es consistente si:

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

  En los exámenes, para probar la consistencia, se deben verificar estas dos condiciones.

2. Intervalos de Confianza (IC)

Un IC es un rango de valores, calculado a partir de la muestra, que se espera que contenga al parámetro poblacional con un cierto **nivel de confianza** $(1-\alpha)$. La interpretación correcta es que si repitiéramos el experimento muchas veces, el $(1-\alpha)\,\%$ de los intervalos así construidos contendrían el verdadero y fijo valor del parámetro.

2.1. IC para la Media Poblacional μ

Caso A: Varianza σ^2 Conocida

- **Objetivo**: Estimar el valor de μ cuando conocemos la variabilidad de la población.
- **Supuestos**: La población se distribuye de forma Normal, o la muestra es grande ($n \geq 30$) para poder invocar el Teorema Central del Límite.
- **Estadístico Pivote**: $Z = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$, que tiene una distribución Normal estándar, $N(0,1)$.
- **Fórmula del Intervalo**:

$$\overline{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Caso B: Varianza σ^2 Desconocida

- **Subcaso: Muestra Grande ($n \geq 30$)**: El TCL nos permite aproximar la distribución de la media muestral a una normal. Se reemplaza σ por su estimación, la desviación estándar muestral s.

$$\overline{X} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$$

- **Subcaso: Muestra Pequeña ($n < 30$)**: Este caso requiere un supuesto más fuerte.
  - **Supuesto Fundamental**: La población de origen **debe** ser Normal.
  - **Estadístico Pivote**: $T = \frac{\overline{X}-\mu}{S/\sqrt{n}}$, que sigue una distribución t de Student con $n-1$ grados de libertad.
  - **Fórmula del Intervalo**:

$$\overline{X} \pm t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}$$

2.2. IC para la Diferencia de Medias ($\mu_1 - \mu_2$)

Caso A: Muestras Independientes

- **Supuestos Clave**: Ambas poblaciones son normales y las muestras son independientes. La elección de la fórmula depende de si las varianzas son iguales o no.
- **Subcaso: Varianzas desconocidas pero asumidas iguales ($\sigma_1^2 = \sigma_2^2$)**. Un escenario muy común en los exámenes.
  - **Procedimiento**: Se utiliza la **varianza combinada (pooled)**, S_p^2, que es un promedio ponderado de las varianzas muestrales, para obtener una estimación más robusta de la varianza común σ^2.
  - **Estadístico Pivote**: $T = \frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{S_p\sqrt{1/n_1+1/n_2}} \sim t_{n_1+n_2-2}$.
  - **Fórmula del Intervalo**:

$$(\overline{X}_1 - \overline{X}_2) \pm t_{\alpha/2,n_1+n_2-2}S_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$$

  donde la varianza combinada se calcula como $S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$.

Caso B: Muestras Pareadas (Datos Dependientes) Este es un escenario crucial y muy evaluado.

- **Contexto Típico**: Mediciones de .antes y después.en los mismos individuos (ej. una dieta); comparación de dos métodos en las mismas unidades experimentales (ej. dos compiladores en los mismos programas).
- **Supuesto**: La diferencia de los valores de cada par, $D_i = X_{antes,i} - X_{después,i}$, sigue una distribución Normal.
- **Procedimiento Detallado**:
  1. Para cada par, calcular la diferencia D_i. Esto transforma el problema de dos muestras en un problema de una sola muestra (la muestra de diferencias).
  2. Calcular la media de estas diferencias, $\overline{D}$, y su desviación estándar, S_D.
  3. Construir el IC para la media de las diferencias, μ_D, usando la distribución t de Student con $n-1$ grados de libertad, donde n es el número de pares.
- **Fórmula del Intervalo**:

$$\overline{D} \pm t_{\alpha/2,n-1} \frac{S_D}{\sqrt{n}}$$

2.3. IC para Proporciones y Varianzas

IC para una Proporción p

- **Supuesto**: La muestra debe ser grande para que la aproximación Normal sea válida ($n\hat{p} > 10$ y $n(1-\hat{p}) > 10$).
- **Fórmula**:

$$\hat{p} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

IC para la Varianza σ^2

- **Supuesto Crítico**: La población de origen debe ser Normal.
- **Estadístico Pivote y Fórmula**: Se basa en la distribución Ji-cuadrado (χ^2).

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

3. Pruebas de Hipótesis

Una prueba de hipótesis es un procedimiento formal para usar la evidencia de una muestra para decidir entre dos afirmaciones contrapuestas sobre una población.

3.1. La Estructura de una Prueba de Hipótesis

1. **Paso 1: Formular las Hipótesis**.
   - **Hipótesis Nula (H_0)**: Es la hipótesis de "no cambio", "no efecto.º el status quo. Se formula con una igualdad (o $\geq, \leq$). Ejemplo: $H_0 : \mu = 4{,}5$ kg (la dieta reduce el peso en 4.5 kg en promedio).
   - **Hipótesis Alternativa (H_1)**: Es la afirmación que el investigador quiere probar. Determina si la prueba es unilateral ($<, >$) o bilateral ($\neq$). Ejemplo: $H_1 : \mu \neq 4{,}5$ kg.
2. **Paso 2: Establecer el Nivel de Significancia (α)**. Es la probabilidad de cometer un Error de Tipo I (rechazar H_0 cuando es verdadera), que es el error que se controla directamente. Usualmente $\alpha = 0{,}05$ o $\alpha = 0{,}01$.
3. **Paso 3: Identificar y Calcular el Estadístico de Prueba**. Es un valor calculado a partir de la muestra que mide qué tan lejos están los datos de lo que se esperaría si H_0 fuera cierta. Su fórmula depende del parámetro en cuestión y de los supuestos.
4. **Paso 4: Tomar la Decisión**.
   - **Método del p-valor (Preferido en los exámenes)**: Se calcula el p-valor, que es la probabilidad de observar un estadístico de prueba tan extremo o más que el calculado, asumiendo que H_0 es cierta.
     - Si p-**valor** $< \alpha$, la muestra observada es muy improbable bajo H_0. La evidencia es fuerte en contra de H_0, por lo que se la **rechaza**.
     - Si p-**valor** $\geq \alpha$, la muestra es compatible con H_0. No hay evidencia suficiente para rechazarla, por lo tanto **no se rechaza** H_0.
5. **Paso 5: Concluir en el Contexto del Problema**. La conclusión nunca es "se acepta H_0", sino "no hay evidencia estadística suficiente para rechazar H_0". La conclusión debe responder a la pregunta original del problema.

3.2. Guía Detallada de Pruebas Comunes en Exámenes

Prueba T para una Media μ (Pob. Normal, σ^2 desconocida)

- **Objetivo**: Probar una afirmación sobre μ (ej: el tiempo de acceso medio es de 15 ms).
- **Hipótesis**: $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ (o unilaterales).
- **Estadístico de Prueba**: $T = \frac{\overline{X}-\mu_0}{S/\sqrt{n}} \sim t_{n-1}$ bajo H_0.
- **Cálculo del p-valor**: Para $H_1 : \mu < \mu_0$, p-valor $= P(T < t_{obs})$. Para $H_1 : \mu > \mu_0$, p-valor $= P(T > t_{obs})$. Para $H_1 : \mu \neq \mu_0$, p-valor $= 2 \cdot P(T > |t_{obs}|)$.

Prueba T para Muestras Pareadas (μ_D)

- **Objetivo**: Comparar dos condiciones en las mismas unidades (ej: eficacia de una dieta, comparación de dos compiladores).

- **Hipótesis**: Se formulan sobre la diferencia media, μ_D. Ej: $H_0 : \mu_D = 4{,}5$ vs $H_1 : \mu_D \neq 4{,}5$.
- **Procedimiento**: Se calculan las diferencias D_i y se realiza una prueba T para una sola muestra sobre estas diferencias.
- **Estadístico de Prueba**: $T = \frac{\overline{D}-\Delta_0}{S_D/\sqrt{n}} \sim t_{n-1}$ bajo H_0.

Prueba Z para la Diferencia de Proporciones ($p_1 - p_2$)

- **Objetivo**: Determinar si la proporción de un evento difiere entre dos poblaciones independientes (ej: efectividad de dos tipos de empaques).
- **Hipótesis**: Generalmente se prueba si son iguales: $H_0 : p_1 - p_2 = 0$ vs $H_1 : p_1 - p_2 \neq 0$ (o unilaterales).
- **Estadístico de Prueba**: Como bajo H_0 se asume que $p_1 = p_2 = p$, se usa una estimación combinada de la proporción, $\hat{p} = \frac{x_1+x_2}{n_1+n_2}$.

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}}$$

3.3. La Relación entre Pruebas de Hipótesis e Intervalos de Confianza

Esta dualidad es una herramienta poderosa y una forma muy común de resolver pruebas de hipótesis en los exámenes.

- **Principio Fundamental**: Para un nivel de significancia α, una prueba de hipótesis **bilateral** ($H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$) se puede resolver construyendo un intervalo de confianza para θ con nivel $(1 - \alpha)\,\%$.
- **Regla de Decisión a través del IC**:
  - Si el valor hipotético θ_0 de la hipótesis nula **está contenido** dentro del intervalo de confianza, entonces **no se puede rechazar** H_0. La evidencia muestral es compatible con el valor θ_0.
  - Si el valor θ_0 **cae fuera** del intervalo de confianza, entonces se **rechaza** H_0. La evidencia muestral sugiere que el verdadero valor del parámetro es diferente de θ_0.
- **Ejemplo de Aplicación (típico de examen)**: Se pide probar $H_0 : \sigma^2 = 0{,}025$ vs $H_1 : \sigma^2 \neq 0{,}025$ con $\alpha = 0{,}01$. Para resolverlo, se construye un IC del 99 % para σ^2. Si 0.025 está en el intervalo, no se rechaza H_0. Si no lo está, se rechaza.