Tema 5. Contraste de hipótesis (I)

Dr. David Castilla Espino
CA UNED de Huelva, "Profesor Dr. José Carlos Vílchez Martín"

Introducción

- Bienvenida
- **Objetivos** pedagógicos:
  - Conocer el concepto de hipótesis estadística
  - Conocer y estimar los distintos errores que se cometen a la hora de realizar un contraste de hipótesis, así como la potencia de un contraste de hipótesis simples.
  - Conocer las fases de elaboración de un contraste de hipótesis estadística.

Introducción

- Comunicación:
  - Online: Chat de webconferencia.


  - Offline: Foro del grupo de tutoría correspondiente.
- Referencias:
  - Casas, J.M. & P. Gutiérrez (2011): Estadística II: Inferencia Estadística. Editorial Universitaria Ramón Areces, Madrid. 231-299.
  - Novales (1997), Ruiz Maya & Martín Pliego (1999),...

Introducción

- Antecedentes
  - Inferencia estadística (Tema 1). Utiliza datos muestrales para llevar a cabo estimaciones, tomar decisiones, realizar predicciones, comprobar hipótesis u otras generalizaciones acerca de un conjunto de datos más grande denominado población.

Control de calidad

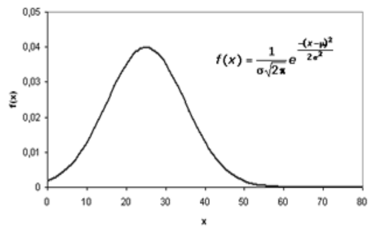

Introducción

- Antecedentes:
  - Procedimientos de la inferencia estadística:
    - Estimación de parámetros poblacionales (Temas 2-4).
    - Contrastación de hipótesis estadística acerca de la población (Temas 5-7).

Dr. David Castilla Espino
CA UNED Huelva

Conceptos generales

- **Hipótesis estadística**: Es una conjetura sobre alguna característica desconocida de la población.
- **Contraste de hipótesis estadística**: Es el procedimiento por el que se verifica una hipótesis estadística mediante el empleo de una muestra. Existen dos tipos de contrastes:
  - **Paramétricos**. Contrastan hipótesis sobre el valor que toman los parámetros de distribuciones poblacionales conocidas (Tema 6). Ej. Contrastar hipótesis sobre μ y σ de una familia de distribuciones Normal.



Dr. David Castilla Espino
CA UNED Huelva

Conceptos generales

- **Contraste de hipótesis estadística** (…):
  - **No paramétricos**: Contrastan hipótesis sobre otras características de la distribución distintas de los parámetros: forma, localización, aleatoriedad,… (Tema 7).
- Tipología de hipótesis estadísticas:
  - **Simples**: La hipótesis estadística se refiere a un solo valor del parámetro poblacional.

$$H_0 : \theta = \theta_0$$

  - **Compuestas**: La hipótesis estadística se refiere a una región del espacio paramétrico y puede tomar por lo tanto varios valores diferentes:

$$H_1 : \theta > \theta_0 \qquad H_1 : \theta < \theta_0 \qquad H_1 : \theta \neq \theta_0$$

Conceptos generales

- Tipología de hipótesis estadísticas (…):
  - **Hipótesis nula (H_0)**: Es aquello que se supone cierto a priori. Representa el *status quo*. Siempre ha de incluir la igualdad. Se contrasta comparando la discrepencia existente entre su valor y el estimado mediante una muestra. Si la discrepancia es pequeña se acepta, en caso contrario se rechaza.
  - **Hipótesis alternativa (H_1)**: Es el complementario de la hipótesis nula. Es aquello que aceptaremos en caso de que la hipótesis nula sea rechazada.
- Tipología de contrastes hipótesis paramétricos:
  - **Bilaterales (o de dos colas)**: Son aquellos en los que la hipótesis alternativa es compuesta y no especifica una dirección. En estos casos, la H_1 toma la forma: H_1: $\theta \neq \theta_0$.

Conceptos generales

- Tipología de contrastes hipótesis paramétricos (…):
  - **Unilaterales (o de una sola cola)**: Son aquellos en los que la hipótesis alternativa es compuesta y especifica una dirección. En estos casos, la H_1 toma la forma: H_1: $\theta > \theta_0$ (a la derecha) ó H_1: $\theta < \theta_0$ (a la izquierda).
- Formule las hipótesis nula y alternativas e indique si el contraste es bilateral o unilateral en los casos que se describen a continuación:
  - Un fabricante de bombillas afirma que, en promedio, la duración de cada bombilla es al menos 1000 horas.

$$H_0: \mu \geq 1000$$
$$H_1: \mu < 1000$$

Conceptos generales

- Formule las hipótesis nula y alternativas en los casos que se describen a continuación (…):
  - Una empresa recibe un lote de productos. A priori se considera el envío válido, salvo que se compruebe mediante un muestreo que hay más 5% de piezas defectuosas.

$$H_0: p \leq 0.05$$
$$H_1: p > 0.05$$

  - Un investigador quiere saber si el salario medio de los hombres y las mujeres son iguales o no.

$$H_0: \mu_x - \mu_y = 0$$
$$H_1: \mu_x - \mu_y \neq 0$$

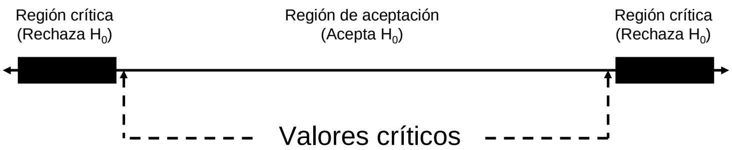
Dr. David Castilla Espino
CA UNED Huelva

Región crítica y de aceptación

- Esquema general de todo contraste de hipótesis:
  - Se establecen las hipótesis del contraste. Ej. piezas defectuosas:
    - H_0: $p \leq 0,05$ (se supone inicialmente cierta)
    - H_1: $p > 0,05$
  - Se recoge información muestral para verificar la hipótesis nula. Ej.:
    - p muestral = 0,04
  - Se establece una regla de decisión basada en la discrepancia entre los valores muestrales y los de la H_0. Ej.:
    - Si p muestral ≤ 0,07 ➔ **Acepta H_0**
    - Si p muestral > 0,07 ➔ Rechaza H_0 ➔ Región crítica

Región crítica y de aceptación

- **Región crítica**: Está constituida por el conjunto de muestras para las que se rechaza la H_0.
- **Región de aceptación**: Está constituida por el conjunto de muestras para las que se acepta la H_0.
- Ej.: Regiones crítica y de aceptación de un contraste bilateral.



Dr. David Castilla Espino
CA UNED Huelva

Errores y potencia de un contraste

		Estados de la naturaleza (Hipótesis cierta)	
		H_0	H_1
Decisión	H_0	Decisión Correcta	**Error de tipo II (β)**
	H_1	**Error de tipo I (α)**	Decisión Correcta

α = P(Rechazar H_0|H_0 cierta) = Nivel de significación

Nivel de confianza =1-α = P(No rechazar H_0|H_0 cierta)

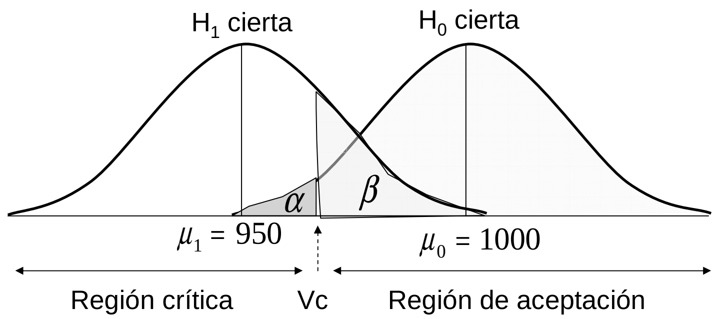
β = P(Aceptar H_0|H_0 falsa)

Potencia del contraste =1-β = P(Rechazar H_0|H_0 falsa)

Dr. David Castilla Espino
CA UNED Huelva

Errores y potencia de un contraste

- Ej.: Considere las siguientes hipótesis simples en relación con la media de una población Normal (μ; σ):
  - H_0: μ_0 = 1000
  - H_1: μ_1 = 950



Dr. David Castilla Espino
CA UNED Huelva

Errores y potencia de un contraste

- **Ejemplo**: Sea una variable aleatoria que se distribuye según un distribución N(μ,σ=27). Sobre el parámetro μ de esta distribución se desea contrastar la hipótesis nula H_0: μ=110 frente a la alternativa H_1: μ=130, mediante una muestra aleatoria simple de tamaño 81, siendo la región crítica el intervalo {Media(x)≥114}. Determine los errores tipo I y II, así como la potencia del contraste.

  Solución:
  - Calculo del error tipo I (α):

$$\alpha = P[\bar{x} \geq 114/H_0(\mu = 110)] = P\left(z \geq \frac{114-110}{27/\sqrt{81}} \right) = P\left(z \geq \frac{4}{3} \right) = 0{,}0912$$

  - Calculo del error tipo II (β)

$$\beta = P[\bar{x} < 114/H_1(\mu = 130)] = P\left(z < \frac{114-130}{27/\sqrt{81}} \right) = P\left(z < \frac{-26}{3} \right) \approx 0$$

Errores y potencia de un contraste

- **Solución (…)**:
  - Calculo de la potencia del contraste:

$$1-\beta = 1 - P[\bar{x} < 114/H_1(\mu = 130)] = 1 - P\left(z < \frac{114-130}{27/\sqrt{81}} \right) \approx 1$$

Observese que para un error tipo I (α), conforme el tamaño muestral crece, la varianza del estimador disminuye y consecuentemente la potencia del contraste (1-β) mejora.

Errores y potencia de un contraste

- **Ejemplo**: Los errores de fabricación de un cierto proceso se distribuyen de acuerdo con la función de densidad f(x;θ)= $\theta e^{-\theta x}$, para x≥0 y θ>0. Sobre el parámetro de esta función de densidad se desea contrastar la hipótesis nula H_0: θ=1 frente a la alternativa H_1: θ=2, mediante una muestra aleatoria de tamaño 1, siendo la región crítica el intervalo $\{0 \le x_1 \le K\}$. Calcule los errores tipo I y II en función de K, así como la relación entre estos dos tipos de errores.

  Solución:
  - Calculo del error tipo I (α):

$$\alpha = P\left[0 \le x_1 \le K / H_0(\theta = 1)\right] = \int_0^K e^{-x_1}\,dx = 1 - e^{-K}$$

  de donde se puede deducir que K=-ln(1- α)

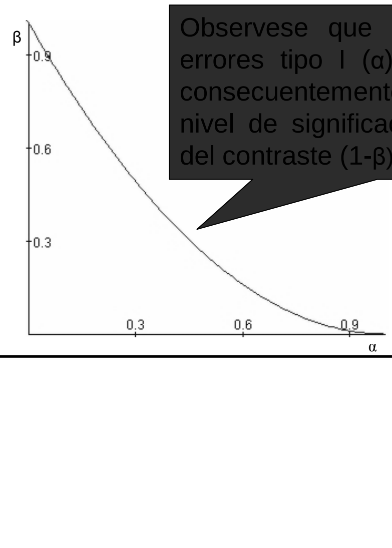
Dr. David Castilla Espino
CA UNED Huelva

Errores y potencia de un contraste

- **Solución (…)**:
  - Calculor del error tipo II (β):

$$\beta = P\left[x_1 > K / H_1(\theta = 2)\right] = \int_K^{+\infty} 2e^{-2x_1}\,dx = e^{-2K}$$

  de modo que β= $e^{2\ln(1-\alpha)}$ y su representación gráfica:



Observese que la relación entre los errores tipo I (α) y II(β) es inversa y consecuentemente la relación entre el nivel de significación (α) y la potencia del contraste (1-β) es directa.

Dr. David Castilla Espino
CA UNED Huelva

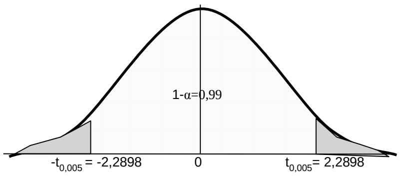
Fases de un contraste

- 1. Formular las hipótesis del contraste. Ej.:
  - H_0: $\mu = \mu_0 = 3$
  - H_1: $\mu \neq \mu_0 = 3$
- 2. Determinar el estadístico de prueba apropiado. Debe cumplir:
  - Función de probabilidad conocida dada H_0.
  - Debe contener el valor del parámetro contrastado.
  - Salvo el parámetro, el resto de terminos son conocidos o estimables.
  - Ej.: Sup. que la población es Normal: $t_{\exp} = \dfrac{\bar{x} - \mu_0}{S_c / \sqrt{n}} \to t_{n-1}$
- 3. Se selecciona el nivel de significación:
  - $\alpha = 1\%$.
- 4. Determinar la región crítica o región de rechazo.

$$P\left(\left|t_{17}\right| > t_{\alpha/2}\right) = 0{,}01 \Rightarrow t_{0,005} = 2{,}898 \Rightarrow t_{\exp} > 2{,}898 \text{ y } t_{\exp} < -2{,}898.$$

Fases de un contraste

- 5. Seleccionar aleatoriamente la muestra (n=18) y calcular el estadístico de prueba.
  - Ej.: $t_{\exp} = \dfrac{\bar{x} - \mu_0}{S_c / \sqrt{n}} = \dfrac{3{,}911 - 3}{0{,}04486 / \sqrt{18}} = 18{,}247$
- 6. Dar la regla de decisión y su interpretación:
  - Si $-2{,}898 \leq t_{\exp} \leq 2{,}898$ ➔ Acepta H_0
  - Si $t_{\exp} < -2{,}898$ ó $t_{\exp} > 2{,}898$ ➔ **Rechaza H_0**



Sumario

- Se han definido contraste de hipótesis estadísticas y se ha distinguido entre hipótesis simples y compuestas, así como nula y alternativa. Además se ha distinguido entre contrastes de hipótesis unilaterales y bilaterales.
- Se han introducido los distintos errores que se pueden cometer a la hora de realizar un cotrastes de hipótesis (α y β). Igualmente se ha introducido los conceptos de nivel de confianza y potencia de un contraste. Se ha ejemplificado el cálculo de estas magnitudes en distintos casos.
- Se han detallado y ejemplificado las fases de un contraste de hipótesis estadísticas.

Dr. David Castilla Espino
CA UNED Huelva