Inferencia Estadística

Testeo de hipótesis

Gabriel Martos Venturini
gmartos@utdt.edu

UTDT

UNIVERSIDAD TORCUATO DI TELLA

Outline

Outline

1. Introducción
   - Motivación y set–up general
   - Función de test y región de rechazo
2. Algunos test habitualmente utilizados en la práctica
3. Otros métodos para construir test y sus propiedades
4. Testeo múltiple

Outline

Motivación

- $\underline{X} \equiv \{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x;\theta)$ con $\theta \in \Theta$,
  - Estimador puntual (suficiencia, eficiencia, consistencia, etc).
  - Intervalos: Acotamos θ con garantías a "largo plazo".
- Testeo de Hipótesis: Nos preguntamos $\underbrace{\theta \in \Theta_0 \subset \Theta}_{\text{hipótesis estadística}}$?

Example (¿la moneda es justa?)

La variable aleatoria X asume el valor 1 si al tirar la moneda sale cara, y 0 en otro caso. Consideraremos $\{X_1, \ldots, X_n\} \overset{iid}{\sim} \text{Bern}(\theta)$ para construir una regla que nos permita decidir si $\theta \in \Theta_0 = \{\frac{1}{2}\}$ ó $\theta \in \Theta_0^C = [0, \frac{1}{2}) \cup (\frac{1}{2}, 1]$.

- Un test de hipótesis es un procedimiento estadístico (una regla de decisión) cuyo riesgo es cuantificable con el que para cada realización $\underline{X} = \underline{x}$ (datos) decido si la evidencia empírica avala $\theta \in \Theta_0$ ó $\theta \in \Theta_0^C$.

Set–up general

- Modelo estadístico $f(x;\theta)$, donde $\theta \in \Theta$ es desconocido.
  - En algunos casos no necesitamos especificar un modelo paramétrico.
- Espacio de hipótesis:
  - Hipótesis nula: $H_0 : \theta \in \Theta_0 \subset \Theta$.
  - Hipótesis alternativa: $H_1 : \theta \in \Theta_0^C = \Theta_1 \subset \Theta$.
  - Es decir que: $\Theta_0 \cap \Theta_1 = \emptyset$ y $\Theta_0 \cup \Theta_1 = \Theta$.
- Tipos de contrastes que vamos a considerar
  - **Unilateral** vs. **Unilateral**:
  $$H_0 : \theta \leq (\geq)\, \theta_0 \text{ vs. } H_1 : \theta > (<)\, \theta_0.$$
  - **Simple** vs. **Bilateral**:
  $$H_0 : \theta = \theta_0 \text{ vs. } H_1 : \theta \neq \theta_0.$$

Outline

Función de test

- Una vez planteadas H_0 y H_1, construiremos $\delta(\underline{X}) \in \{0, 1\}$ como regla para decidir si la evidencia empírica ($\underline{X} = \underline{x}$) está en favor de H_0 ó H_1.

$$\boxed{\text{Regla de decisión: } \delta(\underline{x}) = 1 \Rightarrow \text{ Rechazo } H_0.}$$

- En general las funciones de test dependen de la muestra aleatoria a través de un estadístico (de contraste) T_n y de la región de rechazo R:

$$\delta(T_n(\underline{X}), R) = \begin{cases} 1 \text{ si } T_n \in R \Rightarrow & \text{Rechazo } H_0, \\ 0 \text{ si } T_n \notin R \Rightarrow & \text{No rechazo } H_0. \end{cases}$$

- R es un subconjunto del rango de T_n, denominado **Región crítica o de Rechazo**. Podríamos reescribir la función de test como sigue:

$$\delta(T_n, R) = \mathbb{1}_R(T_n).$$

  - Ejemplo ilustrativo: Modelo normal y contraste para la media.

- ¿Cómo elegimos T_n y definimos R?

Función de test

- En general elegimos como estadísticos de contraste estimadores (consistentes) del parámetro que involucramos en la hipótesis nula.
- Idealmente deberíamos poder, al menos, aproximar la distribución de dicho estadístico para lograr construir una función de test razonable.
  - Ejemplo: Si estoy construyendo un test para $\theta \equiv E(X)$, resultará razonable utilizar como estadístico de contraste $\overline{X}_n$ (consistente para θ y aproximadamente normal si $n \gg 0$).
- Una vez identificado un estadístico de contraste apropiado para el problema en cuestión, para determinar R necesitamos analizar los tipos de errores que podemos cometer–analizar el riesgo de $\delta(T_n, R)$–.
- Para simplificar la notación, en adelante escribimos $\delta \equiv \delta(T_n, R)$, entendiendo que δ depende implícitamente tanto de T_n como de R.

Definiendo y midiendo errores

- Para determinar R tendremos que identificar errores:

$\downarrow$ Decisión / Estado $\longrightarrow$	$\theta \in \Theta_0\,(H_0)$	$\theta \in \Theta_1\,(H_1)$
No rechazo H_0	No hay error	Error tipo II
Rechazo H_0	Error tipo I	No hay error

Definiendo y midiendo errores

- Para determinar R tendremos que identificar errores:

$\downarrow$ Decisión / Estado $\longrightarrow$	$\theta \in \Theta_0\,(H_0)$	$\theta \in \Theta_1\,(H_1)$
No rechazo H_0	No hay error	Error tipo II
Rechazo H_0	Error tipo I	No hay error

- ... y medir la probabilidad de cometer dichos errores:

	$\theta \in \Theta_0\,(H_0)$	$\theta \in \Theta_1\,(H_1)$
No rechazo H_0	$P_{\theta\in\Theta_0}[T_n(\underline{X}) \notin R]$	$P_{\theta\in\Theta_1}[T_n(\underline{X}) \notin R]$
Rechazo H_0	$P_{\theta\in\Theta_0}[T_n(\underline{X}) \in R]$	$P_{\theta\in\Theta_1}[T_n(\underline{X}) \in R]$

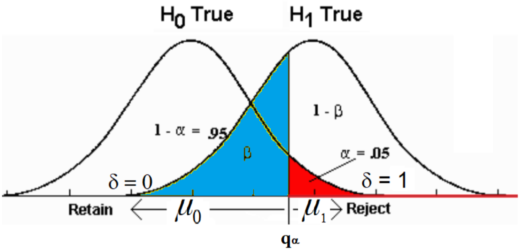
- La "función de riesgo del test" la podemos escribir como:

$$\text{Riesgo}(\delta,\theta) = \begin{cases} P_{\theta\in\Theta_0}(\delta = 1) \equiv \alpha(R,\theta), \\ P_{\theta\in\Theta_1}(\delta = 0) \equiv \beta(R,\theta). \end{cases}$$

¿Podemos elegir R para reducir ambas probabilidades?

- Ilustración: $X \sim N(\mu, \sigma_o^2)$ y $T_n = \overline{X}_n$ para el test:

$$H_0 : \mu = \mu_0 \text{ vs. } H_1 : \mu = \mu_1, \quad \text{con } \mu_1 > \mu_0 \text{ y } \alpha(R, \mu_0) = 0.05.$$



- Si $\uparrow q_\alpha$ ("reduzco" R) entonces $\downarrow \alpha$ y $\uparrow \beta$ (y viceversa).

- NO puedo reducir ambas probabilidades al mismo tiempo.

$$\text{Riesgo}(\delta, \mu) = \mathbb{1}_{\mu_0}(\mu)\alpha(R, \mu) + \mathbb{1}_{\mu_1}(\mu)\beta(R, \mu).$$

Trade–off: α vs β

- Decidimos testear $H_0 : \theta \in \Theta_0$ con $\delta(T_n, R)$.
- ¿Podemos manipular R de tal forma que reducimos la probabilidad de cometer uno de los errores manteniendo constante el otro? (rta: NO)
- Para disminuir $\alpha(R, \theta)$ cambiamos R por $\underline{R_* \subset R}$, luego:

$$\alpha(R_*, \theta) = P_{\theta \in \Theta_0}(T_n \in R_*) < P_{\theta \in \Theta_0}(T_n \in R) = \alpha(R, \theta).$$

- Pero si $R_* \subset R$ entonces $\underline{R_*^C \supset R^C}$, y por tanto:

$$\beta(R_*, \theta) = P_{\theta \in \Theta_1}(T_n \in R_*^C) > P_{\theta \in \Theta_1}(T_n \in R^C) = \beta(R, \theta).$$

Por tanto si cambiamos R por $R_* \subset R$ tal que $\alpha(R_*, \theta) < \alpha(R, \theta)$, entonces debe ocurrir necesariamente que $\beta(R_*, \theta) > \beta(R, \theta)$.

- ¿Cómo elegimos R?
  - Acotamos la probabilidad de cometer el ET I:

$$\text{Elijo } R \text{ tal que: } \sup_{\theta\in\Theta_0} P_\theta\Big(\underbrace{T_n \in R}_{\delta=1}\Big) \leq \alpha_0.$$

  - La **significatividad** α_0 del test es un parámetro elegido por el usuario.
- Por lo tanto, la probabilidad de cometer el error tipo II (NO rechazar H_0 cuando ésta es falsa), queda 'descontrolada' (no acotada).
- Cuando NO podemos rechazar H_0 corremos el riesgo de cometer el ET II. En este escenario conviene ser cautos respecto de la validez de H_0, ya que no controlamos (acotamos) la probabilidad de cometer el ETII.
- Ilustración del caso $X \sim N(\mu, \sigma_o^2)$ y $H_0 : \mu \leq \mu_0$.
  - Elegimos R tal que: $\sup_{\mu\in(-\infty,\mu_0]} P_{\mu\leq\mu_0}(\overline{X}_n \in R) \leq 0.05 = \alpha_0$.

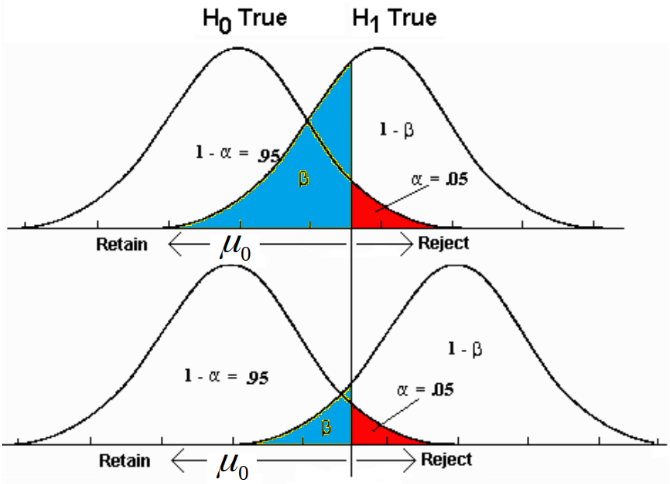


Figure: Esta R garantiza que la probabilidad de cometer el ET I es como máximo 0.05, pero no podemos decir nada respecto de la probabilidad de cometer ET II.

- **Potencia**$(\theta) = 1 - \beta(\theta)$, para $\theta \in \Theta_1$ (entendemos que R esta fija).

Sobre cómo definir H_0 y H_1 en la práctica

De acuerdo con la discusión anterior, deberíamos poner en H_0 aquella hipótesis cuyo **incorrecto rechazo genere el mayor perjuicio**.

- Discutamos un ejemplo:
  - Se sabe que cuando el nivel promedio de alcohol en sangre esta por encima de μ_0 el riesgo de accidente vehicular es elevado. Cuando en la vía pública se controla a los conductores deberíamos contrastar:

$$H_0 : \mu \leq \mu_0 \text{ ó } H_0 : \mu \geq \mu_0?$$

- No siempre es evidente que falso rechazo generaría mayor perjuicio.
- En la práctica el criterio utilitarista a seguir es poner en H_1 aquello que quiero demostrar (y lograr rechazar H_0 con los datos).
- Poner en H_0 aquello que suponemos es cierto o se cumple en general.
  - Aquella hipótesis que nadie cuestionaría si la tomamos como cierta.

Consecuencias del ET I

- $H_0 : \mu \leq \mu_0$ (el conductor está sobrio).
  - El incorrecto rechazo tiene consecuencias sobre el conductor: Paga una multa elevada, le quitan su licencia, y/o se queda un par de horas al costado del camino (hasta que no se pueda rechazar H_0).
- $H_0 : \mu \geq \mu_0$ (el conductor está ebrio).
  - El incorrecto rechazo tiene consecuencias colectivas potencialmente severas como accidentes de transito con costos de vidas o lesiones graves para el conductor, sus acompañantes o terceras personas.
  - Como sólo podemos controlar (acotar) con seguridad la probabilidad de cometer el ETI con el test, yo propondría ésta como hipótesis nula.
    - ★ Elegimos α_0 pequeño para tener una baja frecuencia de incorrectos rechazos (inexorablemente algunos falsos rechazos van a ocurrir).

Outline

Outline

Test Exactos vs Asintóticos

- **Exactos**: Valen para "*n* pequeño" porque conocemos la distribución exacta que tienen los estadísticos de contraste.
  - Ejemplo: Contraste para μ en población $N(\mu, \sigma_0^2)$.

$$\overline{X}_n \sim N(\mu, \sigma_0^2/n).$$

- **Asintóticos**: Basados en aproximaciones asintóticas (válidos para $n \gg 0$) de las distribuciones de los estadísticos de contraste.
  - Ejemplo: Contraste para $E(X)$ en población cuya distribución es desconocida (o que no podemos asumir como normal).

$$\overline{X}_n \sim_a N(E(X), \mathrm{Var}(X)/n).$$

$X \sim N(\mu, \sigma_o^2)$ (varianza conocida)

- Para testear en cada uno de los casos:

$$H_0\colon\ \mu \geq \mu_0 \qquad\qquad H_0\colon\ \mu = \mu_0 \qquad\qquad H_0\colon\ \mu \leq \mu_0$$
$$H_1\colon\ \mu < \mu_0 \qquad\qquad H_1\colon\ \mu \neq \mu_0 \qquad\qquad H_1\colon\ \mu > \mu_0$$

utilizamos el estadístico de contraste $\overline{X}_n \sim N(\mu, \sigma_o^2/n)$.

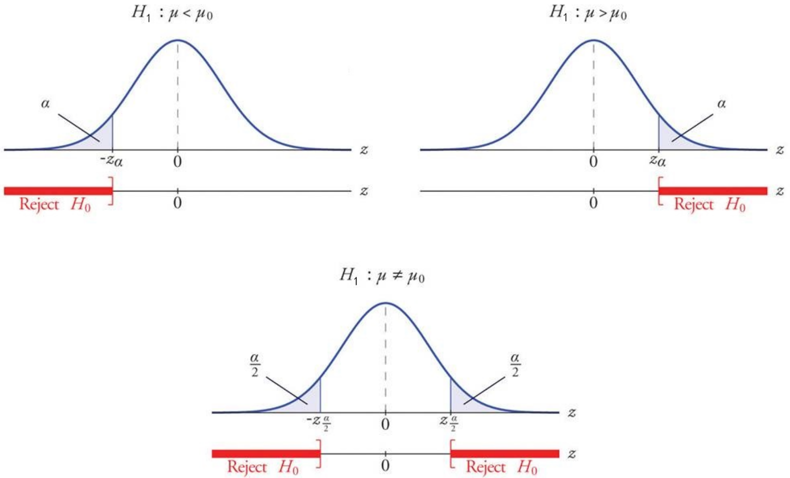
- Fijamos la signficatividad α_o y definimos las regiones de rechazo R:

$$\{\overline{X}_n : \overline{X}_n \leq \mu_0 - z_{\alpha_o}\frac{\sigma_o}{\sqrt{n}}\} \quad \{\overline{X}_n : |\overline{X}_n - \mu_0| \geq z_{\alpha_o/2}\frac{\sigma_o}{\sqrt{n}}\} \quad \{\overline{X}_n : \overline{X}_n \geq \mu_0 + z_{\alpha_o}\frac{\sigma_o}{\sqrt{n}}\}$$

- Garantizamos en los 3 casos que: $\sup_{\mu \in H_0} P[\overline{X}_n \in R] \leq \alpha_o$.

- Notación: De aquí en adelante utilizo $\alpha \equiv \alpha_0$ para referirme al nivel de significación del test y evitar α_0 que resulta engorroso al escribir R.

Las regiones de rechazo se pueden escribir en términos de $Z = \frac{\overline{X}_n - \mu_0}{\sigma_o/\sqrt{n}}$:



$$\{Z : \underbrace{\frac{\overline{X}_n - \mu_0}{\sigma_o/\sqrt{n}}}_{Z} \leq -z_\alpha\} \quad \{Z : \underbrace{\left|\frac{\overline{X}_n - \mu_0}{\sigma_o/\sqrt{n}}\right|}_{|Z|} \geq z_{\alpha/2}\} \quad \{Z : \underbrace{\frac{\overline{X}_n - \mu_0}{\sigma_o/\sqrt{n}}}_{Z} \geq z_\alpha\}$$

Probabilidad de cometer el ET II y Potencia

$$\beta(\mu) = P_{\mu \neq \mu_0}\left(\sqrt{n}(\mu_0 - \mu)/\sigma_o - z_{\alpha/2} \leq Z \leq \sqrt{n}(\mu_0 - \mu)/\sigma_o + z_{\alpha/2}\right)$$

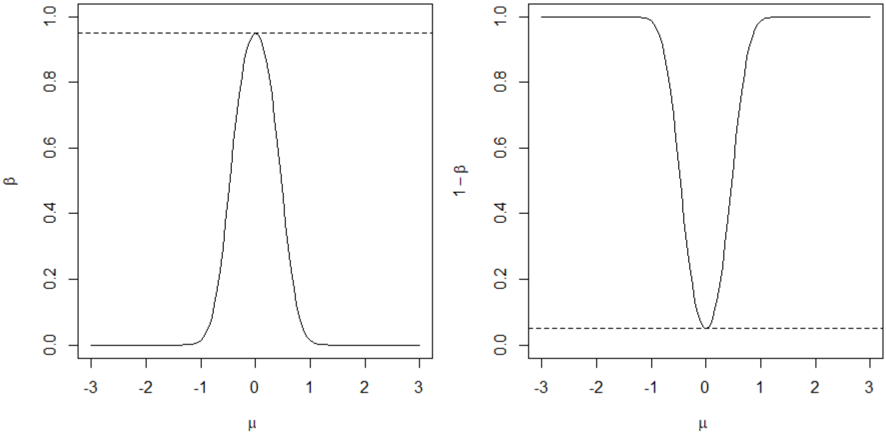


Figure: Ejemplo $H_0 : \mu = 0$, $n = 10$, y $\alpha = 0.05$.

Dualidad entre test e intervalos

- Consideremos el test:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

- Para una significatividad α, la región de rechazo será:

$$R_\alpha = \{\overline{X}_n : |\overline{X}_n - \mu_0| \geq z_{\alpha/2} \frac{\sigma_o}{\sqrt{n}}\}$$

- Por lo tanto la región de <u>no rechazo</u> R_α^C se puede escribir como:

$$R_\alpha^C = \{\overline{X}_n : \overline{X}_n - z_{\alpha/2} \frac{\sigma_o}{\sqrt{n}} \leq \mu_0 \leq \overline{X}_n + z_{\alpha/2} \frac{\sigma_o}{\sqrt{n}}\} \equiv \mathrm{IC}_{1-\alpha}(\mu).$$

- **Corolario**: Para un test de significatividad α, si μ_0 cae
  - Dentro del intervalo de confianza $1 - \alpha \Rightarrow$ no rechazamos H_0.
  - Fuera del intervalo de confianza $1 - \alpha \Rightarrow$ rechazamos H_0.

Ejercicio en clase

- El MECON pretende estimar el salario medio percibido por los trabajadores (para, por ejemplo, redefinir el nivel del salario mínimo).
- Se sabe (asume) que el salario mensual de los trabajadores argentinos W (expresado en unidades de 10mil) se distribuye $N(\mu_W, \sigma_W^2 = 20^2)$.
- Con una significatividad del 5% testear si el salario medio (μ_W) de la economía está por encima de los 40m pesos mensuales.
  - Al plantear el test asume que el MECON pretende demostrar que en efecto el salario medio está por arriba de los 40m.
  - Economistas del ministerio encuestaron a $n = 100$ trabajadores representativos del país (elegidos al azar) y estimaron que $\overline{w}_n = 44$m.
  - Cómo concluyes el test?
- Computa $\beta(\mu)$ y la función de potencia del test.

$$\begin{cases} H_0 : \mu_W \leq 40 \\ H_1 : \mu_W > 40 \end{cases}$$

- La región de rechazo (ver slide § 20) es:

$$\mathrm{R} = \{\overline{W}_n \;:\; \overline{W}_n \geq \mu_0 + z_\alpha \sigma_0/\sqrt{n} = 40 + 1.645 \times 20/\sqrt{100} = 43.29\}.$$

- En términos de $Z = 10(\overline{W}_n - 40)/20$, $\mathrm{R} = \{Z \;:\; Z \geq z_\alpha = 1.645\}$.
- Como $\bar{w}_n = 44 \in R$ (ó $z = 2 \in R$), luego rechazamos H_0.
- Con un nivel de significación del 5% concluimos que el salario medio de los argentinos es superior a los 40m pesos mensuales.

$$\beta(\mu) = P_{\mu \in \Theta_1}\Big[\frac{\overline{W}_n - \mu}{\sigma_o/\sqrt{n}} \leq \frac{\mu_0 - \mu}{\sigma_0/\sqrt{n}} + z_\alpha\Big] = P_{\mu > 40}\Big[Z \leq \frac{(40-\mu)}{2} + 1.645\Big].$$

Luego

$$1 - \beta(\mu) = P\big[Z \geq 21.645 - \mu/2\big].$$

Contraste para μ en población $N(\mu, \sigma^2)$

- Para cualquiera de los siguientes conjuntos de hipótesis:

$$H_0\colon\ \mu \geq \mu_0 \qquad H_0\colon\ \mu = \mu_0 \qquad H_0\colon\ \mu \leq \mu_0$$
$$H_1\colon\ \mu < \mu_0 \qquad H_1\colon\ \mu \neq \mu_0 \qquad H_1\colon\ \mu > \mu_0$$

- El estadístico de contraste es:

$$T_n = \frac{\overline{X}_n - \mu_0}{S_n/\sqrt{n}} \sim t_{n-1}$$

- Una vez elegimos α, las regiones de rechazo serán:

$$\{\overline{X}_n : \overline{X}_n \leq \mu_0 - t_\alpha \frac{S_n}{\sqrt{n}}\} \quad \{\overline{X}_n : |\overline{X}_n - \mu_0| \geq t_{\alpha/2} \frac{S_n}{\sqrt{n}}\} \quad \{\overline{X}_n : \overline{X}_n \geq \mu_0 + t_\alpha \frac{S_n}{\sqrt{n}}\}$$

- Test exacto (vale para n pequeño).

Durante el mes pasado, en farmacias el precio de un medicamento seguía una distribución normal de media \$1780. Este mes, en una muestra de 16 farmacias, se estimó que $\overline{x}_n = \$1900$ y $s_n = \$250$. ¿Aumentó el precio medio del medicamento?

1. En primer lugar planteamos el test:

$$\begin{cases} H_0 : \mu \leq \mu_0 = 1780 \\ H_1 : \mu > \mu_0 = 1780 \end{cases}$$

2. Elije α (el tamaño del test) y determina la región crítica.

$$\alpha = 0.05 \text{ y } t_{\alpha,15} = 1.753 \Rightarrow R_\alpha = \left\{ T_n = \frac{\overline{X}_n - \mu_0}{S_n/\sqrt{n}} > 1.753 \right\}$$

3. Con los datos de la muestra $T_n = \frac{1900-1780}{250/\sqrt{16}} = t_{obs} = 1.92$.

4. Rechazo H_0: Con $\alpha = 0.05$, hay evidencia de aumento precio medio.

- En vez de reportar la decisión sobre el test, forzando a quién lee los resultados a tomar el mismo criterio respecto del valor de significación que escogimos, podemos reportar el *p*–valor que se define como:

Definition

El *p*-**valor** es el valor de significación más pequeño con el cual la **hipótesis nula puede ser rechazada** dados los datos de la muestra.

- El *p*-valor cambia cuando cambian los datos.
- Podríamos pensar en el p–valor como el resultado de:

$$\min \alpha \qquad \text{s.a: } \delta(\text{datos}) = 1.$$

- Ej. ant: Vamos $\downarrow \alpha$ (*corriendo a la derecha* R) hasta que no podemos seguir achicando R sin dejar de rechazar H_0 (dado $\overline{x}_n = \$1900$).

Regla para calcular el p-valor en la práctica

Si T_n es el estadístico de contraste del test, y con los datos de la muestra se tiene que el valor del estadístico de contraste es de t_{obs}, luego:

- Si $H_0 : \theta \leq \theta_0$, entonces: p–val $= P\big(T_n \geq t_{obs}\big)$.
- Si $H_0 : \theta \geq \theta_0$, entonces: p–val $= P\big(T_n \leq t_{obs}\big)$.
- Si $H_0 : \theta = \theta_0$, entonces: p–val $= 2\min\{P\big(T_n \leq t_{obs}\big), T_n \geq t_{obs}\big)\}$.
- RECHAZAS H_0 cuando:
  - El p–valor $\leq \alpha$ (significación que eliges para el contaste).

Remark: Notar que la cantidad t_{obs} depende de los datos. Por lo tanto, si cambian los datos con los que realizo el test, es altamente plausible que se modifique la cantidad t_{obs} y con ello el p–valor.

- Neyman, Pearson y Fisher (3 estadísticos del siglo XX) generaron una disputa encarnada y hostil en la historia de la estadística por su desacuerdo en definir un procedimiento para contrastar hipótesis.
- Neyman y Pearson pensaban el problema como la elección entre dos hipótesis específicas (y contrapuestas) H_0 vs H_1. Para cada problema, desarrollaron una regla de decisión que funcionara razonablemente bien en repetidas aplicaciones de la regla.
- Para Fisher, en cambio, el problema era desafiar una hipótesis nula H_0 sin tener en mente ninguna alternativa concreta. La evidencia en contra de H_0 que se recogía de los datos se resumía en un numero al que llamó p-valor, pero no se establecía ninguna regla de decisión (cada persona podía reaccionar distinto frente al mismo p-valor).
- Veamos en un ejemplo las consecuencias de los diferencias enfoques.

Una situación hipotética...

Una consultora estudia el tiempo promedio diario que dedican los habitantes de la ciudad a utilizar plataformas de streamming (parámetro μ). La consultora recomendará utilizar como medio de publicidad dichas plataformas si $\mu > 1$. Para los datos de una muestra se alcanza a rechazar $H_0 : \mu \leq 1$ con un nivel de significativad de $\alpha = 0.05$, sin embargo NO se puede rechazar H_0 si $\alpha < 0.05$ (por ejemplo $\alpha = 0.0499$).

- Es un poco absurdo que la recomendación de la consultora sea distinta para $\alpha = 0.05$ que para $\alpha = 0.0499$.
- Para evitar este tipo de situaciones, en vez de reportar una decisión específica (que depende de un α concreto) es recomendable indicar el p–valor y que quién recibe esta información decida si esta cantidad es evidencia suficiente o no para decidir si alcanza para rechazar H_0.

Algunas críticas al rol central del p–valor en inferencia

- Criticas al excesivo peso que tiene esta regla de decisión sobre otros factores que resultan igual o más relevantes en el análisis de los datos.
- P values are just the tip of the iceberg.
- Calidad de las mediciones (datos atípicos).
- Validación del modelo subyacente bajo H_0.
- Mi consejo es que siempre seas bastante cauto respecto a como concluyes tus análisis.
- Check de estudios similares para contrastar resultados con evidencia relacionada.
  - Meta–análisis: p–curve and p–Hacking.

Contraste para σ^2 en población normal

- Sea $\{X_1, \ldots, X_n\} \overset{iid}{\sim} N(\mu, \sigma^2)$.
- Para cualquiera de los siguientes conjuntos de hipótesis:

$$H_0\colon\ \sigma^2 = \sigma_0^2 \qquad H_0\colon\ \sigma^2 \geq \sigma_0^2 \qquad H_0\colon\ \sigma^2 \leq \sigma_0^2$$
$$H_1\colon\ \sigma^2 \neq \sigma_0^2 \qquad H_1\colon\ \sigma^2 < \sigma_0^2 \qquad H_1\colon\ \sigma^2 > \sigma_0^2$$

- El estadístico de contraste es:

$$T_n \equiv \frac{(n-1)S_n^2}{\sigma_0^2} \sim \chi^2_{n-1} \text{ (chi–cuadrada con n-1 grad. libertad).}$$

- Para garantizar que la probabilidad de cometer el ET I está acotada por α, definimos las siguientes regiones de rechazo R:

$$\{S_n^2 : T_n \in [0, c_{1-\alpha/2}] \cup [c_{\alpha/2}, \infty)\} \quad \{S_n^2 : T_n \in [0, c_{1-\alpha}]\} \quad \{S_n^2 : T_n \in [c_\alpha, \infty)\}$$

- Donde c_α es la constante que verifica $P(\chi^2_{n-1} > c_\alpha) = \alpha$.

- Una ONG financió un proyecto de potabilización de agua y para certificar la potabilidad, se tomaron muestras durante $n = 30$ días consecutivos. Con estos datos se estimó que la varianza del nivel de impurezas presente por cada ml^3 de agua es $s_n^2 = 2.62(p/ml^3)^2$.
- Los niveles promedio diarios observados de contaminación son acordes a los requerimientos del proyecto; pero se requirió demostrar también que la varianza del nivel de impurezas no exceda las $4(p/ml^3)^2$ para terminar de certificar el proyecto.
- Asumiendo que la cantidad de partículas por mililitro cúbico sigue una distribución normal, plantee el test correspondiente e interprete los resultados. Asuma que $\alpha = 0.1$ es un valor razonable para este test.

Solución

- Supongamos que hicimos el proyecto (queremos probar $\sigma^2 < 4$):

$$\begin{cases} H_0 : \sigma^2 \geq 4 \\ H_1 : \sigma^2 < 4 \end{cases} \quad \text{y } R_\alpha = \left\{ S_n^2 : \frac{(n-1)S_n^2}{\sigma_0^2} \leq q_{1-\alpha} \right\}.$$

- Como $\alpha = 0.1$, luego $q_{1-\alpha} = 19.767$ y $R_\alpha = \{T_n \leq 19.767\}$.
- Valor observado de $T_n = (29)(2.62)/4 = 18.995$ cae en R_α.
- Para un $\alpha = 0.1$ hay evidencia para afirmar que la variabilidad en las cantidades de impurezas en el agua cumplen con los estándares.
- El p-val se calcula haciendo: $P(T_n \leq t_{obs} = 18.995) \approx 0.078$ (si hubieras elegido un $\alpha = 0.05$, no alcanzarías a rechazar H_0).

Contraste asintótico de localización

- Sea $\{X_1, \ldots, X_n\} \overset{iid}{\sim} X$ con $V(X) < \infty$.
- Para cualquiera de los siguientes conjuntos de hipótesis:

$$\begin{array}{lll} H_0\colon\ E(X) \geq \mu_0 & \qquad H_0\colon\ E(X) = \mu_0 & \qquad H_0\colon\ E(X) \leq \mu_0 \\ H_1\colon\ E(X) < \mu_0 & \qquad H_1\colon\ E(X) \neq \mu_0 & \qquad H_1\colon\ E(X) > \mu_0 \end{array}$$

- El estadístico de contraste es: $T_n = \frac{\overline{X}_n - \mu_0}{S_n/\sqrt{n}} \overset{n \gg 0}{\sim_a} N(0,1)$, y R_α:

$$\{\overline{X}_n : \overline{X}_n \leq \mu_0 - z_\alpha \frac{S_n}{\sqrt{n}}\} \quad \{\overline{X}_n : |\overline{X}_n - \mu_0| \geq z_{\alpha/2} \frac{S_n}{\sqrt{n}}\} \quad \{\overline{X}_n : \overline{X}_n \geq \mu_0 + z_\alpha \frac{S_n}{\sqrt{n}}\}$$

- Cuando $n \gg 0$: $\sup_{\theta \in \Theta_0} P_\theta(\delta = 1) \approx \alpha$ (asintótico).
- Si $\{X_1, \ldots, X_n\} \overset{iid}{\sim} \text{Bern}(p)$: $E(X) = p$, $V(X) = p(1-p)$.

$$T_n = \frac{\widehat{p}_n - p_0}{\sqrt{\widehat{p}_n(1 - \widehat{p}_n)/n}} \overset{n \gg 0}{\sim_a} N(0,1).$$

Caso estudio: Encuesta de Synopsis 2018 sobre la ILE

Según una encuesta de Synopsis (Marzo 2018) realizada sobre 1485 adultos (a nivel nacional y teniendo en cuenta estratos socio–económicos), el 53.3% de los entrevistados estuvieron a favor de la promulgación de la ILE (ley de interrupción legal del embarazo); mientras que el 32.6% de los encuestados reveló no estar de acuerdo y el 13.8% restante NS/NC.

- Con una significación (aproximada) del 1%, ¿Se podría decir que la mayoría de los Argentinos está (en Marzo del 2018) a favor de una ley que eventualmente despenalizare la interrupción voluntaria del embarazo?

Solución

- Los datos sugieren una pequeña ventaja en favor de ILE:

$$\begin{cases} H_0 : p \leq 0.5 \\ H_1 : p > 0.5 \end{cases}$$

- Con $\alpha = 0.01$, la región (asintótica) de rechazo será entonces:

$$R = \{\widehat{p} : T_n = \frac{\widehat{p}_n - p_0}{\sqrt{\widehat{p}_n(1-\widehat{p}_n)/n}} \geq z_{0.01} = 2.32\}.$$

- Calculamos el valor realizado del estadístico de contraste:

$$t_{obs} = \frac{0.533 - 0.500}{\sqrt{\frac{0.533\times(1-0.533)}{1485}}} = 2.548 \in R$$

- Con una significatividad (aproximada) del 1% se rechaza la hipótesis nula. Es decir, hay evidencia estadística a favor de un apoyo mayoritario de una ley de despenalización del aborto (marzo de 2018).

- ¿Cómo calculamos el p-valor?

Outline

De localización en poblaciones independientes (asintótico)

- Queremos testear diferencias de medias en dos poblaciones no relacionadas y que no podemos asumir normales:

$$\begin{array}{lll} H_0\colon\ E(X)-E(Y)\geq c & H_0\colon\ E(X)-E(Y)=c & H_0\colon\ E(X)-E(Y)\leq c \\ H_1\colon\ E(X)-E(Y)<c & H_1\colon\ E(X)-E(Y)\neq c & H_1\colon\ E(X)-E(Y)>c \end{array}$$

- Consideramos $\{X_1,\ldots,X_n\} \overset{iid}{\sim} X$ y $\{Y_1,\ldots,Y_m\} \overset{iid}{\sim} Y$:

$$T_{n,m}=\frac{(\overline{X}_n-\overline{Y}_m)-c}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}} \overset{(H_0)}{\sim_a} N(0,1)$$

- Regiones de rechazo:

$$\{T_{n,m}: T_{n,m}\leq z_{1-\alpha}\} \quad \{T_{n,m}: |T_{n,m}|\geq z_{\alpha/2}\} \quad \{T_{n,m}: T_{n,m}\geq z_{\alpha}\}$$

- Estas regiones de rechazo garantizan que asintóticamente la probabilidad de cometer el Error Tipo I esta acotada por α.

Caso particular de lo anterior

- Sean $\{X_1, \ldots, X_n\} \sim \text{Bern}(p_X)$ y $\{Y_1, \ldots, Y_m\} \sim \text{Bern}(p_Y)$.

$$T_{n,m} = \frac{(\widehat{p}_X - \widehat{p}_Y) - c}{\sqrt{\frac{\widehat{p}_X \times (1 - \widehat{p}_X)}{n} + \frac{\widehat{p}_Y \times (1 - \widehat{p}_Y)}{m}}} \overset{(H_0)}{\sim_a} N(0,1)$$

- Último ejercicio práctico de la G4.

Contraste asintóticos de localización muestras apareadas

- Consideramos $\{D_1 = X_1 - Y_1, \ldots, D_n = X_n - Y_n\} \overset{iid}{\sim} (X, Y)$.
- En general (X, Y) representan mediciones sobre un mismo individuo en dos momentos del tiempo (ej: antes y después de un tratamiento).
- Nos interesan testear sobre el parámetro $E(D) = E(X - Y) = \mu_D$:

$$\begin{array}{lll} H_0\colon\ \mu_D \geq c & H_0\colon\ \mu_D = c & H_0\colon\ \mu_D \leq c \\ H_1\colon\ \mu_D < c & H_1\colon\ \mu_D \neq c & H_1\colon\ \mu_D > c \end{array}$$

- Estadístico de contraste para estos test será:

$$T_n = \frac{\overline{D}_n - c}{S_D/\sqrt{n}} \overset{(H_0)}{\sim_a} N(0, 1)$$

- Regiones de rechazo:

$$\{T_n : T_n \leq z_{1-\alpha}\} \quad \{T_n : |T_n| \geq z_{\alpha/2}\} \quad \{T_n : T_n \geq z_\alpha\}$$

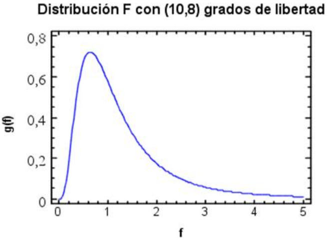
- En muestras pequeñas utilizar t–Student (y chequear normalidad).

Prueba para comparar varianzas (test exacto)

- $\{X_1, \ldots, X_n\} \overset{iid}{\sim} \mathrm{N}(\mu_X, \sigma_X^2)$ y $\{Y_1, \ldots, Y_m\} \overset{iid}{\sim} \mathrm{N}(\mu_Y, \sigma_Y^2)$

$$H_0\colon \sigma_X^2 = \sigma_Y^2 \qquad H_0\colon \sigma_X^2 \geq \sigma_Y^2 \qquad H_0\colon \sigma_X^2 \leq \sigma_Y^2$$
$$H_1\colon \sigma_X^2 \neq \sigma_Y^2 \qquad H_1\colon \sigma_X^2 < \sigma_Y^2 \qquad H_1\colon \sigma_X^2 > \sigma_Y^2$$

- Estadístico de contraste: $F_{n,m} = \frac{S_X^2}{S_Y^2} \sim F(n-1, m-1)$.



- Regiones de rechazo:

$$\{F_{n,m} : F_{n,m} \in [q_{1-\alpha/2}, q_{\alpha/2}]^C\} \quad \{F_{n,m} : F_{n,m} \in [0, q_{1-\alpha}]\} \quad \{F_{n,m} : F_{n,m} \in [q_\alpha, \infty)\}$$

donde q_α: $P(F(n-1, m-1) > q_\alpha) = \alpha$.

Ejemplo

Se desea comparar la variabilidad en dos métodos de engorde de novillos basados en las cebada forrajeras A y B. Con ese fin se seleccionó un lote de 30 animales lo más parejo posible, dividiéndolo en dos grupos de 15 animales cada uno. Los animales pastorearon durante el mismo lapso de tiempo en potreros (idénticos) sembrados cada uno con dichas variedades. Se registraron los aumentos de peso de todos los novillos, información que, luego de su procesamiento, arrojó los siguientes resultados:

$$\begin{array}{rcc} \text{Incremento promedio de pesos: } & \overline{X}_A = 25\text{kg} & \overline{X}_B = 26,2\text{kg} \\ \text{Variabilidad registrada en el incremento: } & S_A = 3,4\text{kg} & S_B = 2,8\text{kg} \end{array}$$

- Con una significatividad del 5%, ¿se puede asegurar que la variabilidad en los dos métodos de engorde son diferentes?

Solución

- Planteemos un test bilateral para intentar demostrar que los dos métodos tienen de engorde tienen varianzas distintas:

$$\begin{cases} H_0 : \sigma_A^2 = \sigma_B^2 \Rightarrow \sigma_A^2/\sigma_B^2 = 1 \\ H_1 : \sigma_A^2 \neq \sigma_B^2 \Rightarrow \sigma_A^2/\sigma_B^2 \neq 1. \end{cases}$$

- Con $\alpha = 0.05$, la región de rechazo será entonces:

$$R_\alpha = \{F \leq 0.335 \cup F \geq 2.978\}.$$

- Comando $qf(\alpha/2, n_X - 1, n_Y - 1)$ en R para cuantiles.
- El valor realizado del estadístico de contraste es:

$$F = 3.4^2/2.8^2 = 1.47 \notin R_\alpha.$$

- A un nivel de significatividad del 5% NO alcanzamos a rechazar la hipótesis nula: NO hay evidencia estadística en favor de que en la varianza de los dos métodos de engorde sea diferente.

Outline

Outline

- Dada $\{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x;\theta)$, consideremos el siguiente contraste

$$\begin{cases} H_0 : \theta = \theta_0, \\ H_1 : \theta \neq \theta_0. \end{cases}$$

- Los EMV son asintóticamente normales:

$$\sqrt{n}(\hat{\theta}_n - \theta) \to_F N(0, i(\theta)^{-1}).$$

- De la normalidad asintótica de los EMV (+ Slutsky):

$$T_n = \frac{\widehat{\theta}_n - \theta}{\widehat{\text{se}}} \overset{n \gg 0}{\sim_a} N(0,1), \text{ donde } \widehat{\text{se}} = \left[ni(\widehat{\theta}_n)\right]^{-1/2}.$$

- La región de rechazo del test de Wald de significancia asintótica α se define como: $R_\alpha = \{\widehat{\theta}_n : |(\widehat{\theta}_n - \theta_0)/\widehat{\text{se}}| \geq z_{\alpha/2}\}$.

- Ejemplo: Parámetro λ en modelo Poisson.

- La función de potencia (aproximada) del test:

$$1-\beta(\theta) = 1-\Phi\left(\frac{\theta_0-\theta}{\widehat{\mathsf{se}}}+z_{\alpha/2}\right)+\Phi\left(\frac{\theta_0-\theta}{\widehat{\mathsf{se}}}-z_{\alpha/2}\right), \quad \theta\in\Theta_1.$$

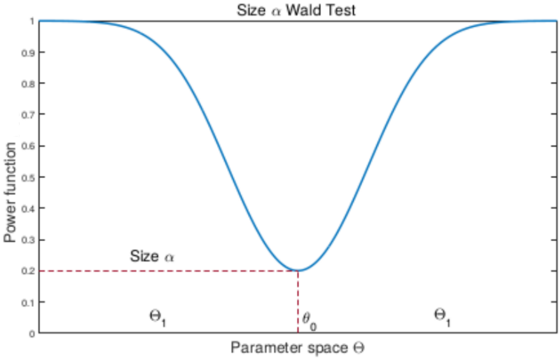


Figure: Curva de potencia para el test de tamaño $\alpha = 0.2$.

  - Como $\downarrow \widehat{\mathsf{se}}$ cuando $\uparrow n$: La potencia es alta cuando n es grande.

Consideraciones finales sobre el test Wald

- Recordar que es un test asintótico (se usa cuando $n \gg 0$).
- Relación con los Intervalos de Confianza de Wald:
  - EMV: Un intervalo con nivel de confianza asintótico $1-\alpha$ para θ tiene la siguiente estructura:

$$\mathrm{IC}_{1-\alpha}(\theta) : \left[\widehat{\theta}_n - z_{\alpha/2}\widehat{\mathrm{se}},\, \widehat{\theta}_n + z_{\alpha/2}\widehat{\mathrm{se}}\right].$$

  - Luego rechazamos $H_0 : \theta = \theta_0 \iff \theta_0 \notin \mathrm{IC}_{\alpha}(\theta)$ (dualidad).

Outline

- $\underline{X} \equiv \{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x;\theta)$ y $L(\theta|\underline{X}) = \Pi_{i=1}^{n} f(X_i;\theta)$.
- Para $H_0 : \theta \in \Theta_0 \subset \Theta$ consideramos el estadístico de contraste

$$\Lambda_n(\underline{X}) \equiv \Lambda_n = \frac{\sup_{\theta\in\Theta_0} L(\theta|\underline{X})}{\sup_{\theta\in\Theta} L(\theta|\underline{X})}$$

- La región de rechazo tiene la forma $R = \{\Lambda_n \leq c\}$, donde $0 \leq c \leq 1$.
- Fijamos α y elegimos c_α tal que $\sup_{\theta\in\Theta_0} P_\theta(\Lambda_n \leq c_\alpha) \leq \alpha$.
- En la práctica suele ocurrir que resulta imposible (o demasiado complicado) determinar la distribución exacta de Λ_n.
- Ejemplo: Hipótesis unilateral en el modelo Poisson.
- Utilizamos aproximaciones asintóticas de la distribución de Λ_n para determinar c (valen para los modelos de la familia exponencial).

Aproximaciones asintóticas de la distribución de Λ_n

- Llamemos $\widehat{\theta}_0 \equiv \arg\max L(\theta|\underline{X})$ sa $\theta \in \Theta_0$ y $\widehat{\theta}_n \equiv \arg\max L(\theta|\underline{X})$.
- Bajo *condiciones de regularidad* sobre el modelo estadístico:

$$-2\ln\Lambda_n = -2(\ell(\widehat{\theta}_0|\underline{X}) - \ell(\widehat{\theta}_n|\underline{X})) \to_F \chi^2_{(1)}$$

- Cuando $n \gg 0$, la región de rechazo tiene la siguiente forma:

$$R_\alpha = \{\underline{X} : \ -2\ln\Lambda_n(\underline{X}) \geq q_\alpha\}, \text{ donde } q_\alpha \text{ cuantíl } \alpha \text{ de una } \chi^2_{(1)}.$$

- Este contraste tiene una significancia asintótica α:

$$\lim_{n\to\infty} \underbrace{\sup_{\theta\in\Theta_0}}_{\text{Bajo } H_0} P_\theta\big(\underbrace{-2\ln\Lambda_n \geq q_\alpha}_{\text{Rechazar } H_0}\big) \leq \alpha.$$

- El resultado vale para contrastes *multivariantes* (los grados de libertad de la χ^2 con la que aproximamos la distribución de Λ_n dependen de $\dim(\Theta)$ y de como configuramos H_0).

Resumen LRT ($n \gg 0$)

1. Especifica tu hipótesis nula $H_0 : \boldsymbol{\theta} \in \Theta_0 \subset \Theta$ y elige α.
   - $r = \dim(\Theta)$.
     - Caso más general: $r = \dim(\Theta)$ - parámetros libres bajo H_0.
   - Calcula q_α, el cuantíl α de una $\chi^2_{(r)}$.
2. Con los datos ($\underline{X} = \underline{x}$) computa $\widehat{\boldsymbol{\theta}}_0$ (EMV–restringido) y $\widehat{\boldsymbol{\theta}}_n$ (EMV).
3. Calcular: $\lambda_n(\underline{x}) = \dfrac{L(\widehat{\boldsymbol{\theta}}_0|\underline{x})}{L(\widehat{\boldsymbol{\theta}}|\underline{x})}$ (realización de Λ_n).
   - Si $-2\ln\lambda_n(\underline{x}) \geq q_\alpha \Rightarrow$ rechazar H_0.
   - Caso contrario, retener H_0.

Outline

Optimalidad

- $\underline{X} \equiv \{X_1, \ldots, X_n\} \overset{iid}{\sim} f(x;\theta)$ y $H_0 : \theta \in \Theta_0$ (vs $H_1 : \theta \in \Theta_1$).
- Fijemos $\alpha \in (0,1)$ y definamos:

$$\mathcal{D}(\Theta_0, \alpha) := \left\{ \delta : \underline{X} \to \{0,1\} \,\middle|\, \sup_{\theta \in \Theta_0} P_\theta[\delta = 1] \leq \alpha \right\}.$$

- **Optimalidad**: Un test δ para H_0 vs H_1 es óptimo a nivel α si
  1- $\delta \in \mathcal{D}(\Theta_0, \alpha)$.
  2- Para todo $\delta' \in \mathcal{D}(\Theta_0, \alpha)$ y $\theta \in \Theta_1$: $P_\theta[\delta = 0] \leq P_\theta[\delta' = 0]$.
- Un test es óptimo si garantiza que como mucho comete un error tipo I con probabilidad α, y al mismo tiempo tiene la menor probabilidad de cometer el error tipo II para cualquier valor de $\theta \in \Theta_1$.
- La existencia de estos test depende de la estructura de Θ_0 y Θ_1, así como del modelo de probabilidad subyacente para los datos.

1. $\underline{X} \equiv (X_1, \ldots, X_n) \overset{iid}{\sim} f(x;\theta)$ y $H_0 : \theta = \theta_0$ vs $H_1 : \theta = \theta_1$.

$$\Lambda_n = \frac{L(\theta_0|\underline{X})}{L(\theta_1|\underline{X})}$$

- Si podemos encontrar una constante c tal que $P_{\theta_0}[\Lambda_n \leq c] = \alpha$, el test de ratio de verosimilitud es óptimo (tiene la potencia más alta).

- En familias exponenciales, el test se construye a partir del estadístico suficiente como veremos en el siguiente ejemplo: $N(\theta, \sigma_*^2 = 1)$.

$$c = e^{n(\mu_0 - \mu_1) z_\alpha - n(\mu_0^2 - \mu_1^2)/2} \text{ (asumiendo que } \mu_0 < \mu_1\text{)}.$$

- Si X sigue un modelo discreto, puede no existir un test óptimo.
  - Contra ejemplo: Poisson.

2. $f(x;\theta)$ en familia exponencial y $H_0 : \theta \leq (\geq) \theta_0$ vs $H_1 : \theta > (<) \theta_1$.

- Ejemplos y estructura general del test en VP §4.3.2.

Outline

El problema

- Consideremos testear en simultáneo $H_0^1, \ldots, H_0^m$ (independencia).
- Si cada test tiene una significatividad α:

$$
\begin{aligned}
\text{FWER}^1 = P(\text{al menos 1 ET I}) &= 1 - P(\text{Ningún ET I}) \\
&\overset{ind}{=} 1 - \Pi_{i=1}^{m}(1 - P(\text{ET I en test } i)) \\
&= 1 - (1-\alpha)^m
\end{aligned}
$$

- Si m es grande la probabilidad de hacer al menos un falso rechazo es virtualmente 1 (FWER $\approx$ 1). ¿Cómo evitar este contratiempo?
- Bonferroni: Significatividad individual de $\frac{\alpha}{m} \to$ FWER $\approx \alpha$.
- Procedimiento conservador: $m \gg 0$ no logramos rechazar ninguna H_0.

[1] Family-Wise Error Rate

Benjamini–Hochberg

- $R = R_f + R_v$ (el total de rechazos es igual a la suma de los falsos rechazos más los verdaderos rechazos). Solo R es observable.

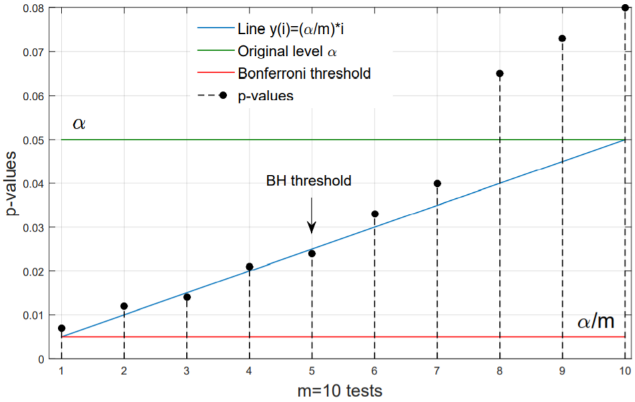
$$\text{False Discovery Rate } = E(R_f/R)$$

- BH: Método general para acotar la FDR$\leq \alpha$ (vos elegís α).
- Consideramos los p-valores ordenados de los múltiples tests $p_{(1)} \leq \cdots \leq p_{(m)}$, que renombramos como $H_0^{(1)}, \ldots, H_0^{(m)}$. Luego:

$$i^* = \max\left\{ i = 1, \ldots, m : p_{(i)} \leq \frac{\alpha}{m} \frac{i}{\beta_m} \right\}$$

- Rechazando $H_0^{(1)}, \ldots, H_0^{(i^*)}$ el método garantiza que FDR$\leq \alpha$.
- $\beta_m = 1$ si los test son independientes y $\beta_m = \sum_{i=1}^{m} 1/i$ en otro caso.

- Ejemplo: $p_{(1)} = 0.007, p_{(2)} = 0.012, p_{(3)} = 0.014, p_{(4)} = 0.021, p_{(5)} = 0.024, p_{(6)} = 0.033, p_{(7)} = 0.04, p_{(8)} = 0.065, p_{(9)} = 0.073, p_{(10)} = 0.08.$



- Benjamini and Hochberg (1995) "*Controlling the false discovery rate: a practical and powerful approach to multiple testing*," Journal of the Royal Statistical Society. Series B.