Cap.8. Test de Hipótesis Estadísticas

Estadística Computacional ILI 280 - San Joaquín

Prof. Ricardo Ñanculef Alegría
jnancu@inf.utfsm.cl

Departamento de Informática UTFSM
Campus San Joaquín - II 2015

Departamento de Informática
Universidad Técnica Federico Santa María

Outline

I. Conceptos Base

Conceptos Base

Hipótesis

Afirmación acerca de una característica desconocida (parámetro) de la población o fenómeno que se esta estudiando. Más precisamente, una afirmación acerca de una característica de la v.a. estudiada de la población o fenómeno.

Ejemplo

"La <u>concentración media</u> de sodio en botellas de agua mineral de la marca A (μ_1) es menor que en botellas de la marca B (μ_2)". Anotamos:

$$\mathbf{H_0}: \ \mu_1 < \mu_2$$

Ejemplo

"La <u>proporción</u> de personas que votarán por el candidato A es mayor que la proporción de personas que votarán por el candidato B (p_2) ". Anotamos:

$$\mathbf{H_0}: \ p_1 > p_2$$

Conceptos Base

Test o Contraste de Hipótesis

Procedimiento para decidir si la evidencia experimental sostiene la hipótesis o la contradice. Dos decisiones posibles

1. **Se rechaza la hipótesis:** Los datos experimentales muestran que es muy improbable que H_0 sea cierta y por lo tanto dejamos de creer en ella.
2. **No se rechaza la hipótesis:** No hay información suficiente para rechazar H_0, y por lo tanto podemos/debemos continuar creyendo en ella.

Conceptos Base

Test o Contraste de Hipótesis

Todo contraste consta de dos hipótesis: H_0, la hipótesis nula y H_1, la hipótesis alternativa.

Hipótesis Alternativa (H_1)

Hipótesis que sustituirá la hipótesis Nula H_0 cuando ésta es rechazada.

Ejemplo

- Variable Aleatoria X: concentración de plomo en las botellas de agua mineral vendidas por cierta empresa.
- Parámetro. μ: valor esperado (media poblacional).
- Hipótesis

$$\mathbf{H_0}: \ \mu = 0$$
$$\mathbf{H_1}: \ \mu \neq 0$$

Conceptos Base

Test o Contraste de Hipótesis

Procedimiento para decidir si la evidencia experimental sostiene la hipótesis o la contradice. Dos decisiones posibles

1. **Se rechaza la hipótesis:** Los datos experimentales muestran que es muy improbable que H_0 sea cierta y por lo tanto dejamos de creer en ella.
2. **No se rechaza la hipótesis:** No hay información suficiente para rechazar H_0, y por lo tanto podemos/debemos continuar creyendo en ella.

- Analogía: Tal como en un juicio, una persona es inocente hasta que se demuestra lo contrario, en un contraste, la hipótesis nula es **verdadera hasta que no se pruebe lo contrario con abundante evidencia**.

Conceptos Base

¿Cómo formular la Hipótesis Nula y Alternativa?

- H_0 representa aquello que se cree actualmente y que debiésemos creer si no podemos demostrar lo contrario.
- H_1 representa la contradicción de H_0 y con frecuencia aquello que nos interesa demostrar.

Ejemplo

- Variable Aleatoria X: concentración de plomo en las botellas de agua mineral vendidas por cierta empresa.
- Parámetro. μ: valor esperado (media poblacional).
- Hipótesis

$$\mathbf{H_0}: \ \mu = 0$$
$$\mathbf{H_1}: \ \mu \neq 0$$

Conceptos Base

¿Cómo formular la Hipótesis Nula y Alternativa?

- Definir una variable aleatoria X (o más de una) que corresponde a aquello que podemos medir de la población.
- Identificar un parámetro θ que representa la característica de X sobre la que queremos declarar algo.
- Formular las hipótesis como afirmaciones acerca de los valores de θ.
- H_0 representa aquello que se cree actualmente y que debiésemos creer si no podemos demostrar lo contrario.

Ejemplo

- Variable Aleatoria X: concentración de plomo en las botellas de agua mineral vendidas por cierta empresa.
- Parámetro. μ: valor esperado (media poblacional).
- Hipótesis

$$\mathbf{H_0}: \quad \mu = 0$$

$$\mathbf{H_1}: \quad \mu \neq 0$$

Conceptos Base

¿Cómo Testar la Hipótesis?

1. **Prior Belief:** H_0 se asumirá verdadera hasta que se pruebe lo contrario.
2. **Estadístico de Contraste:** Encontrar un estadístico T cuya distribución, asumiendo H_0 verdadera, disponiendo de una muestra, sea completamente conocida.
3. **Región de Aceptación:** Definir un I.C. para T asumiendo H_0 es verdadera.
4. **Región Crítica (o de Rechazo):** El complemento del I.C. anterior representa el conjunto de valores para T que son improbables cuando H_0 es verdadera.
5. **Contrastar:** Si observamos que, vistos los datos experimentales, T toma valores dentro de la Región Crítica, rechazaremos H_0 y aceptaremos H_1. En otro caso no podemos rechazar H_0.
6. **Importante:** La región crítica elegida debe ser "compatible" con H_1.

Conceptos Base

Ejemplo - Botellas - 1

- Variable Aleatoria X: concentración de sodio (log del %) en botellas de agua mineral de cierta marca.
- Parámetro. μ: valor esperado (media poblacional).
- Hipótesis

$$\mathbf{H_0}: \ \mu = 0$$
$$\mathbf{H_1}: \ \mu \neq 0$$

- Para estudiar la veracidad de la hipótesis, recogemos una muestra de $n = 25$ botellas y medimos la concentración promedio de sodio.
- Supongamos que observamos un valor de $\bar{X} = 0.8$. ¿Es creíble H_0?

Conceptos Base

Ejemplo - Botellas - 1

- Asumiendo $X \sim \mathcal{N}(\mu, \sigma^2)$ con σ conocida, sabemos que

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

- Por lo tanto, con probabilidad $1 - \alpha$

$$Z \in \left[z_{1-\alpha/2}, z_{\alpha/2}\right]$$

- Si asumimos H_0 cierta, $\mu = 0$. Por lo tanto, con probabilidad $1 - \alpha$,

$$\frac{\bar{X}}{\sigma/\sqrt{n}} \in \left[z_{1-\alpha/2}, z_{\alpha/2}\right]$$

- Con $\alpha = 0.05$, $\sigma = 1$ y $n = 25$, tenemos que con probabilidad 0.95,

$$\frac{\bar{X}}{1/5} \in\in [-1.96, 1.96] \equiv \bar{X} \in [-0.392, 0.392]$$

Conceptos Base

Ejemplo - Botellas - 1

- ....con $\alpha = 0.05$, $\sigma = 1$ y $n = 25$, tenemos que con probabilidad 0.95,

$$\bar{X} \in \left[\pm 1.96 \frac{1}{5}\right] \equiv [-0.392, 0.392]$$

- ¿Qué sucede si experimentalmente, observamos que $\bar{X} = 0.8$?. Del razonamiento anterior, tenemos si H_0 fuese cierta

$$\bar{X} \notin [-0.392, 0.392]$$

  con probabilidad a lo más 0.05.
- Por lo tanto, si H_0 es cierta, el resultado $\bar{X} = 0.8$ es muy raro (ocurre menos del 5 % de las veces).
- Por lo tanto, o somos muy desafortunados o H_0 es falsa.
- Optamos por rechazar H_0 y aceptar H_1. Se dice que H_0 se rechaza a un nivel de significancia $\alpha = 0.05$.

Conceptos Base

Ejemplo - Botellas - 1

- Procedimiento Estándar: En vez de despejar el promedio y verificar si éste se encuentra efectivamente en el IC obtenido asumiendo H_0 correcta, podemos verificar directamente si el estadístico de contraste Z se encuentra en el IC esperado para él:

$$¿\ Z \in \left[z_{1-\alpha/2}, z_{\alpha/2}\right]?$$

- Para $\alpha = 0.05$

$$¿\ Z \in [-1.96, 1.96]?$$

- En nuestro caso y asumiendo H_0 cierta

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{0.8 - 0.0}{1/\sqrt{25}} = 4$$

- Como $4 \notin [-1.96, 1.96]$, optamos por rechazar H_0 y aceptar H_1.
- Se dice que H_0 se rechaza a un nivel de significancia $\alpha = 0.05$.

Conceptos Base

Ejemplo - Botellas - 2

- Supongamos que ahora nuestra hipótesis alternativa es ligeramente diferente:
- Hipótesis

$$\mathbf{H_0}: \quad \mu = 0 \quad (\text{concentración} = 1\,\%)$$
$$\mathbf{H_1}: \quad \mu < 0 \quad (\text{concentración} < 1\,\%)$$

- Para estudiar la veracidad de la hipótesis, recogemos una muestra de $n = 20$ botellas y medimos la concentración promedio de sodio.
- Supongamos que observamos un valor de $\bar{X} = 0.8$. ¿Es creíble H_0?

Conceptos Base

Ejemplo - Botellas - 2

- .... tenemos que con probabilidad 0.95,

$$\bar{X} \in \left[\pm 1.96 \frac{1}{5} \right] \equiv [-0.392, 0.392]$$

- Por lo tanto, si H_0 es cierta, el resultado $\bar{X} = 0.8$ es muy inusual (ocurre menos del 5 %) de las veces.
- Por lo tanto, optamos por rechazar H_0 y aceptar H_1.
- Pero $H_1 : \mu < 0$!!. Si H_1 es cierta, $\bar{X} = 0.8$ es aún más inusual!
- No debiéramos permitir que el resultado $\bar{X} = 0.8$ apoye H_1.

Conceptos Base

Ejemplo - Botellas - 2

- Rediseñados el procedimiento para el sistema
- Hipótesis

$$\mathbf{H_0}: \quad \mu = 0$$
$$\mathbf{H_1}: \quad \mu < 0$$

- Asumiendo $X \sim \mathcal{N}(\mu, \sigma^2)$ con σ conocida, sabemos que

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

- Por lo tanto, con probabilidad $1 - \alpha$

$$Z \in [z_{1-\alpha}, \infty]$$

- Ahora construimos un IC asimétrico para $\bar{X}$ que cubra lo mejor posible la parte donde se ubica Z si la hipótesis alternativa fuese falsa.

Conceptos Base

Ejemplo - Botellas - 2

- ... con probabilidad $1-\alpha$

$$Z \in [z_{1-\alpha}, \infty]$$

- Si H_0 es cierta,

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\bar{X}}{\sigma/\sqrt{n}} \in [z_{1-\alpha}, \infty]$$

- ....con $\alpha = 0.05$, $\sigma = 1$ y $n = 25$, tenemos que con probabilidad 0.95,

$$\bar{X} \in \left[-1.65\frac{1}{5}, \infty\right] \equiv [-0.33, \infty]$$

- ¿Qué sucede si experimentalmente, observamos que $\bar{X} = 0.8$?. Del razonamiento anterior, tenemos que con probabilidad 0.95, H_0 es compatible con este resultado experimental.
- Por lo tanto, no podemos rechazar H_0!.

Conceptos Base

Ejemplo - Botellas - 2

- Procedimiento Estándar: En vez de despejar el promedio y verificar si éste se encuentra efectivamente en el IC obtenido asumiendo H_0 correcta, podemos verificar directamente si el estadístico de contraste Z se encuentra en el IC esperado para él:

$$¿\ Z \in [z_{1-\alpha}, \infty]\ ?$$

- Para $\alpha = 0.05$

$$¿\ Z \in [-1.65, \infty]?$$

- En nuestro caso y asumiendo H_0 cierta

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{0.8 - 0.0}{1/\sqrt{25}} = 4$$

- Como $4 \in [-1.65, \infty]$, no podemos rechazar H_0.

Conceptos Base

Conceptos Clave

- Hipótesis Nula H_0 versus Hipótesis Alternativa H_1.
- Estadístico de Contraste (T).
- Región de Aceptación: IC para T a un nivel de significación α.
- Región Crítica: Valores de Z que de ser observados llevarán al rechazo de H_0 y la aceptación de H_1.

Conceptos Base

Tipos y Probabilidad de Errores

	H_0 es cierta	H_1 es cierta
Se rechaza H_0	**Error Tipo 1**	Bien
No se rechaza H_0	Bien	**Error Tipo 2**

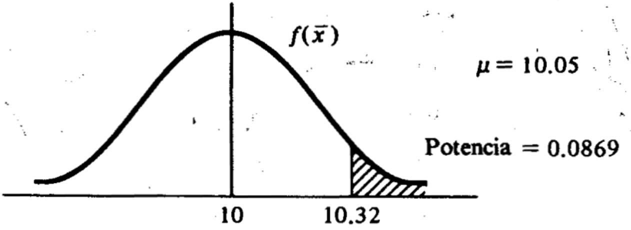
- Probabilidades de Error:

$$\alpha = P\,(\text{Rechazar } H_0 \mid H_0 \text{ Verdadera}) = P\,(\text{Error Tipo 1})$$
$$\beta = P\,(\text{No Rechazar } H_0 \mid H_1 \text{ Verdadera}) = P\,(\text{Error Tipo 2})$$

- Significanción o tamaño del test: α (prob. de rechazar H_0 cuando H_0 es cierta).
- Poder o potencia del test: $1 - \beta$ (prob. de aceptar H_1 cuando H_1 es cierta).
- Queremos un test minimizar el tamaño y maximizar la potencia.
- Típicamente un error tipo 1 es mucho más grave que un error tipo 2 porque H_0 representa aquello que se debiese creer por defecto.

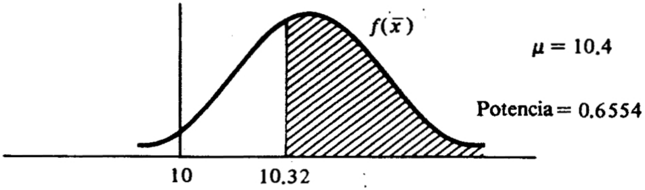
Conceptos Base

- El valor de α será fijado a un determinado valor (nivel de significación).
- El valor de β dependerá de la distribución de Z bajo H_1.
- Supongamos que rechazamos H_0 cuando $Z > 10.32$.
- Supongamos que bajo H_1, $Z \sim \mathcal{N}(10, \sigma^2)$.
- La potencia sería:



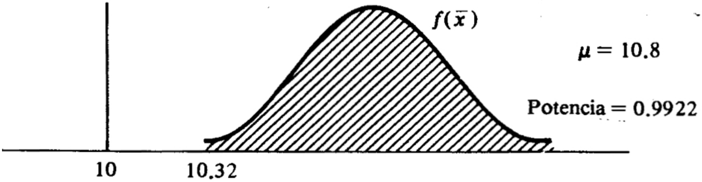
Conceptos Base

- El valor de $1-\beta$ (potencia) depende de la distribución de Z bajo H_1.
- Supongamos que rechazamos H_0 cuando $Z > 10.32$.
- Supongamos que bajo H_1, $Z \sim \mathcal{N}(10.4, \sigma^2)$.
- La potencia sería:



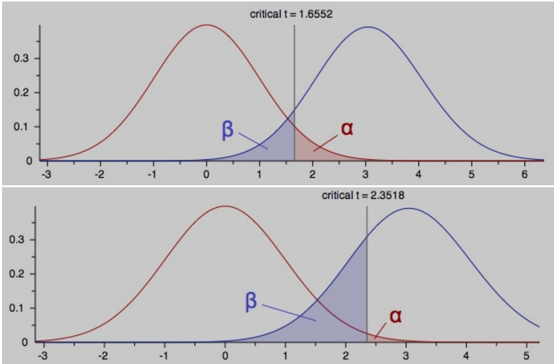
Conceptos Base

- El valor de β depende de la distribución de Z bajo H_1.
- Supongamos que rechazamos H_0 cuando $Z > 10.32$.
- Supongamos que bajo H_1, $Z \sim \mathcal{N}(10.8, \sigma^2)$.
- La potencia sería:



Conceptos Base

- Típico tradeoff entre α y β.
- Curva Roja: Distribución del estadístico de contraste (Z) si H_0 fuese verdadera.
- Curva Roja: Distribución del estadístico de contraste (Z) si H_1 fuese verdadera.
- Errores si rechazamos cuando $Z > 1.6552$ versus errores si rechazamos cuando $Z > 2.3518$.



Conceptos Base

Criterio Estándar

- Fijar la probabilidad de un error tipo 1 máxima tolerada (α) y buscar maximizar la potencia (poder).

Test Uniformemente más Potentes

Consideremos todo el conjunto de contrastes con tamaño $\leq \alpha$ para un determinado problema. Si dentro de esta familia existe un test que maximiza la potencia (poder), éste se denomina contraste uniformemente más potente. A veces existe a veces no.

Outline

II. Contrastes Clásicos para la Media de Una Población

Contrastes Clásicos

Supuestos Clave

- Si X e Y denotan las variables estudiadas, asumiremos que $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ e $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ con parámetros posiblemente desconocidos.
- El último supuesto se puede levantar si usamos el Teorema del Límite Central y declaramos que los procedimientos son asintóticamente correctos.
- Existen métodos modernos de contraste (computacionales) que no requieren asumir normalidad, pero que no podemos estudiar por razones de tiempo.
- Asumiremos que las muestras involucradas son todas IID.
- Denotaremos las muestras $\underline{X} = (X_1, X_2, \ldots, X_{n_X})$ y $\underline{Y} = (Y_1, Y_2, \ldots, Y_{n_Y})$.

Caso 1. Test para una Media con Varianza Conocida

1.1 Hipótesis Bilateral

$$\mathbf{H_0}: \ \mu_X = \gamma_0$$
$$\mathbf{H_1}: \ \mu_X \neq \gamma_0$$

1.1 Estadístico de Contraste

Asumiendo σ_X^2 conocida,

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n_X}} \overset{\text{bajo } H_0}{\sim} \frac{\bar{X} - \gamma_0}{\sigma_X/\sqrt{n_X}} \overset{\text{bajo } H_0}{\sim} \mathcal{N}(0,1)$$

1.1 Región Crítica

Rechazamos H_0 con significancia α si

$$Z < z_{1-\alpha/2} \ \text{ó} \ Z > z_{\alpha/2}$$

Caso 1. Test para una Media con Varianza Conocida

1.2 Hipótesis Unilateral 1

$$\mathbf{H_0}: \ \mu_X = \gamma_0$$
$$\mathbf{H_1}: \ \mu_X > \gamma_0$$

1.2 Estadístico de Contraste

Asumiendo σ_X^2 conocida,

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X/\sqrt{n_X}} \ \underbrace{\sim}_{\text{bajo } H_0} \ \frac{\bar{X} - \gamma_0}{\sigma_X/\sqrt{n_X}} \ \underbrace{\sim}_{\text{bajo } H_0} \ \mathcal{N}(0,1)$$

1.2 Región Crítica

Rechazamos H_0 con significancia α si

$$Z > z_\alpha$$

Caso 1. Test para una Media con Varianza Conocida

1.3 Hipótesis Unilateral 2

$$\mathbf{H_0}: \ \mu_X = \gamma_0$$
$$\mathbf{H_1}: \ \mu_X < \gamma_0$$

1.3 Estadístico de Contraste

Asumiendo σ_X^2 conocida,

$$Z = \frac{\bar{X} - \mu_X}{\sigma_X / \sqrt{n_X}} \underset{\sim}{\text{bajo } H_0} \frac{\bar{X} - \gamma_0}{\sigma_X / \sqrt{n_X}} \underset{\sim}{\text{bajo } H_0} \mathcal{N}(0,1)$$

1.3 Región Crítica

Rechazamos H_0 con significancia α si

$$Z < z_{1-\alpha}$$

Caso 2. Test para una Media con Varianza Desconocida

2.1 Hipótesis Bilateral

$$\mathbf{H_0}: \ \mu_X = \gamma_0$$
$$\mathbf{H_1}: \ \mu_X \neq \gamma_0$$

2.1 Estadístico de Contraste

Sin asumir σ_X^2 conocida,

$$T = \frac{\bar{X} - \mu_X}{S_X/\sqrt{n_X}} \underset{\text{bajo } H_0}{\sim} \frac{\bar{X} - \gamma_0}{S_X/\sqrt{n_X}} \underset{\text{bajo } H_0}{\sim} \mathcal{T}_{n_X - 1} \ \text{ con } \ S_X^2 = \frac{\sum_{i=1}^{n_X}(X_i - \bar{X})^2}{n_X - 1}$$

2.1 Región Crítica

Rechazamos H_0 con significancia α si

$$T < t_{1-\alpha/2}^{n_X-1} \ \text{ ó } \ T > t_{\alpha/2}^{n_X-1} \ ,$$

con $t_{\alpha/2}^{n_X-1}$ el percentil de la distribución T-Student con $n_X - 1$ grados de libertad que acumula probabilidad $1 - \alpha/2$. Análogamente $t_{1-\alpha/2}^{n_X-1}$.

Caso 2. Test para una Media con Varianza Desconocida

2.2 Hipótesis Unilateral 1

$$\mathbf{H_0}: \ \mu_X = \gamma_0$$
$$\mathbf{H_1}: \ \mu_X > \gamma_0$$

2.2 Estadístico de Contraste

Sin asumir σ_X^2 conocida,

$$T = \frac{\bar{X} - \mu_X}{S_X/\sqrt{n_X}} \underset{\widetilde{}}{\text{bajo } H_0} \frac{\bar{X} - \gamma_0}{S_X/\sqrt{n_X}} \overset{\text{bajo } H_0}{\sim} \mathcal{T}_{n_X-1} \ \text{ con } \ S_X^2 = \frac{\sum_{i=1}^{n_X}(X_i - \bar{X})^2}{n_X - 1}$$

2.2 Región Crítica

Rechazamos H_0 con significancia α si

$$T > t_\alpha^{n_X-1}$$

con $t_\alpha^{n_X-1}$ el percentil de la distribución T-Student con $n_X - 1$ grados de libertad que acumula probabilidad $1-\alpha$.

Caso 2. Test para una Media con Varianza Desconocida

2.3 Hipótesis Unilateral 2

$$\mathbf{H_0}: \ \mu_X = \gamma_0$$
$$\mathbf{H_1}: \ \mu_X < \gamma_0$$

2.3 Estadístico de Contraste

Sin asumir σ_X^2 conocida,

$$T = \frac{\bar{X} - \mu_X}{S_X/\sqrt{n_X}} \overset{\text{bajo } H_0}{\sim} \frac{\bar{X} - \gamma_0}{S_X/\sqrt{n_X}} \overset{\text{bajo } H_0}{\sim} \mathcal{T}_{n_X - 1} \quad \text{con} \quad S_X^2 = \frac{\sum_{i=1}^{n_X} (X_i - \bar{X})^2}{n_X - 1}$$

2.3 Región Crítica

Rechazamos H_0 con significancia α si

$$T < t_{1-\alpha}^{n_X - 1}$$

con $t_{1-\alpha}^{n_X-1}$ el percentil de la distribución T-Student con $n_X - 1$ grados de libertad que acumula probabilidad α.

Outline

III. Contrastes Clásicos para las Medias de dos Poblaciones

Ejemplo

Dos Poblaciones

- Sea X el tiempo de vida de las personas que viven en la X región.
- Sea Y el tiempo de vida de las personas que viven en la región Metropolitana.
- Supongamos que $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$.
- Supongamos que $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$.

Ejemplo

Dos Poblaciones

"El tiempo de vida medio de las personas que viven en la X región es mayor que tiempo de vida medio de las personas que viven en la región Metropolitana".

- Posibles Hipótesis:

$$\mathbf{H_0}: \; \mu_X = \mu_Y$$
$$\mathbf{H_1}: \; \mu_X > \mu_Y$$

Ejemplo

Dos Poblaciones

"El tiempo de vida medio de las personas que viven en la X región es mayor que tiempo de vida medio de las personas que viven en la región Metropolitana".

Diseño del Test/Contraste

- $\underline{X} = (X_1, X_2, \ldots, X_{n_X})$ una muestra IID para X.
- $\underline{Y} = (Y_1, Y_2, \ldots, Y_{n_Y})$ una muestra IID para Y.
- Medimos $\Delta = \bar{X} - \bar{Y}$ y calculamos el **Estadístico de Contraste:**

$$Z = \frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}}$$

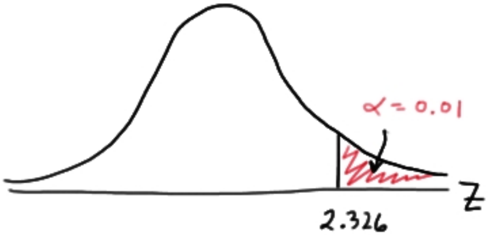
- Asumiendo H_0 verdadera, $Z \sim \mathcal{N}(0,1)$[a].
- **Región Crítica**: $Z > z_\alpha$

[a] asumiendo $Var[X] = \sigma_X^2$, $Var[X] = \sigma_Y^2$ conocidas.

Ejemplo

Diseño del Test/Contraste

- **Estadístico de Contraste:** $Z \sim \mathcal{N}(0,1)$



- **Región Crítica**: $Z > z_\alpha$.
- Ojo: interesa la distribución de Z condicional a H_0.

Ejemplo

Dos Poblaciones

En vez de estudiar el sistema:

$$\begin{aligned} \mathbf{H_0} &: \ \mu_X = \mu_Y \\ \mathbf{H_1} &: \ \mu_X > \mu_Y \ , \end{aligned}$$

que es equivalente a:

$$\begin{aligned} \mathbf{H_0} &: \ \mu_X - \mu_Y = 0 \\ \mathbf{H_1} &: \ \mu_X - \mu_Y > 0 \ , \end{aligned}$$

consideraremos una version más general:

$$\begin{aligned} \mathbf{H_0} &: \ \mu_X - \mu_Y = \delta_0 \\ \mathbf{H_1} &: \ \mu_X - \mu_Y > \delta_0 \ , \end{aligned}$$

y sus variantes ...

Caso 3. Test de Dos Medias con Varianzas Conocidas

3.1 Hipótesis Bilateral

$$\begin{array}{lll} \mathbf{H_0}: \ \mu_X = \mu_Y & \longrightarrow & \mathbf{H_0}: \ \mu_X - \mu_Y = \delta_0 \\ \mathbf{H_1}: \ \mu_X \neq \mu_Y & & \mathbf{H_1}: \ \mu_X - \mu_Y \neq \delta_0 \end{array}$$

3.1 Estadístico de Contraste

Asumiendo σ_X^2, σ_Y^2 conocidas,

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \overset{\text{bajo } H_0}{\sim} Z = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{N}(0,1)$$

3.1 Región Crítica

Rechazamos H_0 con significancia α si

$$Z < z_{1-\alpha/2} \ \text{ ó } \ Z > z_{\alpha/2}$$

Caso 3. Test de Dos Medias con Varianzas Conocidas

3.2 Hipótesis Unilateral 1

$$\begin{array}{ll} \mathbf{H_0}: \ \mu_X = \mu_Y \\ \mathbf{H_1}: \ \mu_X > \mu_Y \end{array} \quad \longrightarrow \quad \begin{array}{ll} \mathbf{H_0}: \ \mu_X - \mu_Y = \delta_0 \\ \mathbf{H_1}: \ \mu_X - \mu_Y > \delta_0 \end{array}$$

3.2 Estadístico de Contraste

Asumiendo σ_X^2, σ_Y^2 conocidas,

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \overset{\text{bajo } H_0}{\sim} Z = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{N}(0,1)$$

3.2 Región Crítica

Rechazamos H_0 con significancia α si

$$Z > z_\alpha$$

Caso 3. Test de Dos Medias con Varianzas Conocidas

3.3 Hipótesis Unilateral 2

$$\begin{array}{llcll} \mathbf{H_0}: & \mu_X = \mu_Y & \longrightarrow & \mathbf{H_0}: & \mu_X - \mu_Y = \delta_0 \\ \mathbf{H_1}: & \mu_X < \mu_Y & & \mathbf{H_1}: & \mu_X - \mu_Y < \delta_0 \end{array}$$

3.3 Estadístico de Contraste

Asumiendo σ_X^2, σ_Y^2 conocidas,

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \overset{\text{bajo } H_0}{\sim} Z = \frac{(\bar{X} - \bar{Y}) - \delta_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{N}(0,1)$$

3.3 Región Crítica

Rechazamos H_0 con significancia α si

$$Z < z_{1-\alpha}$$

Caso 4. Test de Dos Medias con Varianzas Homocedásticas

4 Caso Varianzas Desconocidas

Análogamente a lo que hemos venido haciendo en el caso de varianzas desconocidas, consideramos en este caso el estadístico de contraste que corresponde al cuociente entre el estadístico Z y el estadístico χ^2 que usamos para la varianza muestral.

4 Estadístico de Contraste

Sin asumir σ_X^2, σ_Y^2 conocidas, pero asumiendo que $\sigma_X^2 = \sigma_Y^2$

$$T = \frac{\left(\bar{X} - \bar{Y}\right) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} T = \frac{\left(\bar{X} - \bar{Y}\right) - \delta_0}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{T}_{n_X + n_Y - 2}$$

$$S_p^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2} \quad S_X^2 = \frac{\sum_{i=1}^{n_X}(X_i - \bar{X})^2}{n_X - 1} \quad S_Y^2 = \frac{\sum_{j=1}^{n_Y}(Y_j - \bar{Y})^2}{n_Y - 1}$$

Caso 4. Test de Dos Medias con Varianzas Homocedásticas

4.1 Hipótesis Bilateral

$$\begin{array}{llll} \mathbf{H_0}: & \mu_X = \mu_Y & \longrightarrow & \mathbf{H_0}: \; \mu_X - \mu_Y = \delta_0 \\ \mathbf{H_1}: & \mu_X \neq \mu_Y & & \mathbf{H_1}: \; \mu_X - \mu_Y \neq \delta_0 \end{array}$$

4.1 Estadístico de Contraste

Sin asumir σ_X^2, σ_Y^2 conocidas, pero asumiendo que $\sigma_X^2 = \sigma_Y^2$

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{T}_{n_X + n_Y - 2}$$

4.1 Región Crítica

Rechazamos H_0 con significancia α si

$$T < t_{1-\alpha/2}^{n_X+n_Y-2} \text{ ó } T > t_{\alpha/2}^{n_X+n_Y-2},$$

con $t_{1-\alpha/2}^{n_X+n_Y-2}$ el percentil de la distribución T-Student con $n_X + n_Y - 2$ grados de libertad que acumula probabilidad $\alpha/2$.

Caso 4. Test de Dos Medias con Varianzas Homocedásticas

4.2 Hipótesis Unilateral 1

$$\begin{array}{ll} \mathbf{H_0}: \ \mu_X = \mu_Y \\ \mathbf{H_1}: \ \mu_X > \mu_Y \end{array} \quad \longrightarrow \quad \begin{array}{ll} \mathbf{H_0}: \ \mu_X - \mu_Y = \delta_0 \\ \mathbf{H_1}: \ \mu_X - \mu_Y > \delta_0 \end{array}$$

4.2 Estadístico de Contraste

Sin asumir σ_X^2, σ_Y^2 conocidas, pero asumiendo que $\sigma_X^2 = \sigma_Y^2$

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{T}_{n_X + n_Y - 2}$$

4.2 Región Crítica

Rechazamos H_0 con significancia α si

$$T > t_\alpha^{n_X + n_Y - 2},$$

con $t_\alpha^{n_X + n_Y - 2}$ el percentil de la distribución T-Student con $n_X + n_Y - 2$ grados de libertad que acumula probabilidad $1 - \alpha$.

Caso 4. Test de Dos Medias con Varianzas Homocedásticas

4.3 Hipótesis Unilateral 2

$$\begin{array}{ll} \mathbf{H_0}: & \mu_X = \mu_Y \\ \mathbf{H_1}: & \mu_X < \mu_Y \end{array} \longrightarrow \begin{array}{ll} \mathbf{H_0}: & \mu_X - \mu_Y = \delta_0 \\ \mathbf{H_1}: & \mu_X - \mu_Y < \delta_0 \end{array}$$

4.3 Estadístico de Contraste

Sin asumir σ_X^2, σ_Y^2 conocidas, pero asumiendo que $\sigma_X^2 = \sigma_Y^2$

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{T}_{n_X + n_Y - 2}$$

4.3 Región Crítica

Rechazamos H_0 con significancia α si

$$T < t_{1-\alpha}^{n_X + n_Y - 2}$$

con $t_{1-\alpha}^{n_X + n_Y - 2}$ el percentil de la distribución T-Student con $n_X + n_Y - 2$ grados de libertad que acumula probabilidad α.

Caso 4. Test de Dos Medias con Varianzas Heterocedásticas

4 Caso Varianzas Desconocidas Heterocedásticas

Con respecto al caso anterior cambia sólo el número de grados de libertad de la T involucrada.

4 Estadístico de Contraste

Sin asumir σ_X^2, σ_Y^2 conocidas, y con $\sigma_X^2 \neq \sigma_Y^2$

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{T}_\nu$$

donde (fórmula de Welch) $\nu = \dfrac{\left(\frac{S_X^2}{n_X} + \frac{S_Y^2}{n_Y}\right)^2}{\frac{\left(\frac{S_X^2}{n_X}\right)^2}{n_X - 1} + \frac{\left(\frac{S_Y^2}{n_Y}\right)^2}{n_Y - 1}}$

Caso 4. Test de Dos Medias con Varianzas Heterocedásticas

4.1 Hipótesis Bilateral

$$\begin{array}{ll} \mathbf{H_0}: & \mu_X = \mu_Y \\ \mathbf{H_1}: & \mu_X \neq \mu_Y \end{array} \quad \longrightarrow \quad \begin{array}{ll} \mathbf{H_0}: & \mu_X - \mu_Y = \delta_0 \\ \mathbf{H_1}: & \mu_X - \mu_Y \neq \delta_0 \end{array}$$

4.1 Estadístico de Contraste

Sin asumir σ_X^2, σ_Y^2 conocidas, y con $\sigma_X^2 \neq \sigma_Y^2$

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{T}_\nu$$

4.1 Región Crítica

Rechazamos H_0 con significancia α si

$$T < t^\nu_{1-\alpha/2} \ \text{ó} \ T > t^\nu_{\alpha/2} \ ,$$

con $t^\nu_{1-\alpha/2}$ el percentil de la distribución T-Student con ν grados de libertad (fórmula de Welch) que acumula probabilidad $\alpha/2$.

Caso 4. Test de Dos Medias con Varianzas Heterocedásticas

4.2 Hipótesis Unilateral 1

$$\begin{array}{lll} \mathbf{H_0}: \ \mu_X = \mu_Y & \longrightarrow & \mathbf{H_0}: \ \mu_X - \mu_Y = \delta_0 \\ \mathbf{H_1}: \ \mu_X > \mu_Y & & \mathbf{H_1}: \ \mu_X - \mu_Y > \delta_0 \end{array}$$

4.2 Estadístico de Contraste

Sin asumir σ_X^2, σ_Y^2 conocidas, y con $\sigma_X^2 \neq \sigma_Y^2$

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{T}_\nu$$

4.2 Región Crítica

Rechazamos H_0 con significancia α si

$$T > t_\alpha^\nu \ ,$$

con t_α^ν el percentil de la distribución T-Student con ν grados de libertad (fórmula de Welch) que acumula probabilidad $1 - \alpha$.

Caso 4. Test de Dos Medias con Varianzas Heterocedásticas

4.3 Hipótesis Unilateral 2

$$\begin{array}{lll} \mathbf{H_0}: \ \mu_X = \mu_Y & \longrightarrow & \mathbf{H_0}: \ \mu_X - \mu_Y = \delta_0 \\ \mathbf{H_1}: \ \mu_X < \mu_Y & & \mathbf{H_1}: \ \mu_X - \mu_Y < \delta_0 \end{array}$$

4.3 Estadístico de Contraste

Sin asumir σ_X^2, σ_Y^2 conocidas, y con $\sigma_X^2 \neq \sigma_Y^2$

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} T = \frac{(\bar{X} - \bar{Y}) - \delta_0}{S\sqrt{\frac{1}{n_X} + \frac{1}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{T}_\nu$$

4.3 Región Crítica

Rechazamos H_0 con significancia α si

$$T < t_{1-\alpha}^{\nu}$$

con $t_{1-\alpha}^{\nu}$ el percentil de la distribución T-Student con ν grados de libertad (fórmula de Welch) que acumula probabilidad α.

Recordemos ...

Supuestos Clave

- Si X e Y denotan las variables estudiadas, asumiremos que

$$X \sim \mathcal{N}(\mu_X, \sigma_X^2)$$
$$Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$$

con parámetros posiblemente desconocidos.

- Denotaremos las muestras

$$\underline{X} = (X_1, X_2, \ldots, X_{n_X})$$
$$\underline{Y} = (Y_1, Y_2, \ldots, Y_{n_Y})$$

- Asumiremos que las muestras involucradas son todas IID.
- Definimos $S_p^2 = \frac{(n_X-1)S_X^2+(n_Y-1)S_Y^2}{n_X+n_Y-2}$ $\quad S_X^2 = \frac{\sum_{i=1}^{n_X}(X_i-\bar{X})^2}{n_X-1}$ $\quad S_Y^2 = \frac{\sum_{j=1}^{n_Y}(Y_j-\bar{Y})^2}{n_Y-1}$

Outline

IV. Contrastes Clásicos para Una y Dos Varianzas

Caso 5. Test para el valor de la Varianza

5.1 Hipótesis Bilateral

$$\mathbf{H_0}: \ \sigma_X^2 = \sigma_0^2$$
$$\mathbf{H_1}: \ \sigma_X^2 \neq \sigma_0^2$$

5.1 Estadístico de Contraste

$$\chi = \frac{(n_X - 1)S_X^2}{\sigma_0^2} \ \overset{\text{bajo } H_0}{\sim} \ \chi^2_{n_X - 1} \ \text{ con } \ S_X^2 = \frac{\sum_{i=1}^{n_X}(X_i - \bar{X})^2}{n_X - 1}$$

5.1 Región Crítica

Rechazamos H_0 con significancia α si

$$\chi < \chi^{n_X - 1}_{1-\alpha/2} \ \text{ ó } \ \chi > \chi^{n_X - 1}_{\alpha/2} \ ,$$

con $\chi^{n_X - 1}_{1-\alpha/2}$ el percentil de la distribución Chi-cuadrado con $n_X - 1$ grados de libertad que acumula probabilidad $\alpha/2$.

Caso 5. Test para el valor de la Varianza

5.2 Hipótesis Unilateral 1

$$\mathbf{H_0}: \ \sigma_X^2 = \sigma_0^2$$
$$\mathbf{H_1}: \ \sigma_X^2 > \sigma_0^2$$

5.2 Estadístico de Contraste

$$\chi = \frac{(n_X - 1)S_X^2}{\sigma_0^2} \underset{\sim}{\text{bajo } H_0} \chi^2_{n_X - 1} \quad \text{con} \quad S_X^2 = \frac{\sum_{i=1}^{n_X}(X_i - \bar{X})^2}{n_X - 1}$$

5.2 Región Crítica

Rechazamos H_0 con significancia α si

$$\chi > \chi_\alpha^{n_X - 1},$$

con $\chi_\alpha^{n_X - 1}$ el percentil de la distribución Chi-cuadrado con $n_X - 1$ grados de libertad que acumula probabilidad $1 - \alpha$.

Caso 5. Test para el valor de la Varianza

5.3 Hipótesis Unilateral 2

$$\mathbf{H_0}: \ \sigma_X^2 = \sigma_0^2$$
$$\mathbf{H_1}: \ \sigma_X^2 < \sigma_0^2$$

5.3 Estadístico de Contraste

$$\chi = \frac{(n_X - 1)S_X^2}{\sigma_0^2} \underbrace{\sim}_{\text{bajo } H_0} \chi^2_{n_X - 1} \quad \text{con} \quad S_X^2 = \frac{\sum_{i=1}^{n_X}(X_i - \bar{X})^2}{n_X - 1}$$

5.3 Región Crítica

Rechazamos H_0 con significancia α si

$$\chi < \chi_{1-\alpha}^{n_X - 1}$$

con $\chi_{1-\alpha}^{n_X-1}$ el percentil de la distribución Chi-cuadrado con $n_X - 1$ grados de libertad que acumula probabilidad α.

Caso 6. Test para comparar dos Varianzas (Homocedasticidad)

6.1 Hipótesis Bilateral

$$\mathbf{H_0} : \ \sigma_X^2 = \sigma_Y^2$$
$$\mathbf{H_1} : \ \sigma_X^2 \neq \sigma_Y^2$$

6.1 Estadístico de Contraste

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \ \overset{\text{bajo } H_0}{\sim} \ \frac{S_X^2}{S_Y^2} \ \overset{\text{bajo } H_0}{\sim} \ F_{n_X-1,n_Y-1}$$

6.1 Región Crítica

Rechazamos H_0 con significancia α si

$$F < f_{1-\alpha/2}^{n_X-1,n_Y-1} \ \text{ó} \ F > f_{\alpha/2}^{n_X-1,n_Y-1},$$

con $f_{1-\alpha/2}^{n_X-1,n_Y-1}$ el percentil de la distribución F con $n_X - 1$ y $n_Y - 1$ grados de libertad que acumula probabilidad $\alpha/2$.

Caso 6. Test para comparar dos Varianzas (Homocedasticidad)

6.2 Hipótesis Unilateral 1

$$\mathbf{H_0}: \ \sigma_X^2 = \sigma_Y^2$$

$$\mathbf{H_1}: \ \sigma_X^2 > \sigma_Y^2$$

6.2 Estadístico de Contraste

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \overset{\text{bajo } H_0}{\sim} \frac{S_X^2}{S_Y^2} \overset{\text{bajo } H_0}{\sim} F_{n_X-1,n_Y-1}$$

6.2 Región Crítica

Rechazamos H_0 con significancia α si

$$F > f_\alpha^{n_X-1,n_Y-1} \ ,$$

con $f_\alpha^{n_X-1,n_Y-1}$ el percentil de la distribución F con $n_X - 1$ y $n_Y - 1$ grados de libertad que acumula probabilidad $1 - \alpha$.

Caso 6. Test para comparar dos Varianzas (Homocedasticidad)

6.3 Hipótesis Unilateral 1

$$\mathbf{H_0}: \ \sigma_X^2 = \sigma_Y^2$$

$$\mathbf{H_1}: \ \sigma_X^2 < \sigma_Y^2$$

6.3 Estadístico de Contraste

$$F = \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \overset{\text{bajo } H_0}{\sim} \frac{S_X^2}{S_Y^2} \overset{\text{bajo } H_0}{\sim} F_{n_X-1, n_Y-1}$$

6.3 Región Crítica

Rechazamos H_0 con significancia α si

$$F < f_{1-\alpha}^{n_X-1, n_Y-1} ,$$

con $f_{1-\alpha}^{n_X-1, n_Y-1}$ el percentil de la distribución F con $n_X - 1$ y $n_Y - 1$ grados de libertad que acumula probabilidad α.

Outline

V. Contrastes Clásicos para Una y Dos Proporciones

Caso 7. Test para Una Proporción

Estrategia General para Una Proporción

Si deseamos estimar la proporción de una determinada población que satisface un determinado criterio, podemos considerar una variable aleatoria Bernoulli X que toma el valor 1 con probabilidad p y 0 con probabilidad $1-p$. $X(\omega)=1$ representa que el individuo ω satisface el criterio. Si tomamos ahora una muestra IID de X, digamos $X_1, X_2, \ldots, X_n$, y consideramos la v.a.

$$\Sigma_X = \sum_i X_i$$

sabemos de nuestro estudio de v.a. que $\Sigma_X \sim \text{Bin}(n,p)$ con $E[\Sigma_X] = np$ y $\text{Var}(\Sigma_X) = np(1-p)$. Si p no está demasiado lejos de 0.5, esta v.a. se puede aproximar como una v.a. normalmente distribuida con valor esperado y varianza equivalentes, i.e. $\Sigma \approx \mathcal{N}(np, np(1-p))$. De aquí, es fácil ver que

$$\hat{p}_X = \frac{\Sigma_X}{n} \approx \mathcal{N}\left(p, \frac{p(1-p)}{n}\right) .$$

Caso 7. Test para Una Proporción

7.1 Hipótesis Bilateral

$$\mathbf{H_0}: \; p_X = p_0$$
$$\mathbf{H_1}: \; p_X \neq p_0$$

7.1 Estadístico de Contraste

$$Z = \frac{\hat{p}_X - p}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X}}} \underset{\text{bajo } H_0}{\sim} \frac{\hat{p}_X - p_0}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X}}} \underset{\text{bajo } H_0}{\sim} \mathcal{N}(0,1)$$

7.1 Región Crítica

Rechazamos H_0 con significancia α si

$$Z < z_{1-\alpha/2} \text{ ó } Z > z_{\alpha/2} \; ,$$

con $z_{1-\alpha/2}$ el percentil de la distribución $\mathcal{N}(0,1)$ que acumula probabilidad $\alpha/2$ y con $z_{\alpha/2}$ el percentil de la distribución $\mathcal{N}(0,1)$ que acumula probabilidad $1-\alpha/2$.

Caso 7. Test para Una Proporción

7.2 Hipótesis Unilateral 1

$$\mathbf{H_0}: \; p_X = p_0$$
$$\mathbf{H_1}: \; p_X > p_0$$

7.2 Estadístico de Contraste

$$Z = \frac{\hat{p}_X - p}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X}}} \underset{\sim}{\text{bajo } H_0} \frac{\hat{p}_X - p_0}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X}}} \underset{\sim}{\text{bajo } H_0} \mathcal{N}(0,1)$$

7.2 Región Crítica

Rechazamos H_0 con significancia α si

$$Z > z_\alpha \; ,$$

con z_α el percentil de la distribución $\mathcal{N}(0,1)$ que acumula probabilidad $1-\alpha$.

Caso 7. Test para Una Proporción

7.3 Hipótesis Bilateral

$$\mathbf{H_0}: \quad p_X = p_0$$
$$\mathbf{H_1}: \quad p_X < p_0$$

7.3 Estadístico de Contraste

$$Z = \frac{\hat{p}_X - p}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X}}} \overset{\text{bajo } H_0}{\sim} \frac{\hat{p}_X - p_0}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X}}} \overset{\text{bajo } H_0}{\sim} \mathcal{N}(0,1)$$

7.3 Región Crítica

Rechazamos H_0 con significancia α si

$$Z < z_{1-\alpha} \ ,$$

con $z_{1-\alpha}$ el percentil de la distribución $\mathcal{N}(0,1)$ que acumula probabilidad α.

Caso 8. Test para Comparar dos Proporciones

Estrategia General para Dos Proporciones

En este caso tenemos dos poblaciones y por lo tanto dos v.a. Bernoulli X e Y, donde X toma el valor 1 con probabilidad p_X y 0 con probabilidad $1 - p_X$ mientras que Y toma el valor 1 con probabilidad p_Y y 0 con probabilidad $1 - p_Y$. $X(\omega_1) = 1$ representa que el individuo ω_1 de la primera población satisface el criterio estudiado e $Y(\omega_2) = 1$ representa que el individuo ω_2 de la segunda población satisface el criterio estudiado. Si tomamos ahora sendas muestras IID de X e Y, tenemos que

$$\hat{p}_X = \frac{\sum_i X_i}{n_X} =\approx \mathcal{N}\left(p_X, \frac{p_X(1-p_X)}{n_X}\right)$$

$$\hat{p}_Y = \frac{\sum_i Y_i}{n_Y} =\approx \mathcal{N}\left(p_Y, \frac{p_Y(1-p_Y)}{n_Y}\right)$$

De aquí, es fácil ver que

$$\hat{p}_X - \hat{p}_Y =\approx \mathcal{N}\left(p_X - p_Y, \frac{p_X(1-p_X)}{n_X} + \frac{p_Y(1-p_Y)}{n_Y}\right) .$$

Caso 8. Test para Comparar dos Proporciones

8.1 Hipótesis Bilateral

$$\mathbf{H_0}: \; p_X = p_Y$$
$$\mathbf{H_1}: \; p_X \neq p_Y$$

8.1 Estadístico de Contraste

$$Z = \frac{(\hat{p}_X - \hat{p}_Y) - (p_X - p_Y)}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \frac{(\hat{p}_X - \hat{p}_Y)}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{N}(0,1)$$

8.1 Región Crítica

Rechazamos H_0 con significancia α si

$$Z < z_{1-\alpha/2} \text{ ó } Z > z_{\alpha/2} \, ,$$

con $z_{1-\alpha/2}$ el percentil de la distribución $\mathcal{N}(0,1)$ que acumula probabilidad $\alpha/2$ y con $z_{\alpha/2}$ el percentil de la distribución $\mathcal{N}(0,1)$ que acumula probabilidad $1-\alpha/2$.

Caso 8. Test para Comparar dos Proporciones

8.2 Hipótesis Unilateral 1

$$\mathbf{H_0}: \quad p_X = p_Y$$
$$\mathbf{H_1}: \quad p_X > p_Y$$

8.2 Estadístico de Contraste

$$Z = \frac{(\hat{p}_X - \hat{p}_Y) - (p_X - p_Y)}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \frac{(\hat{p}_X - \hat{p}_Y)}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{N}(0,1)$$

8.2 Región Crítica

Rechazamos H_0 con significancia α si

$$Z > z_\alpha \ ,$$

con z_α el percentil de la distribución $\mathcal{N}(0,1)$ que acumula probabilidad $1-\alpha$.

Caso 8. Test para Comparar dos Proporciones

8.3 Hipótesis Bilateral

$$\mathbf{H_0}: \quad p_X = p_Y$$

$$\mathbf{H_1}: \quad p_X < p_Y$$

8.3 Estadístico de Contraste

$$Z = \frac{(\hat{p}_X - \hat{p}_Y) - (p_X - p_Y)}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}} \underset{}{\overset{\text{bajo } H_0}{\sim}} \frac{(\hat{p}_X - \hat{p}_Y)}{\sqrt{\frac{\hat{p}_X(1-\hat{p}_X)}{n_X} + \frac{\hat{p}_Y(1-\hat{p}_Y)}{n_Y}}} \overset{\text{bajo } H_0}{\sim} \mathcal{N}(0,1)$$

8.3 Región Crítica

Rechazamos H_0 con significancia α si

$$Z < z_{1-\alpha} \ ,$$

con $z_{1-\alpha}$ el percentil de la distribución $\mathcal{N}(0,1)$ que acumula probabilidad α.

Hipótesis Nulas Compuestas

¿Cómo manejar hipótesis de la forma?

$$\mathbf{H_0}: \mu_X \geq \delta_0$$
$$\mathbf{H_1}: \mu_X < \delta_0$$

Criterio Estándar

Usar el Sistema:
$$\mathbf{H_0}: \mu_X = \delta_0$$
$$\mathbf{H_1}: \mu_X < \delta_0$$

Motivo

Si H_0 es verdadera $\mu_X \geq \delta_0$. El peor error tipo 1 se observa cuando $\mu_X = \delta_0$.

Hipótesis Nulas Compuestas

¿Cómo manejar hipótesis de la forma?

$$\begin{aligned} \mathbf{H_0} &: \ \mu_X \leq \delta_0 \\ \mathbf{H_1} &: \ \mu_X > \delta_0 \end{aligned}$$

Criterio Estándar

Usar el Sistema:
$$\begin{aligned} \mathbf{H_0} &: \ \mu_X = \delta_0 \\ \mathbf{H_1} &: \ \mu_X > \delta_0 \end{aligned}$$

Motivo

Si H_0 es verdadera $\mu_X \leq \delta_0$. El peor error tipo 1 se observa cuando $\mu_X = \delta_0$.

Hipótesis Nulas Compuestas

¿Cómo manejar hipótesis de la forma?

$$\begin{array}{ll} \mathbf{H_0}: & \mu_X \geq \mu_Y \\ \mathbf{H_1}: & \mu_X < \mu_Y \end{array} \longrightarrow \begin{array}{ll} \mathbf{H_0}: & \mu_X - \mu_Y \geq 0 \\ \mathbf{H_1}: & \mu_X - \mu_Y < 0 \end{array}$$

Criterio Estándar

Usar el Sistema:
$$\begin{array}{ll} \mathbf{H_0}: & \mu_X = \mu_Y \\ \mathbf{H_1}: & \mu_X < \mu_Y \end{array} \longrightarrow \begin{array}{ll} \mathbf{H_0}: & \mu_X - \mu_Y = 0 \\ \mathbf{H_1}: & \mu_X - \mu_Y < 0 \end{array}$$

Motivo

Si H_0 es verdadera $\mu_X - \mu_Y \geq 0$. El peor error tipo 1 se observa cuando $\mu_X - \mu_Y = 0$.

Hipótesis Nulas Compuestas

Analogamente ... un sistema de la forma

$$\mathbf{H_0}: \ \mu_X - \mu_Y \leq \delta_0$$
$$\mathbf{H_1}: \ \mu_X - \mu_Y > \delta_0$$

Se maneja como

$$\mathbf{H_0}: \ \mu_X - \mu_Y = \delta_0$$
$$\mathbf{H_1}: \ \mu_X - \mu_Y > \delta_0$$

Hipótesis Nulas Compuestas

¿Cómo manejar hipótesis de la forma?

$$\begin{aligned} \mathbf{H_0} &: \ \sigma_X^2 \geq \sigma_0^2 \\ \mathbf{H_1} &: \ \sigma_X^2 < \sigma_0^2 \end{aligned}$$

Criterio Estándar

Usar el Sistema:
$$\begin{aligned} \mathbf{H_0} &: \ \sigma_X^2 = \sigma_0^2 \\ \mathbf{H_1} &: \ \sigma_X^2 < \sigma_0^2 \end{aligned}$$

Motivo

Si H_0 es verdadera $\sigma_X^2 \geq \sigma_0^2$. El peor error tipo 1 se observa cuando $\sigma_X^2 = \sigma_0^2$.

Hipótesis Nulas Compuestas

¿Cómo manejar hipótesis de la forma?

$$\begin{aligned} \mathbf{H_0} &: \ \sigma_X^2 \leq \sigma_0^2 \\ \mathbf{H_1} &: \ \sigma_X^2 > \sigma_0^2 \end{aligned}$$

Criterio Estándar

Usar el Sistema:
$$\begin{aligned} \mathbf{H_0} &: \ \sigma_X^2 = \sigma_0^2 \\ \mathbf{H_1} &: \ \sigma_X^2 > \sigma_0^2 \end{aligned}$$

Motivo

Si H_0 es verdadera $\sigma_X^2 \leq \sigma_0^2$. El peor error tipo 1 se observa cuando $\sigma_X^2 = \sigma_0^2$.

Hipótesis Nulas Compuestas

¿Cómo manejar hipótesis de la forma?

$$\mathbf{H_0} : \ \sigma_X^2 \geq \sigma_Y^2$$
$$\mathbf{H_1} : \ \sigma_X^2 < \sigma_Y^2$$

Criterio Estándar

Usar el Sistema:

$$\mathbf{H_0} : \ \sigma_X^2 = \sigma_Y^2$$
$$\mathbf{H_1} : \ \sigma_X^2 < \sigma_Y^2$$

Motivo

Si H_0 es verdadera $\sigma_X^2 \geq \sigma_Y^2$. El peor error tipo 1 se observa cuando $\sigma_X^2 = \sigma_Y^2$.

Hipótesis Nulas Compuestas

¿Cómo manejar hipótesis de la forma?

$$\mathbf{H_0} : \ \sigma_X^2 \leq \sigma_Y^2$$
$$\mathbf{H_1} : \ \sigma_X^2 > \sigma_Y^2$$

Criterio Estándar

Usar el Sistema:

$$\mathbf{H_0} : \ \sigma_X^2 = \sigma_Y^2$$
$$\mathbf{H_1} : \ \sigma_X^2 > \sigma_Y^2$$

Motivo

Si H_0 es verdadera $\sigma_X^2 \leq \sigma_Y^2$. El peor error tipo 1 se observa cuando $\sigma_X^2 = \sigma_Y^2$.

Hipótesis Nulas Compuestas

Análogamente cualquier otro caso!

Temas