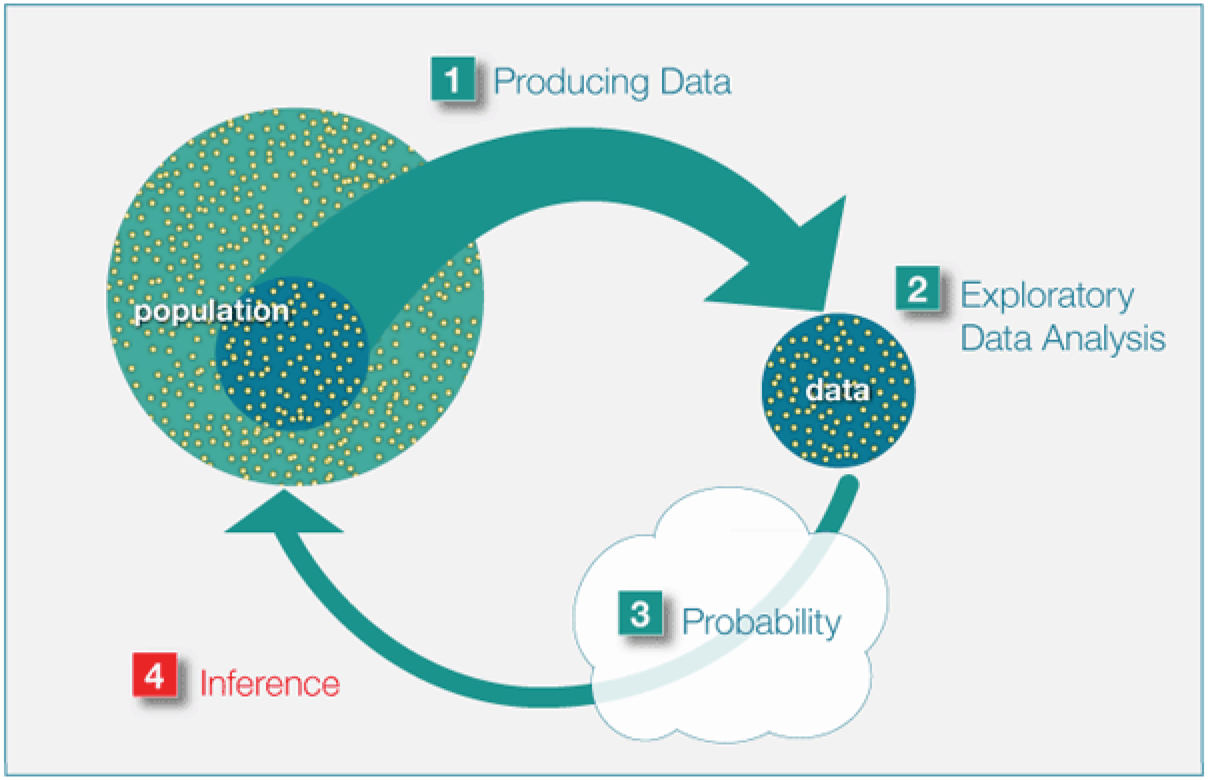
Inferencia Estadística



Profesor: Pedro Saa (pnsaa@uc.cl)
Año: 1-2025

OBJETIVOS DE APRENDIZAJE

- **O1**: Comprender que la estimación puntual y es un estimador de un parámetro poblacional tiene **variabilidad** (error estándar) y la distribución de esta variabilidad construye la **distribución muestral**
- **O2**: Entender que el teorema del límite central es sobre la **distribución muestral** y, que bajo ciertas condiciones, esta es casi **normal**
- **O3**: Definir y explicar que un intervalo de confianza es un **rango plausible de valores** para un **parámetro de población**.
- **O4**: Evaluar cuando es necesario aplicar un test de hipótesis de una o dos colas
- **O5**: Aplicar test de hipótesis para contrastar hipótesis utilizando **intervalos de confianza o valor-p**
- **O6**: Reconocer que toda test de hipótesis conlleva un **error medible (tipo I o tipo II)**

Los **intervalos de confianza** nos permite entregar un **rango de valores** donde estimamos, con un cierto **nivel de confianza**, que el parámetro poblacional se encuentre en él

El error estandar mide la incertidumbre asociado a nuestra **estimación puntal**, también nos provee de una guía para estimar el **intervalo de confianza**

Estimación Puntual

Intervalo de confianza

El **95%** de intervalos de confianza computados **contienen** el promedio poblacional

Test de hipótesis

… también llamado **Contraste de Hipótesis** o **Prueba de Significación**

Caso de estudio

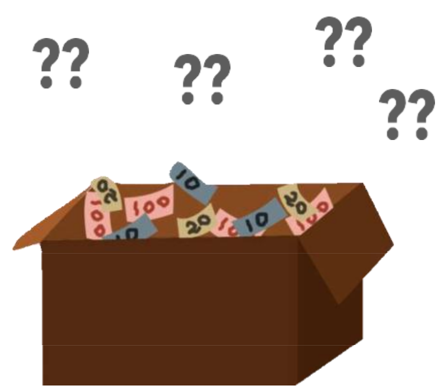
El escéptico y testarudo Dr. Nullsheimer

Planteamiento

Suponga que dos científicos están discutiendo el valor promedio los tickets de una enorme caja que los contiene. Dr. Null (en corto) dice que el promedio es 50, Dr. Alt (en corto) dice que el promedio es distinto a 50.

Luego de mucha discusión deciden ver los datos, toman una muestra al azar de 500 tickets, el promedio resultó ser 48 con una desviación estandar (SD) de 15.3

Dr. Nullsheimer



Dr. Altshuler

Dr. Nullsheimer

Dr. Null: El promedio de muestra es casi 50, tal como te dije que iba ser…

Dr. Alt: Ayy Null, el promedio está realmente bajo 50

Dr. Null: Alt por favor, la desviación es solo 2 y la SD es 15.3, la diferencia es minúscula en relación a las SD, es resultado es simplemente azar

Dr. Alt: Hmm, Null creo que deberías ver la el error estandar (SE) no la SD

Dr. Null: ¿Por qué?

Dr. Alt: Porque la SE nos dice que tan lejos el promedio de la muestra está de su valor esperado, el promedio de la caja

Dr. Altshuler

Dr. Nullsheimer

Dr. Null: Ok ok, ¿cuál es el SE?

Dr. Alt: Estamos de acuerdo que la SD de la caja es 15.3, la SD de los datos?

Dr. Null: Ok si estoy de acuerdo con esto

Dr. Alt: Ok, entonces el SE de la muestra es:

$$SE = \frac{15.3}{\sqrt{500}} \approx 0.7$$

Dr. Null: Ya.. y?

Dr. Alt: El promedio de la muestra es 48, tu dices que debe ser 50, si tu teoría es correcta el promedio de la muestra se encuentra 3 SE bajo el valor esperado

Dr. Null: Espera.. de donde sacaste 3 SE?

Dr. Altshuler

Dr. Alt: Uuff.. volvamos a estadística 1.0:

$$\frac{48 - 50}{0.7} \approx -3$$

Dr. Null: ¿Me estás queriendo decir que 3 SE son demasiados SE para ser explicado meramente por azar?

Dr. Alt: Exactamente, no puedes explicar esta diferencias solamente por azar, la diferencia es real. En otras palabras, el promedio de los tickets de esa caja no es 50, es otro número

Dr. Nullsheimer

Dr. Null: Ok, tienes razón, rechazo la idea que el promedio es 50…

Dr. Altshuler

El marco formal de test de hipótesis consiste en la mirada escéptica formulada como **hipótesis nula (H_0)** y aquella mirada que propone un cambio llamada **hipótesis alternativa (H_A o H_1)**

H_0: El valor promedio de los tickets en la caja es 50 $\mu_1 = 50$

H_1: El valor promedio de los tickets en la caja es distinto a 50 $\mu_1 \neq 50$

El escéptico (investigador) no va abandonar **H_0** a menos que la evidencia a favor de **H_1** sea lo suficientemente grande para rechazar **H_0** en favor de **H_1**

La hipótesis alternativa, según información previa, puede ser formulada como superior ($\mu_1 < \mu_2$), inferior ($\mu_1 > \mu_2$) o diferente ($\mu_1 \neq \mu_2$) al grupo a investigar

La hipótesis nula corresponde a la idea que la diferencia observada se debe al **azar**, en cambio, la hipótesis alternativa corresponde a la idea que la diferencia observada es **real**

La evidencia apunta a rechazar o no la hipótesis nula en favor de la alternativa, sin embargo, no implica que la hipótesis nula sea **verdadera**

Las hipótesis son siempre sobre los **parámetros de población** y no sobre los **parámetros de muestra**

Test de hipótesis usando intervalos de confianza

La idea detrás es definir el área que entendemos como aceptación o rechazo de H_0 mediante a un nivel de confianza usualmente 95%

```
n = 500
prom_muestral = 48
des_est = 15.3

err_std = np.round(des_est / np.sqrt(n), 1) # redondeamos
print(f'Error estándar: {err_std}') #0.7
z_critico = st.norm.ppf(0.975) # 1-alpha/2
corte_inferior = prom_muestral - z_critico * err_std # 46.63
corte_superior = prom_muestral + z_critico * err_std # 49.37
```

$$46.63 < \mu < 49.37$$

Dado que el valor nulo ($\mu 1 = 50$) no se encuentra en el intervalo, rechazamos H_0

Si bien este método es efectivo, no te dice nada de la probabilidad de ocurrencia de los resultados bajo la hipótesis nula

Test de hipótesis usando estadística de prueba Z

Esta fue la estrategia que ocupó Dr. Altshuler para rechazar la hipótesis del Dr. Nullsheimer. Puesto en palabras formales:

$$P(resultados \mid H_0 \ verdadera)$$

$$P(\bar{x} = 48 \mid H_0 : \ \mu = 50)$$

Cumpliendo El teorema del límite central $\rightarrow$

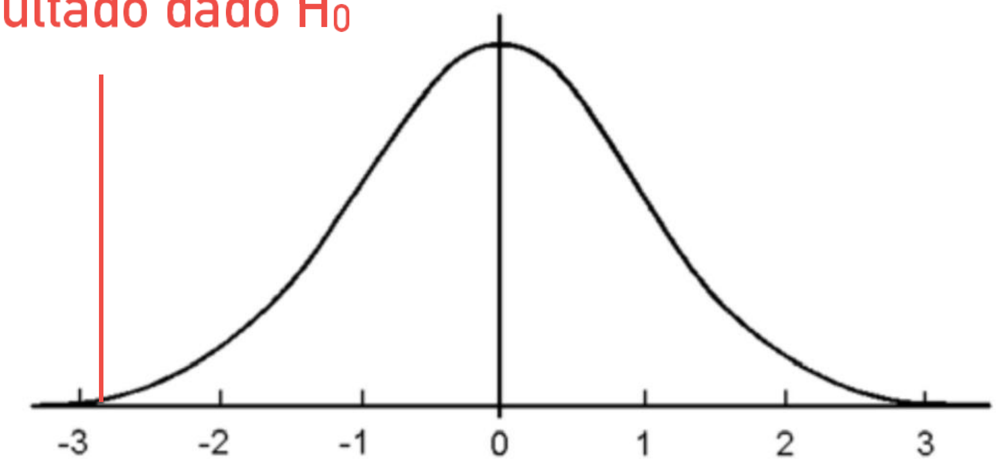
$$\bar{x} \sim N(\mu = 3, \ SE = 0.7)$$

Estadística de prueba Z

$$z = \frac{48 - 50}{0.7} = -2.86$$

$$p - value = P(z = -2.86) = 0.0021$$

Resultado dado H_0



Nivel de significancia (α)

El valor-p cuantifica el peso de la evidencia en contra la hipótesis nula, cuando el valor-p es inferior al **nivel de significancia (α)** rechazamos la hipótesis nula

El **nivel de significancia** se ha definido arbitrariamente en **$\alpha = 5\%$**, esto dice que es improbable ver un resultado que sucede 1 de cada 20 observaciones

El nivel de significancia está tradicionalmente establecido en 5%, pero dependiendo de la disciplina o del caso este puede ser menor, por ejemplo, en medicina se suele utilizar 1%

¿Cómo interpretamos el resultado?

Coloquial

Si el promedio de los tickets de la caja es 48, existe 0.21% de probabilidad que una muestra aleatoria de 500 tickets tenga un promedio de 50. Por tanto, dado que la probabilidad es tan baja (**$p < 0.05$**) decimos que el resultado es **estadísticamente significativo**, por ende rechazamos la hipótesis nula, no podemos concluir que el resultado obtenido es simplemente por azar.

Estadística

Con una prueba estadística z de 2.86 y un valor-p de 0.0021, rechazamos la hipótesis nula a un 5% nivel de significancia. Concluimos que el promedio del valor de los tickets en la caja es distinto a 50.

Prueba de dos o una cola

Pruebas de una cola vs dos colas

El ejemplo revisado hasta ahora se uso el método de dos colas debido a como fue formulada la hipótesis alternativa:

H_1: El valor promedio de los tickets en la caja es distinto a 50 **$\mu_1 \neq 50$**

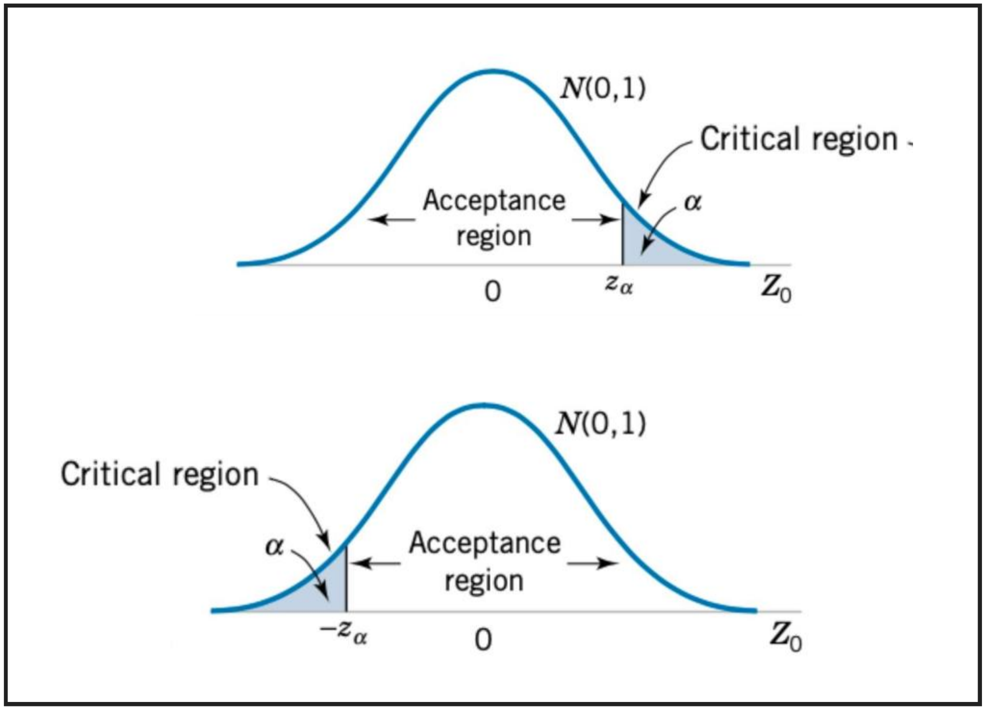
Dado que el promedio pudo haber sido superior o inferior, revisamos la probabilidad que esto ocurra en ambos sentidos para ello en casos debo compararlo con un nivel de significancia: **$\alpha = 5\%/2$** o multiplicar la probabilidad obtenida por 2 y compararla con **α**

$$p - value(\text{una cola}) = P(z = -2.86) = 0.0021$$

$$p - value(\text{dos colas}) = 2 \cdot P(z = -2.86) = 0.0042$$

Pruebas de una cola vs dos colas

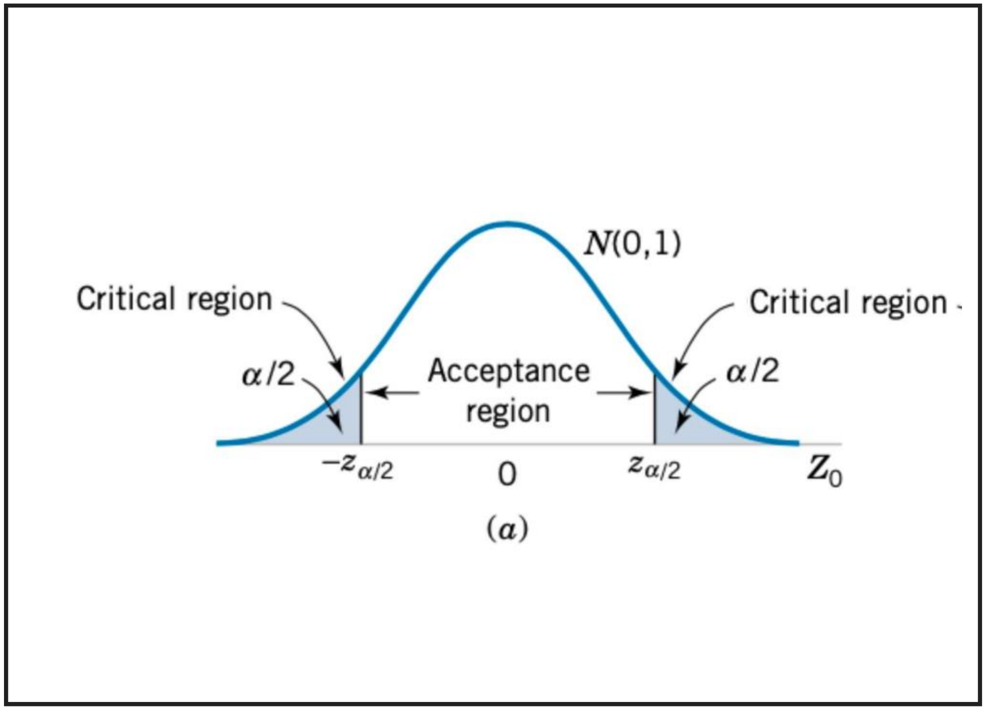
Prueba de una cola



H$_1$: El valor promedio de los tickets en la caja es inferior a 50

$\mu_1 < 50$

Prueba de dos colas



H$_1$: El valor promedio de los tickets en la caja es distinto a 50

$\mu_1 \neq 50$

Caso de estudio

Diferencias entre horas de sueño de estudiantes rurales vs promedio país

Estudio: Una encuesta realizada por la Fundación Nacional del Sueño encontró que los estudiantes duermen, en promedio, 7 horas por noche. Investigadores al ver esta encuesta deciden poner a prueba esta hipótesis, ellos afirman que los estudiantes en colegios rurales duermen más que el promedio nacional. Ellos tomaron una muestra aleatoria de 100 alumnos de colegios rurales, los resultados están en el objeto **sleep_df** Realice una test de hipótesis y exponga sus resultados.

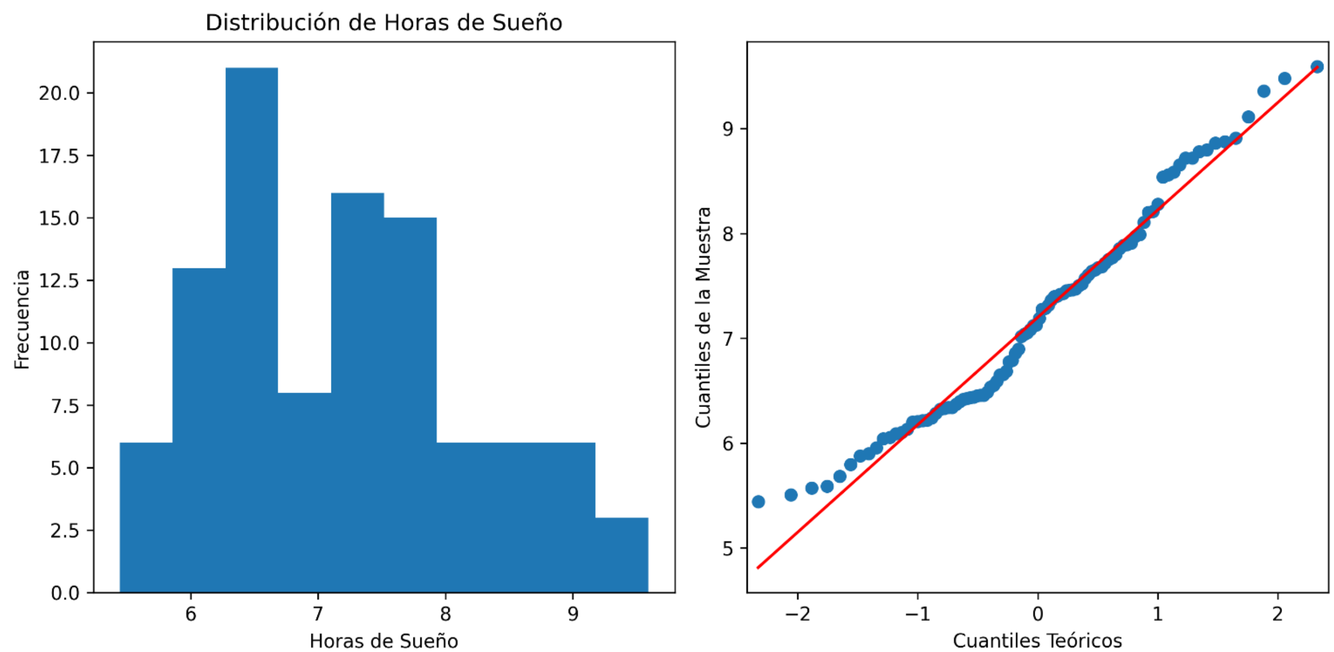
```python
# Realizamos un test de hipotesis
# Hipotesis
# H0: mu = 7
# Ha: mu > 7

# Calculamos el estimador puntual (promedio y error estandar)
alpha = 0.05
n = horas_sueno.shape[0] # 100
prom_muestra = np.mean(horas_sueno) # 7.1998
sd_muestra = np.std(horas_sueno) # 0.9968
se_muestra = sd_muestra / np.sqrt(n) # 0.09968
```

```python
# Independencia: Dado que fue una muestra aleatorioa y menor al 10% de la población
# Podemos asumir independencia
# Normalidad: Se puede ver en el siguiente gráfico

fig, ax = plt.subplots(1,2, figsize=(10,5), dpi = 300)
sm.qqplot(horas_sueno, line='r', ax=ax[1])
ax[1].set_xlabel('Cuantiles Teóricos')
ax[1].set_ylabel('Cuantiles de la Muestra')
ax[0].hist(horas_sueno, bins = 10)
ax[0].set_title('Distribución de Horas de Sueño')
ax[0].set_xlabel('Horas de Sueño')
ax[0].set_ylabel('Frecuencia')
fig.tight_layout()
plt.show()
```

Los datos muestran se muestran levemente asimétricos, pero dado el número de datos (100), es suficientemente grande ($n>30$) podemos asumir normalidad



```
valorz = (prom_muestra - 7)/se_muestra # 2.004446
valorp = 1-st.norm.cdf(valorz, loc = 0, scale = 1) #  0.02251
valorp < alpha # Verdadero
```

Con una prueba estadística Z de 2.0044 y un valor-p = 0.0225, rechazamos la hipótesis nula con un 5% de nivel de significancia. Concluimos que hay suficiente evidencia estadística para decir que estudiantes de colegios rurarles duermen más que el promedio nacional.

Resumen test de hipótesis para un promedio de población

1. **Establecer la hipótesis:**
   - H_0: Hipótesis nula (μ = valor nulo)
   - H_1: Hipótesis alternativa ($\mu >$ o $<$ o $\neq$ el valor nulo)
2. **Calcular el estimador puntual ($\bar{x}$)**
3. **Verificar supuestos y condiciones**
   - Independencia: muestreo o asignación aleatoria
   - Normalidad: histogramas casi normales o también revisar qqplot sino que $n > 30$
4. **Calcular una prueba estadística correspondiente y valor-p**
5. **Tomar una decisión e interpretarla en contexto.**
   - Si valor-p $< \alpha$, rechazo H_0, datos proveen evidencia de H_1
   - Si valor-p $> \alpha$, no rechazo H_0, datos no proveen evidencia de H_1

Errores de decisión

Test de hipótesis **no son infalibles**, es posible **cometer errores de decisión** incluso cuando se siguen correctamente todos los lineamientos de la prueba

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	✔	Type I error
	H_A true	Type 2 error	✔

Se define como **error tipo 1** cuando rechazas la hipótesis nula cuando en realidad era verdadera (**Falso Positivo**)

Se define como **error tipo 2** cuando no rechazas la hipótesis nula cuando en realidad la hipótesis alternativa era verdadera (**Falso negativo**)

En la realidad jamas sabemos si H_0 o H_1 es verdad pero es necesario considerar todas las posibilidades

Otra forma de ver los errores es homologarlo con la **dinámica de un juicio**

H_0: **El acusado es inocente**

H_1: **El acusado es culpable**

¿Qué tipo de errores son en cada caso?

Declarar el acusado inocente cuando en verdad era culpable. **ERROR TIPO 2**

Declarar el acusado culpable cuando en verdad era inocente. **ERROR TIPO 1**

¿Qué tipo de error es peor?

En general esto es bastante subjetivo y depende mucho del tipo de problema que se está analizando, sin embargo, hoy en día existe un consenso en reducir la tasa de error tipo 1.

William Blackstone

"Better that ten guilty persons escape than that one innocent suffer"

La clave es que uno tiene **el control de cada uno de los errores**, sin embargo, es difícil reducir ambos errores (reducir uno generalmente aumenta el otro)

goal: keep α and β low

		Decision	
		fail to reject H_0	reject H_0
Truth	H_0 true	$1 - \alpha$	Type I error, α
	H_A true	Type 2 error, β	$1 - \beta$

Se define como **error tipo 1** cuando rechazas la hipótesis nula cuando en realidad era verdadera, la probabilidad que ocurra es **α** (nivel de significancia)

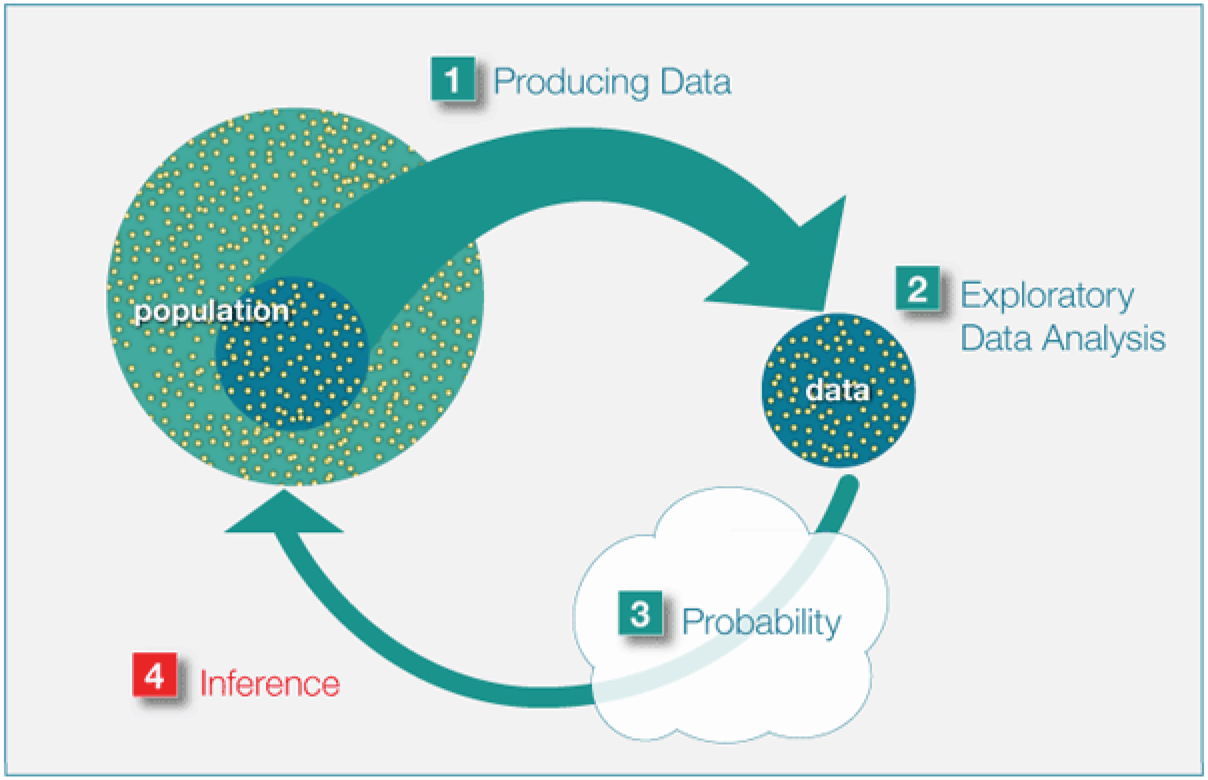
Se define como **error tipo 2** cuando no rechazas la hipótesis nula cuando en realidad la hipótesis alternativa era verdadera, la probabilidad que ocurra es **β**

Se define como el **poder de una prueba** como la probabilidad de correctamente rechazar H_0, esta se calcula como **$1 - \beta$**

Resumen

- Introducimos formalmente la metodología para realizar **tests de hipótesis de una y dos colas**.
- **A la raíz de esta metodología se encuentra la determinación del intervalo de confianza** para el estadístico de interés.
- La aplicación de los **tests de hipótesis** requiere la **definición del nivel de significancia** (confianza) el que determina la probabilidad de **error tipo I.**
- Por otro lado, **el error tipo II está asociado al poder estadístico del contraste de hipótesis.**
- En la práctica, **es difícil en general reducir la probabilidad de error tipo I sin aumentar la probabilidad de error tipo II.**

Inferencia Estadística



Profesor: Pedro Saa (pnsaa@uc.cl)
Año: 1-2025