Test de Hipótesis

Estadística Computacional

Juan Zamora Osorio
juan.zamora@pucv.cl

Instituto de Estadística
Pontificia Universidad Católica de Valparaíso

15 de noviembre de 2023

Test de Hipótesis

Hemos aprendido sobre...

- Describir datos.
- Probabilidades.
- Variables aleatorias.
- Inferencia estadística (paramétros).

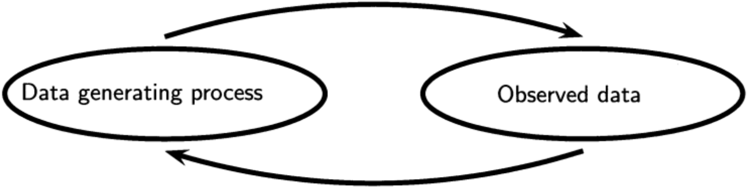
Objetivo

- Comparar *hipótesis* o afirmaciones sobre una característica de una población.
- Características:
  - Valor de un parámetro de la población.
  - Valor de varios parámetros de la población.
  - Forma de la distribución de un parámetro de la población o de la población misma.

Recordar

Probabilidades

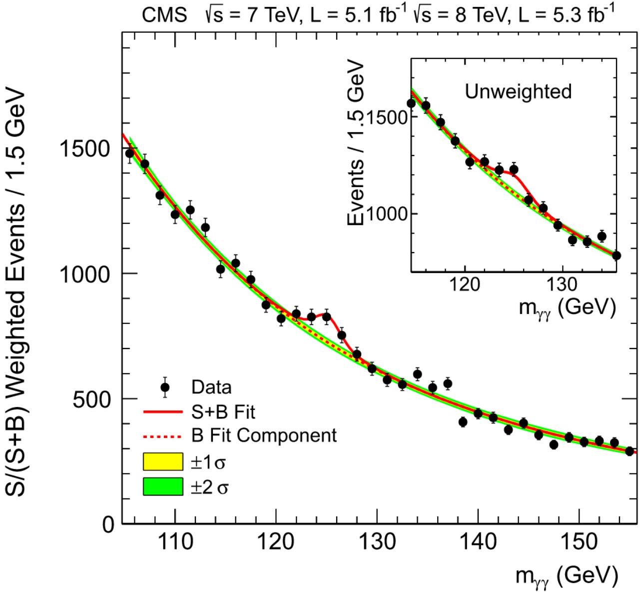
- ¿Dado un proceso que genera datos, cuáles son las propiedades que observaremos?



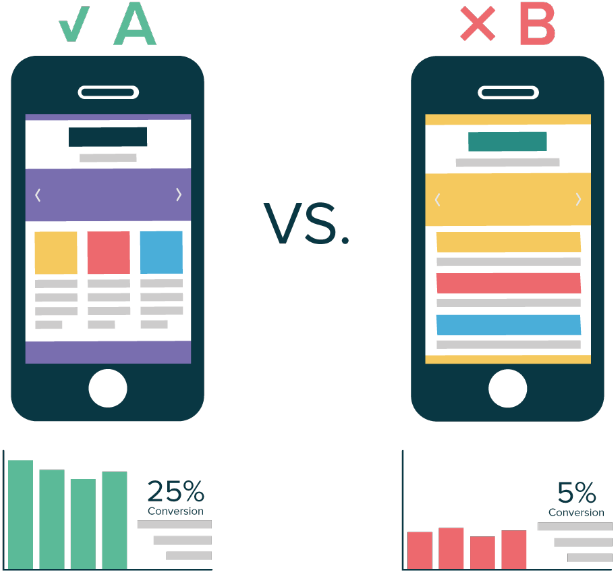
Inferencia estadística

- ¿Dadas las observaciones, qué podemos decir sobre el proceso que genera los datos?

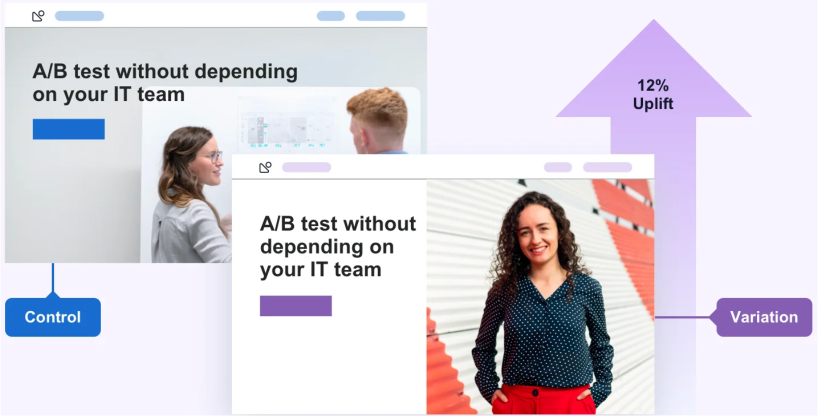
Ejemplo: Descubrimiento Bosson de Higgs



Ejemplo: A/B testing



Ejemplo: A/B testing



Hipótesis

Hipótesis nula H_0

- Una afirmación supuesta como verdad hasta ahora.
- Supuesto a priori.

Hipótesis alternativa H_1 o H_a

- Una afirmación distinta a H_0.
- Generalmente H_1 contradice H_0.

Test de hipótesis

- Hipótesis H_0 debe ser *falsable*.
- Se busca rechazar H_0 en favor de H_1.
- Respuestas: se rechaza, o no se puede rechazar.

Hipótesis

Cuidado

- ¡Rechazar no es lo mismo que aceptar!
- Muchos tests se basan en rechazar o no una hipótesis, sin mirar la alternativa.

Cuidado

- Generalmente es más útil saber qué hipótesis es más plausible (contraste de hipótesis).
- No veremos relaciones de causalidad en este curso.

Incertidumbre

- ¿Podemos cuantificar el grado de rechazo de una hipótesis?
- Si conseguimos más datos, ¿podemos responder?

Hipótesis – región crítica

Recordar: Estadístico

- Sea X una muestra iid univariada.
- Recordar que un estadístico es una función de la muestra.
- El estadístico tiene una distribución que depende de la muestra.

Regla de decisión

- Si el estadístico cae en la *región crítica*, se rechaza H_0.

Ejemplo

- $H_0 =$ la altura media de lo/as estudiantes del departamento de informática es mayor a 1,65 m.
- Se mide la altura de 25 estudiantes de este curso.
- Se calcula el promedio (media muestral $\bar{X}_{25}$).
- La región crítica es el rango de valores de $\bar{X}_{25}$ donde H_0 se rechaza.

Hipótesis – errores

Error de tipo I

- Se rechaza la hipótesis nula cuando en verdad es cierta.
- Se suele llamar *nivel de significancia*.

Error de tipo II

- No se rechaza la hipótesis nula cuando en verdad es falsa.

Hipótesis – errores

Ejemplo

- Supongamos que la altura sigue distribución normal $\mathcal{N}(\mu, 0{,}01)$.
- Supongamos que rechazamos si $\bar{X}_{25} < 1{,}70$.
- Supongamos que μ tiene distribución a priori uniforme entre 0 y 3

(Continuación del ejemplo...)

Probabilidad de error de tipo I

- Por teorema del límite central, sabemos que $\bar{X}_{25} \sim \mathcal{N}\left(\mu, \frac{0{,}01}{25}\right)$.

$$P(\text{error de tipo I}) = P(\text{rechazar } H_0 \mid H_0 \text{ es cierta})$$

$$= P\left(\bar{X}_n < 1{,}70 \mid \mu > 1{,}65\right) = \frac{P\left(\bar{X}_n < 1{,}70, \mu > 1{,}65\right)}{P(\mu > 1{,}65)}$$

$$= \frac{\int_{1{,}65}^{3} \Phi\left(5\frac{1{,}7-\mu}{0{,}1}\right) p(\mu) d\mu}{\int_{1{,}65}^{3} p(\mu) d\mu} = \frac{1}{3 - 1{,}65} \int_{1{,}65}^{3} \Phi(50(1{,}7 - \mu)) d\mu \approx 0{,}037.$$

Hipótesis – errores

Ejemplo

- Supongamos que la altura sigue distribución normal $\mathcal{N}(\mu, 0{,}01)$.
- Supongamos que rechazamos si $\bar{X}_{25} < 1{,}70$.
- Supongamos que μ tiene distribución a priori uniforme entre 0 y 3

(Continuación del ejemplo...)

Probabilidad de error de tipo II

- Por teorema del límite central, sabemos que $\bar{X}_{25} \sim \mathcal{N}\left(\mu, \frac{0{,}01}{25}\right)$.

$$P(\text{error de tipo I}) = P(\text{no rechazar } H_0 \mid H_0 \text{ no es cierta})$$

$$= P\left(\bar{X}_n \geq 1{,}70 \mid \mu \leq 1{,}65\right) = \frac{\int_0^{1{,}65} \left(1 - \Phi\left(5\frac{1{,}7-\mu}{0{,}1}\right)\right) p(\mu) d\mu}{\int_0^{1{,}65} p(\mu) d\mu}$$

$$= \frac{1}{1{,}65 - 0} \int_0^{1{,}65} 1 - \Phi(50(1{,}7 - \mu)) d\mu \approx 0{,}000024.$$

Hipótesis – errores

Errores

- Tipo I: se rechaza la hipótesis nula cuando en verdad es cierta.
- Tipo II: no se rechaza la hipótesis nula cuando en verdad es falsa.

Ejemplo

- $H_0 =$ la altura media de lo/as estudiantes del departamento de informática es mayor a 1,65 m.
  - $P(\text{error de tipo I}) = P(\text{rechazar } H_0 \mid H_0 \text{ es cierta}) \approx 0{,}037$.
  - $P(\text{error de tipo II}) = P(\text{no rechazar } H_0 \mid H_0 \text{ no es cierta}) \approx 0{,}000024$.
- Ambas dependen del umbral de corte t. En el ejemplo, $t = 1{,}7$.
- Podemos variar t de manera de reducir alguno de los dos.
  - Si t aumenta, el error de tipo I aumenta y el error de tipo II disminuye.
  - Si t disminuye, el error de tipo I disminuye y el error de tipo II aumenta.

Ejemplo: recuperación de enfermedad

Probando un medicamento

- Se sabe que si una persona contrae una enfermedad, tiene probabilidad $p = 0{,}25$ de sobrevivir luego de 3 meses.
- Se experimenta un nuevo remedio en 20 personas recién contagiadas.
- Luego de 3 meses se revisa cuántos sobrevivieron.
- La hipótesis nula H_0 es que el remedio no afecta, por lo que la nueva probabilidad de sobrevivir luego de 3 meses es $p = 0{,}25$.
- La hipótesis alternativa H_1 es que la probabilidad es mayor, $p > 0{,}25$.
- ¿Cuántas personas deberían sobrevivir para rechazar H_0?

Ejemplo: recuperación de enfermedad

Probando un medicamento

- Se sabe que si una persona contrae una enfermedad, tiene probabilidad $p = 0{,}25$ de sobrevivir luego de 3 meses.
- Se experimenta un nuevo remedio en 20 personas recién contagiadas.
- La hipótesis nula H_0 es que el remedio no afecta, por lo que la nueva probabilidad de sobrevivir luego de 3 meses es $p = 0{,}25$.
- La hipótesis alternativa H_1 es que la probabilidad es mayor, $p > 0{,}25$.
- ¿Cuántas personas deberían sobrevivir para rechazar H_0?

Desarrollo

- Estadístico: X cantidad de personas que sobrevivieron.
- X sigue una distribución binomial $B(n, p)$.

$$\begin{aligned} P(\text{error de tipo I}) &= P(\text{rechazar } H_0 \mid H_0 \text{ es cierta}) \\ &= P(X \geq t \mid p = 0{,}25) = 1 - F_{X \sim B(20, 0{,}25)}(t-1). \end{aligned}$$

Ejemplo: recuperación de enfermedad

Probando un medicamento

- Se sabe que si una persona contrae una enfermedad, tiene probabilidad $p = 0{,}25$ de sobrevivir luego de 3 meses.
- Se experimenta un nuevo remedio en 20 personas recién contagiadas.
- La hipótesis nula H_0 es que el remedio no afecta, por lo que la nueva probabilidad de sobrevivir luego de 3 meses es $p = 0{,}25$.
- La hipótesis alternativa H_1 es que la probabilidad es mayor, $p > 0{,}25$.
- ¿Cuántas personas deberían sobrevivir para rechazar H_0?

Desarrollo

- Estadístico: X cantidad de personas que sobrevivieron.
- X sigue una distribución binomial $B(n, p)$.

$$\begin{aligned} P(\text{error de tipo II}) &= P(\text{no rechazar } H_0 \mid H_0 \text{ no es cierta}) \\ &= P(X < t \mid p \neq 0{,}25) = F_{X \sim B(20,p)}(t-1). \end{aligned}$$

Ejemplo: recuperación de enfermedad

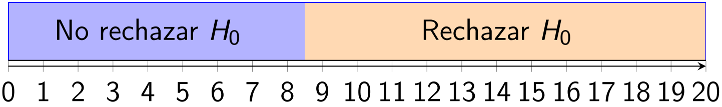
¿Cuál punto de corte t para rechazar?

- $P(\text{error de tipo I}) = P(\text{rechazar } H_0 \mid H_0 \text{ es cierta}) = 1 - F_{X\sim B(20,0{,}25)}(t-1)$.
- $P(\text{error de tipo II}) = P(\text{no rechazar } H_0 \mid H_0 \text{ no es cierta}) = F_{X\sim B(20,p)}(t-1)$.

Ejemplo: $t = 9$

- $P(\text{error de tipo I}) = P(\text{rechazar } H_0 \mid H_0 \text{ es cierta}) \approx 0{,}0409$.
- $P(\text{error de tipo II}) = P(\text{no rechazar } H_0 \mid H_0 \text{ no es cierta}) =$

p	0,25	0,3	0,5	0,75
$P(\text{error de tipo II})$	0,9591	0,8867	0,2517	0,0009



Ejemplo: recuperación de enfermedad

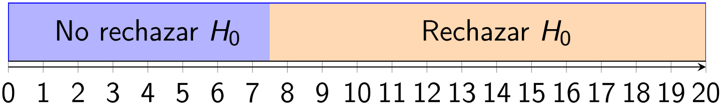
¿Cuál punto de corte t para rechazar?

- $P(\text{error de tipo I}) = P(\text{rechazar } H_0 \mid H_0 \text{ es cierta}) = 1 - F_{X \sim B(20, 0{,}25)}(t-1)$.
- $P(\text{error de tipo II}) = P(\text{no rechazar } H_0 \mid H_0 \text{ no es cierta}) = F_{X \sim B(20,p)}(t-1)$.

Ejemplo: $t = 8$

- $P(\text{error de tipo I}) = P(\text{rechazar } H_0 \mid H_0 \text{ es cierta}) \approx 0{,}1018$.
- $P(\text{error de tipo II}) = P(\text{no rechazar } H_0 \mid H_0 \text{ no es cierta}) =$

p	0,25	0,3	0,5	0,75
$P(\text{error de tipo II})$	0,8982	0,7723	0,1316	0,0002



Ejemplo: peso de recién nacido/as

Comparando países

- Se sabe que recién nacidos en Inglaterra tienen peso con media 3 kg y desviación estándar de 0,5 kg.
- Se cree que la media de peso de recién nacidos en Australia es mayor.
- H_0: la media de peso en Australia es similar a la de Inglaterra, 3 kg.
- H_1: la media de peso en Australia es significativamente mayor a 3 kg.

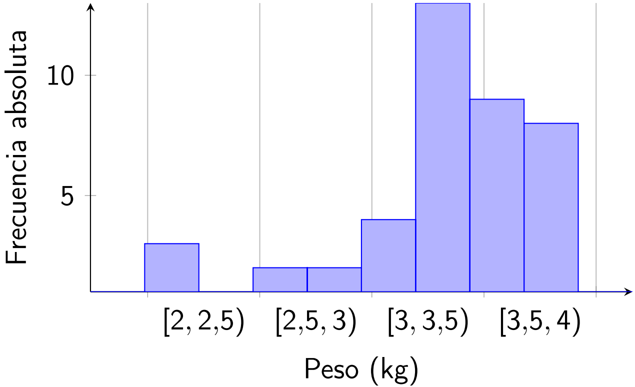
Desarrollo

- $H_0 : \mu = 3$ kg.
- $H_1 : \mu > 3$ kg.
- Se obtiene una muestra de 44 recién nacidos (conjunto *Babyboom*).

Ejemplo: peso de recién nacido/as

44 recién nacidos: conjunto *Babyboom*

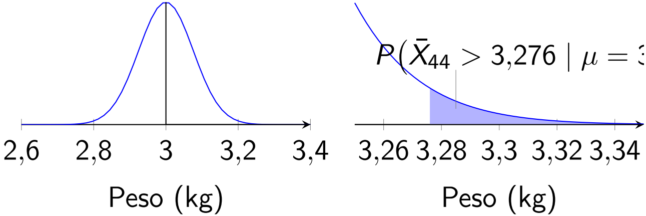
- Media muestral $\bar{x}_{44} \approx 3{,}276$.
- Si H_0 es cierta, $P(\text{error de tipo I}) = P\left(\bar{X}_{44} > 3{,}276 \mid \mu = 3\right)$



Ejemplo: peso de recién nacido/as

44 recién nacidos: conjunto *Babyboom*

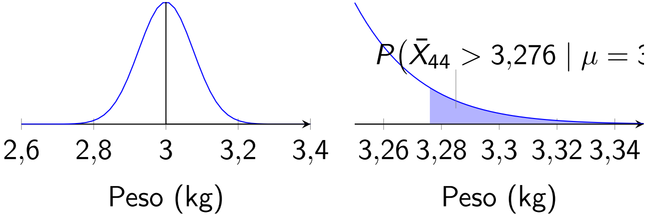
- Media muestral $\bar{x}_{44} \approx 3{,}276$.
- Si H_0 es cierta, $P(\text{error de tipo I}) = P\big(\bar{X}_{44} > 3{,}276 \mid \mu = 3\big)$
- Por teorema del límite central,
  $\bar{X}_{44} \sim \mathcal{N}\Big(\mu, \frac{\sigma^2}{n}\Big) = \mathcal{N}\Big(3, \frac{0{,}25}{44}\Big)$.
- Entonces,
  $P\big(\bar{X}_{44} > 3{,}276 \mid \mu = 3\big) = 1 - F_{\bar{X}_{44} \sim \mathcal{N}\big(3, \frac{1}{176}\big)}(3{,}276)$.
- Es decir, $P\big(\bar{X}_{44} > 3{,}276 \mid \mu = 3\big) \approx 0{,}000125$.



Ejemplo: peso de recién nacido/as

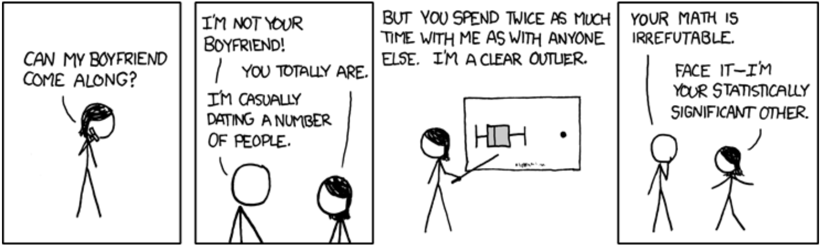
p-valor

- Media muestral $\bar{x}_{44} \approx 3{,}276$.
- $P\big(\bar{X}_{44} > 3{,}276 \mid \mu = 3\big) \approx 0{,}000125$ es llamado *p-valor*.
- El p-valor indica la probabilidad de encontrar un resultado como el obtenido si H_0 es cierta.



Nivel de significancia

- Generalmente se elige de antemano un α máximo.


CAN MY BOYFRIEND COME ALONG?
I'M NOT YOUR BOYFRIEND!
YOU TOTALLY ARE.
I'M CASUALLY DATING A NUMBER OF PEOPLE.
BUT YOU SPEND TWICE AS MUCH TIME WITH ME AS WITH ANYONE ELSE. I'M A CLEAR OUTLIER.
YOUR MATH IS IRREFUTABLE.
FACE IT—I'M YOUR STATISTICALLY SIGNIFICANT OTHER.

Varianza desconocida

Comparando países

- Se sabe que recién nacidos en Inglaterra tienen peso con media 3 kg.
- Se cree que la media de peso de recién nacidos en Australia es mayor.
- H_0: la media de peso en Australia similar a Inglaterra, 3 kg.
- H_1: la media de peso en Australia significativamente $>$ 3 kg.

Desarrollo

- $H_0 : \mu = 3$ kg.
- $H_1 : \mu > 3$ kg.
- Se obtiene una muestra de 44 recién nacidos (conjunto *Babyboom*).
- $\bar{x}_{44} \approx 3{,}276$.
- $\sqrt{s_{44}^2} \approx 0{,}5280$.

Varianza desconocida

Recuerdo

- La variable aleatoria $T = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sqrt{S_n^2}}$ tiene distribución t de Student con $n-1$ grados de libertad.

Desarrollo

- $\bar{x}_{44} \approx 3{,}276$.
- $\sqrt{s_{44}^2} \approx 0{,}5280$.
- $t = \frac{\sqrt{44}(3{,}276 - \mu)}{0{,}5280} \approx 3{,}4674$
- $P(\text{error de tipo I}) = P(T > t \mid \mu = 3) \approx P(T > 3{,}4674)$.
- Es decir, $P(\text{error de tipo I}) = 1 - P(T \leq 3{,}4674) \approx 0{,}0006$.
- Si se considera $\alpha = 0{,}01$, se rechaza H_0.

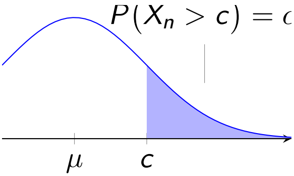
Ejemplo: peso de recién nacido/as

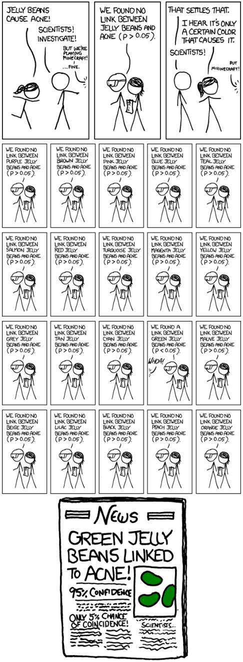
p-valor

- En el caso de varianza conocida, p-valor es $0{,}00013$.
- En el caso de varianza desconocia, p-valor es $0{,}0006$.
- Distribución t de Student tiene colas anchas, en comparación con normal.

Definir región crítica

- Se podría elegir nivel de significancia α de antemano.
- Luego calcular para qué valor de $\bar{X}_n$ se rechaza la hipótesis.





JELLY BEANS CAUSE ACNE!
SCIENTISTS! INVESTIGATE!
BUT WE'RE PLAYING MINECRAFT!
... FINE.
WE FOUND NO LINK BETWEEN JELLY BEANS AND ACNE (P > 0.05).
THAT SETTLES THAT.
I HEAR IT'S ONLY A CERTAIN COLOR THAT CAUSES IT.
SCIENTISTS!
BUT MIIIINECRAFT!
WE FOUND NO LINK BETWEEN PURPLE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BROWN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PINK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLUE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TEAL JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN SALMON JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN RED JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TURQUOISE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN MAGENTA JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN YELLOW JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN GREY JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN TAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN CYAN JELLY BEANS AND ACNE (P > 0.05).
WE FOUND A LINK BETWEEN GREEN JELLY BEANS AND ACNE (P < 0.05).
WHOA!
WE FOUND NO LINK BETWEEN MAUVE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BEIGE JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN LILAC JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN BLACK JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN PEACH JELLY BEANS AND ACNE (P > 0.05).
WE FOUND NO LINK BETWEEN ORANGE JELLY BEANS AND ACNE (P > 0.05).
NEWS
GREEN JELLY BEANS LINKED TO ACNE!
95% CONFIDENCE
ONLY 5% CHANCE OF COINCIDENCE!
SCIENTISTS...

Bilateral (*two sided*)

Comparando países

- Se sabe que recién nacidos en Inglaterra tienen peso con media 3 kg.
- Se cree que media de peso de recién nacidos en Australia es distinta.
- H_0: la media de peso en Australia es similar a la de Inglaterra, 3 kg.
- H_1: la media de peso en Australia es significativamente distinta a 3 kg.

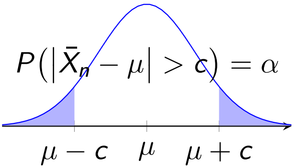
Desarrollo

- $H_0 : \mu = 3$ kg.
- $H_1 : \mu \neq 3$ kg.
- Se obtiene una muestra de 44 recién nacidos (conjunto *Babyboom*).
- $\bar{x}_{44} \approx 3{,}276$.
- $\sqrt{s_{44}^2} \approx 0{,}5280$.

Bilateral (*two sided*)

Desarrollo

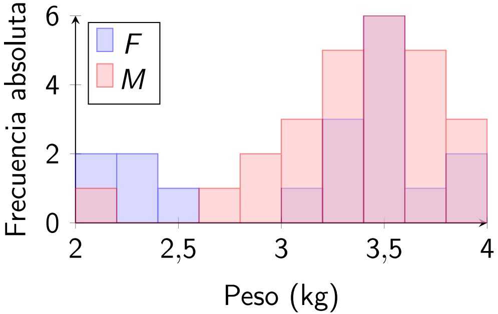
- $H_0 : \mu = 3$ kg.
- $H_1 : \mu \neq 3$ kg.
- $\bar{x}_{44} \approx 3{,}276$, $\sqrt{s^2_{44}} \approx 0{,}5280$.
- Caso varianza desconocida: $t = \frac{\sqrt{44}(3{,}276-\mu)}{0{,}5280} \approx 3{,}4674$.
- $P(|T| > t) = P(T \leq -t) + P(T > t) = 1 + P(T \leq -t) - P(T \leq t)$.
- Así, $P(|T| > t) = 2P(T \leq -|t|) \approx 0{,}0012$.



Ejemplo: peso de recién nacido/as

44 recién nacidos: conjunto *Babyboom*

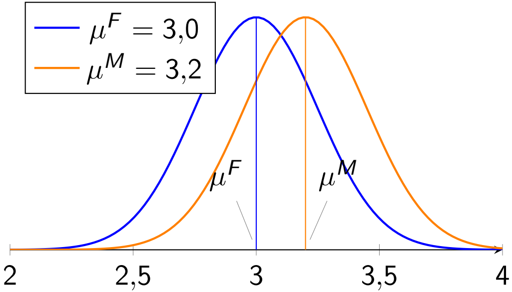
- El peso se puede diferenciar por género (masculino M o femenino F).
- 18 son femenino F, 26 masculino M.
- Pesos promedio $\bar{x}_{18}^{F} \approx 3{,}132$, $\bar{x}_{26}^{M} \approx 3{,}375$.
- ¿Son las medias distintas significativamente?



Comparación de poblaciones

44 recién nacidos: conjunto *Babyboom*

- El peso se puede diferenciar por género (masculino M o femenino F).
- 18 son femenino F, 26 masculino M.
- Pesos promedio $\bar{x}_{18}^{F} \approx 3{,}132$, $\bar{x}_{26}^{M} \approx 3{,}375$.
- ¿Son las medias distintas significativamente?



Comparación de poblaciones

Desarrollo

- $H_0 : \mu^M = \mu^F$.
- $H_1 : \mu^M \neq \mu^F$.
- Pesos promedio $\bar{x}^F_{18} \approx 3{,}132$, $\bar{x}^M_{26} \approx 3{,}375$.

Idea

- Podemos contrastar la diferencia $\mu^M - \mu^F$ y $\bar{X}^M_n - \bar{X}^F_m$.
- $H_0 : \mu^M - \mu^F = 0$.
- $H_1 : \mu^M - \mu^F \neq 0$.
- Diferencia de promedios $\bar{x}^M_{26} - \bar{x}^F_{18} \approx 0{,}2429$.
- Si H_0 es cierta, ¿cómo se distribuye $\bar{X}^M_n - \bar{X}^F_m$?

Comparación de poblaciones

Idea

- $H_0 : \mu^M - \mu^F = 0$.
- $H_1 : \mu^M - \mu^F \neq 0$.
- Diferencia de promedios $\bar{x}_{26}^M - \bar{x}_{18}^F \approx 0{,}2429$.
- Si H_0 es cierta, ¿cómo se distribuye $\bar{X}_n^M - \bar{X}_m^F$?

Caso varianzas conocidas

- $\bar{X}_n^M - \bar{X}_m^F \sim \mathcal{N}\left(\mu^M - \mu^F, \frac{\sigma^{2M}}{n} + \frac{\sigma^{2F}}{m}\right)$
- Se calcula $z = \frac{\bar{X}_n^M - \bar{X}_m^F - \mu^M + \mu^F}{\sqrt{\frac{\sigma^{2M}}{n} + \frac{\sigma^{2F}}{m}}}$.
- Se revisa p-valor $P(|Z| > z)$.

Comparación de poblaciones

En el ejemplo

- Diferencia de promedios $\bar{x}_{26}^{M} - \bar{x}_{18}^{F} \approx 0{,}2429$.
- Supongamos $\sigma^{2M} = 0{,}5$ y $\sigma^{2F} = 0{,}5$.
- $z = \frac{\bar{X}_n^M - \bar{X}_m^F - \mu^M + \mu^F}{\sqrt{\frac{\sigma^{2M}}{n} + \frac{\sigma^{2F}}{m}}} \approx 1{,}1201$.
- p-valor $P(|Z| > z) \approx 0{,}2627$.

Comparación de poblaciones

Caso varianzas desconocidas: *Welch t-test*

- $\frac{\bar{X}_n^M - \bar{X}_m^F - \mu^M + \mu^F}{\sqrt{\frac{S_n^{2M}}{n} + \frac{S_m^{2F}}{m}}} \sim t$ de Student con $\frac{\left(\frac{S_n^{2M}}{n} + \frac{S_m^{2F}}{m}\right)^2}{\frac{\left(\frac{S_n^{2M}}{n}\right)^2}{n-1} + \frac{\left(\frac{S_m^{2F}}{m}\right)^2}{m+1}}$ grados de libertad.
- Se calcula $t = \frac{\bar{X}_n^M - \bar{X}_m^F - \mu^M + \mu^F}{\sqrt{\frac{S^{2M}}{n} + \frac{S^{2F}}{m}}}$.
- Se revisa p-valor $P(|T| > t)$.

Caso varianzas desconocidas: *Student t-test*

- $\frac{\bar{X}_n^M - \bar{X}_m^F - \mu^M + \mu^F}{\sqrt{\frac{(n-1)S_n^{2M} + (m-1)S_m^{2F}}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t$ de Student con $n + m - 2$ grados de libertad.
- Se calcula t.
- Se revisa p-valor $P(|T| > t)$.

Comparación de poblaciones

Caso varianzas desconocidas: ¿cómo elegir?

- Para varianzas similares ($\frac{1}{2} < \frac{\sqrt{s_n^{2M}}}{\sqrt{s_m^{2F}}} < 2$), Student t-test.
- Para varianzas disímiles ($\sqrt{s_n^{2M}} > 2\sqrt{s_m^{2F}}$ o $\sqrt{s_m^{2F}} > 2\sqrt{s_n^{2M}}$), Welch t-test.

¿Basta con p-valores?

The Guardian, viernes 4 de enero de 2002

- Al hacer girar una moneda 250 veces, una moneda de un euro belga terminó en cara 140 veces y sello 110. “Me parece muy sospechoso”, dijo Barry Blight, un profesor de estadística en la Escuela de Economía de Londres. “Si la moneda fuese insesgada la probabilidad de obtener un resultado así de extremo sería menos de un 7 %”.

¿Basta con p-valores?

Ejercicio

- Supongamos moneda sigue una distribución $X \sim \text{Bernoulli}(p)$.
- $H_0 : p = 0{,}5$.
- $H_1 : p \neq 0{,}5$.
- Muestra de tamaño $n = 250$, y se obtienen 140 caras.

p-valor

- Según teorema del límite central, $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
- En este caso $\mu = p$, $\sigma^2 = p(1-p)$.
- Es decir, $\bar{X}_n \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$.
- Se calcula
  $z = \frac{\sqrt{n}(\bar{x}_n - \mu)}{\sqrt{\sigma^2}} = \frac{\sqrt{n}(\bar{x}_n - p)}{\sqrt{p(1-p)}} = \frac{\sqrt{250}\left(\frac{140}{250} - 0{,}5\right)}{0{,}5} \approx 1{,}8974$.
- p-valor sería $P(|Z| > z \mid p = 0{,}5) \approx 0{,}0578$.

¿Basta con p-valores?

Ejercicio

- Supongamos moneda sigue una distribución $X \sim \text{Bernoulli}(p)$.
- $H_0 : p = 0{,}5$.
- $H_1 : p \neq 0{,}5$.
- Muestra de tamaño $n = 250$, y se obtienen 140 caras.

Recordar: teorema de Bayes

$$\underbrace{P(p \mid \text{datos})}_{\text{posterior}} = \frac{\overbrace{P(\text{datos} \mid p)}^{\text{verosimilitud}} \overbrace{P(p)}^{\text{prior}}}{\underbrace{P(\text{datos})}_{\text{evidencia}}}.$$

¿Basta con p-valores?

Ejercicio

- Supongamos moneda sigue una distribución $X \sim \text{Bernoulli}(p)$.
- $H_0 : p = 0{,}5$.
- $H_1 : p \neq 0{,}5$.
- Muestra de tamaño $n = 250$, y se obtienen 140 caras.

Enfoque bayesiano

- Se contrastan las hipótesis, usando la razón de posteriores:
$$\frac{P(H_0 \mid 140)}{P(H_1 \mid 140)} = \frac{P(140 \mid H_0)P(H_0)}{P(140 \mid H_1)P(H_1)}.$$
- En el caso de H_1, se debe elegir un prior para p.
- Si suponemos $P(H_0) = P(H_1)$, tenemos razón de verosimilitudes:
$$\frac{P(H_0 \mid 140)}{P(H_1 \mid 140)} = \frac{P(140 \mid H_0)}{P(140 \mid H_1)} = \text{Factor de Bayes.}$$

¿Basta con p-valores?

Ejercicio

- Supongamos moneda sigue una distribución $X \sim \text{Bernoulli}(p)$.
- $H_0 : p = 0{,}5$.
- $H_1 : p \neq 0{,}5$.
- Muestra de tamaño $n = 250$, y se obtienen 140 caras.

Enfoque bayesiano

- Prior para H_1: $p \sim U(0, 1)$.
- Si suponemos $P(H_0) = P(H_1)$, tenemos razón de verosimilitudes:

$$\frac{P(140 \mid H_1)}{P(140 \mid H_0)} = \frac{\int_0^1 \binom{250}{140} p^{140}(1-p)^{110} dp}{\binom{250}{140} 0{,}5^{250}}$$

$$= \frac{\int_0^1 p^{140}(1-p)^{110} dp}{2^{-250}} = 2^{250} B(141, 111) = 2^{250} \frac{140!110!}{251!} \approx 0{,}48.$$

¿Basta con p-valores?

Ejercicio

- Supongamos moneda sigue una distribución $X \sim \text{Bernoulli}(p)$.
- $H_0 : p = 0{,}5$.
- $H_1 : p \neq 0{,}5$.
- Muestra de tamaño $n = 250$, y se obtienen 140 caras.

Enfoque bayesiano – otro prior

- Prior para H_1: $p \sim \text{Beta}(\alpha, \alpha)$ (prior anterior es $\alpha = 1$).
- Si suponemos $P(H_0) = P(H_1)$, tenemos razón de verosimilitudes:

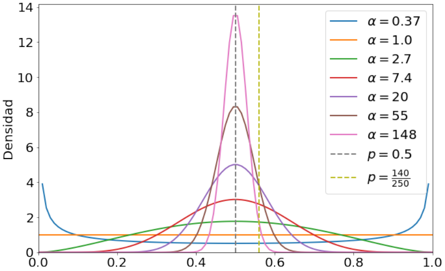
$$\frac{P(140 \mid H_1)}{P(140 \mid H_0)} = \frac{\int_0^1 \frac{1}{B(\alpha,\alpha)} \binom{250}{140} p^{140+\alpha-1}(1-p)^{110+\alpha-1} dp}{\binom{250}{140} 0{,}5^{250}}$$

$$= \frac{2^{250} B(140+\alpha, 110+\alpha)}{B(\alpha,\alpha)} = \frac{2^{250}\Gamma(140+\alpha)\Gamma(110+\alpha)\Gamma(2\alpha)}{\Gamma(250+2\alpha)\Gamma(\alpha)^2}.$$

Enfoque bayesiano

Hipótesis

- $H_0 : p = 0{,}5$.
- $H_1 : p \sim \mathrm{Beta}(\alpha, \alpha)$.

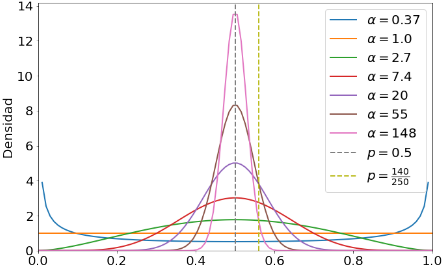
α	$\frac{P(140 \mid H_1)}{P(140 \mid H_0)}$	Prior (distribución beta $\mathrm{Beta}(\alpha, \alpha)$)
0,37	0,25	
1,00	0,48	
2,70	0,82	
7,40	1,3	
20,00	1,8	
55,00	1,9	
148,00	1,7	
403,00	1,3	
1096,00	1,1	

Enfoque bayesiano

Hipótesis

- $H_0 : p = 0{,}5,\ H_1 : p \sim \text{Beta}(\alpha, \alpha),\ H_2 : p = \frac{140}{250}.$

α	$\frac{P(140\mid H_1)}{P(140\mid H_0)}$	Prior (distribución beta $\text{Beta}(\alpha, \alpha)$)
0,37	0,25	
1,00	0,48	
2,70	0,82	
7,40	1,3	
20,00	1,8	
55,00	1,9	
148,00	1,7	
403,00	1,3	
1096,00	1,1	



- $$\frac{P(140\mid H_2)}{P(140\mid H_0)} = \frac{\left(\frac{140}{250}\right)^{140}\left(1-\frac{140}{250}\right)^{110}}{\frac{1}{2^{250}}} = \frac{2^{250}140^{140}110^{110}}{250^{250}} \approx 6{,}1.$$

Pequeño cambio

Ejercicio

- Supongamos moneda sigue una distribución $X \sim \text{Bernoulli}(p)$.
- $H_0 : p = 0{,}5$.
- $H_1 : p \neq 0{,}5$.
- Muestra de tamaño $n = 250$, y se obtienen 141 caras.

p-valor

- Según teorema del límite central, $\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$.
- En este caso $\mu = p$, $\sigma^2 = p(1-p)$.
- Es decir, $\bar{X}_n \sim \mathcal{N}(p, p(1-p))$.
- Se calcula
  $z = \frac{\sqrt{n}(\bar{x}_n - \mu)}{\sqrt{\sigma^2}} = \frac{\sqrt{n}(\bar{x}_n - p)}{\sqrt{p(1-p)}} = \frac{\sqrt{250}\left(\frac{141}{250} - 0{,}5\right)}{0{,}5} \approx 2{,}0239.$
- p-valor sería $P(|Z| > z \mid p = 0{,}5) \approx 0{,}0430$.

Pequeño cambio

Ejercicio

- Supongamos moneda sigue una distribución $X \sim \text{Bernoulli}(p)$.
- $H_0 : p = 0{,}5$.
- $H_1 : p \neq 0{,}5$.
- Muestra de tamaño $n = 250$.

p-valor

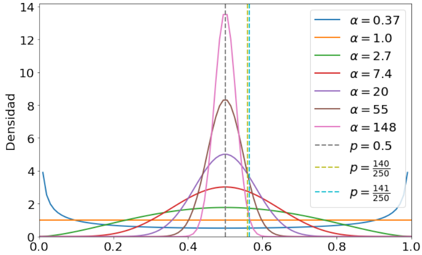
- Si hay 140 caras, p-valor es 0,0578.
- Si hay 141 caras, p-valor es 0,0430.
- Se fija nivel de significancia 0,05.
- En el primer caso, no se rechaza la hipótesis, en el segundo sí.

Enfoque bayesiano

Hipótesis

- $H_0 : p = 0{,}5,\ H_1 : p \sim \text{Beta}(\alpha, \alpha),\ H_2 : p = \frac{140}{250},\ H_3 : p = \frac{141}{250}.$

α	$\frac{P(140\|H_1)}{P(140\|H_0)}$	$\frac{P(141\|H_1)}{P(141\|H_0)}$	Prior (distribución beta $\text{Beta}(\alpha, \alpha)$)
0,37	0,25	0,32	
1,00	0,48	0,61	
2,70	0,82	1,0	
7,40	1,3	1,6	
20,00	1,8	2,2	
55,00	1,9	2,3	
148,00	1,7	1,9	
403,00	1,3	1,4	
1096,00	1,1	1,2	



- $$\frac{P(140|H_2)}{P(140|H_0)} = \frac{2^{250}140^{140}110^{110}}{250^{250}} \approx 6{,}1, \text{ y}$$
$$\frac{P(141|H_3)}{P(141|H_0)} = \frac{2^{250}141^{141}109^{109}}{250^{250}} \approx 7{,}8.$$

Enfoque bayesiano

Cuidado

- A veces se interpreta un p-valor de 0,05 como que la probabilidad en contra la hipótesis nula es 20 contra 1.
- En realidad, la evidencia en este caso está ligeramente a favor de la hipótesis nula, o contra de ella a lo más 2,3 contra 1, dependiendo del prior.

Cuidado

- Los p-valores y niveles de significancia hay que tratarlos con *extremo* cuidado.
- Muchos grandes enredos en la ciencia han sido producto de malas interpretaciones de p-valores.
- Es mucho más común de lo que parece.

¿Basta con p-valores?

Ejemplo

- Un científico lanza una moneda 12 veces y obtiene la secuencia *cccsccccsccs*.
- La secuencia contiene 3 s y 9 c.
- ¿Es la moneda sesgada?

p-valor

- Sea $n = 12$ y r la cantidad de s.
- r sigue una distribución binomial $B(12, p)$.
- Sea $H_0 : p = 0{,}5$.
- $P(r \leq 3 \mid p = 0{,}5) = \frac{1}{2^{12}}\left[\binom{12}{0} + \binom{12}{1} + \binom{12}{2} + \binom{12}{3}\right] \approx 0{,}073$.
- Bilateral: $P(r \leq 3 \text{ o } r \geq 9 \mid p = 0{,}5) \approx 0{,}146$.

¿Basta con p-valores?

Ejemplo

- El científico dice que la variable aleatoria no es r, pues antes de lanzar la moneda, ya había decidido que lo haría hasta que apareciesen 3 s. La variable aleatoria debe ser n.
- La secuencia contiene 3 s y 9 c.
- ¿Es la moneda sesgada?

p-valor – caso 2

- Sea $r = 3$ y n la cantidad de lanzamientos hasta observar 3 s.
- n sigue una distribución con masa $P(n \mid p = 0{,}5, r) = \binom{n-1}{r-1} \left(\frac{1}{2}\right)^n$.

$$P(n \geq 12 \mid p = 0{,}5) = \sum_{n=12}^{+\infty} \frac{1}{2^n} \binom{n-1}{2} = \sum_{n=12}^{+\infty} \frac{(n-1)(n-2)}{2^{n+1}} \approx 0{,}03.$$

¿Basta con p-valores?

Ejemplo

- La secuencia contiene 3 s y 9 c.
- ¿Es la moneda sesgada?

Enfoque bayesiano – caso 1

- Supongamos prior para p como $\text{Beta}(\alpha, \alpha)$.
- Verosimilitud binomial $r \sim B(12, p)$.

$$p(p \mid r = 3) = \frac{P(r = 3 \mid p)p(p)}{P(r = 3)}$$

$$= \frac{\binom{12}{3} p^3(1-p)^9 \Gamma(2\alpha) p^{\alpha-1}(1-p)^{\alpha-1}}{\Gamma(\alpha)^2 P(r = 3)}$$

$$\propto \frac{\Gamma(2\alpha) p^{2+\alpha}(1-p)^{8+\alpha}}{\Gamma(\alpha)^2}.$$

¿Basta con p-valores?

Ejemplo

- La secuencia contiene 3 s y 9 c.
- ¿Es la moneda sesgada?

Enfoque bayesiano – caso 2

- Supongamos prior para p como Beta(α, α).
- Verosimilitud para $n \sim \binom{n-1}{r-1} p^r(1-p)^{n-r} = \frac{(n-1)(n-2)p^3(1-p)^9}{2}$.

$$p(p \mid n = 12) = \frac{P(n = 12 \mid p)p(p)}{P(n = 12)}$$

$$= \frac{(12-1)(12-2)p^3(1-p)^9\Gamma(2\alpha)p^{\alpha-1}(1-p)^{\alpha-1}}{2\Gamma(\alpha)^2 P(n = 12)}$$

$$\propto \frac{\Gamma(2\alpha)p^{2+\alpha}(1-p)^{8+\alpha}}{\Gamma(\alpha)^2}.$$

¿Basta con p-valores?

Enfoque bayesiano

- ¡No depende de la decisión de cómo se tomó la muestra!

Para convencerse

- Un espía revisa resultados obtenidos para cada lanzamiento en la medida que ocurren.
- ¿La inferencia del espía en cada lanzamiento debería depender del criterio de término del experimento?

Para convencerse 2

- Luego del lanzamiento 12, alguien entra al laboratorio y roba la moneda.
- Ya no se pueden realizar más experimentos, por lo que aunque se haya querido, no se pueden hacer más lanzamientos.
- ¿Cómo se calcula el p-valor en ese caso?