

Introducción a la estadística

Bases indispensables y uso de

Olivier Devineau

`olivier.devineau@fcdarwin.org.ec`

Fundación Charles Darwin

Taller interno, 27–30 abril 2010

Noción de test estadístico

Distribución de probabilidad

Distribuciones

Generalidades

Binomial

Poisson

Normal

Otras

Procedimiento

¿Cuál test?

- Representación de las probabilidades asociadas con los estados posibles de una variable aleatoria

Ejemplo: X = número de hijos en una familia de 2 niños

- $2\varphi, (1\sigma, 1\varphi), (1\varphi, 1\sigma), 2\sigma$
 - $p(X = 0 \sigma) = 1/4$
 - $p(X = 1 \sigma) = 1/4 + 1/4$
 - $p(X = 2 \sigma) = 1/4$
- $\left. \begin{array}{l} \\ \\ \end{array} \right\} \sum p(X) = 1$

Distribución binomial

Definición

- Serie de n intentos independientes
- Cada intento \rightarrow Éxito / Fracaso
- Probabilidad de éxito: p

Distribución binomial

Definición

- Serie de n intentos independientes
- Cada intento \rightarrow **Éxito** / **Fracaso**
- Probabilidad de éxito: p

Distribuciones

Generalidades

Binomial

Poisson

Normal

Otras

Procedimiento

¿Cuál test?

Distribución binomial

Definición

- Serie de n intentos independientes
- Cada intento \rightarrow Éxito / Fracaso
- Probabilidad de éxito: p

Distribución binomial

Definición

- Serie de n intentos independientes
 - Cada intento \rightarrow Éxito / Fracaso
 - Probabilidad de éxito: p
- Distribución discontinua
 - $X \rightsquigarrow \mathcal{B}(n, p)$
 - $P(r) = \binom{n}{r} p^r (1 - p)^{n-r}$

Distribuciones

Generalidades

Binomial

Poisson

Normal

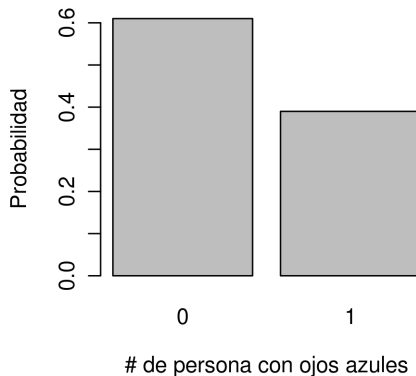
Otras

Procedimiento

¿Cuál test?

Distribución Binomial (2)

- 39% de los habitantes tienen ojos azules
- $X \sim \mathcal{B}(1, 0.39)$



Distribuciones

Generalidades

Binomial

Poisson

Normal

Otras

Procedimiento

¿Cuál test?

Distribución Binomial (2)

- 39% de los habitantes tienen ojos azules
- $X \sim \mathcal{B}(3, 0.39)$



Distribuciones

Generalidades

Binomial

Poisson

Normal

Otras

Procedimiento

¿Cuál test?

Distribución Binomial (2)

- 39% de los habitantes tienen ojos azules
- $X \sim \mathcal{B}(5, 0.39)$



Distribuciones

Generalidades

Binomial

Poisson

Normal

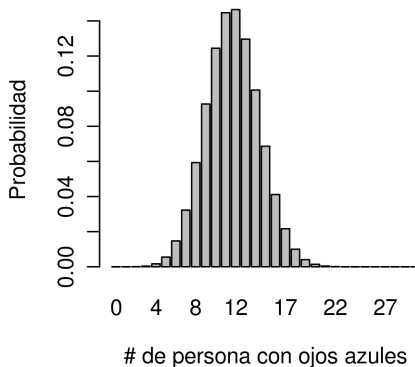
Otras

Procedimiento

¿Cuál test?

Distribución Binomial (2)

- 39% de los habitantes tienen ojos azules
- $X \sim \mathcal{B}(50, 0.39)$



Distribución binomial

¿Cuándo se aplica?

- Porcentaje de mortalidad
- Tasa de infección
- Proporción: sexos, respuesta a un tratamiento, intenciones de voto . . .

Se necesita saber cuantos individuos hay en categoría *éxito* y cuantos hay en categoría *fracaso*

Distribuciones

Generalidades

Binomial

Poisson

Normal

Otras

Procedimiento

¿Cuál test?

Distribución de Poisson

Definición

- Cuantas veces un evento **raro** ocurre por unidad de tiempo/espacio

Distribución de Poisson

Definición

- Cuantas veces un evento **raro** ocurre por unidad de tiempo/espacio

- Distribución discontinua

- $X \rightsquigarrow \mathcal{P}(\lambda)$

- $P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$

Distribución de Poisson

Definición

- Cuantas veces un evento **raro** ocurre por unidad de tiempo/espacio
- Distribución discontinua
- $X \rightsquigarrow \mathcal{P}(\lambda)$
- $P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$

Distribución de Poisson

Definición

- Cuantas veces un evento **raro** ocurre por unidad de tiempo/espacio
- Distribución discontinua
- $X \rightsquigarrow \mathcal{P}(\lambda)$
- $P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$

Distribución normal

Definición

- Teorema del límite central
- Suficientes muestras \rightarrow medias \rightarrow distribución normal

Distribuciones

Generalidades

Binomial

Poisson

Normal

Otras

Procedimiento

¿Cuál test?

Distribución normal

Definición

- Teorema del límite central
- Suficientes muestras \rightarrow medias \rightarrow distribución normal

Distribución normal

Definición

- Teorema del límite central
- Suficientes muestras \rightarrow medias \rightarrow distribución normal
- Distribución continua
 - $X \rightsquigarrow \mathcal{N}(\mu, \sigma)$
 - $f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$

Distribución normal

Definición

- Teorema del límite central
- Suficientes muestras \rightarrow medias \rightarrow distribución normal

- Distribución continua

- $X \rightsquigarrow \mathcal{N}(\mu, \sigma)$

- $$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

Distribución normal

Definición

- Teorema del límite central
- Suficientes muestras \rightarrow medias \rightarrow distribución normal
- Distribución continua
- $X \rightsquigarrow \mathcal{N}(\mu, \sigma)$
- $f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$

Distribuciones

Generalidades

Binomial

Poisson

Normal

Otras

Procedimiento

¿Cuál test?

Distribución normal

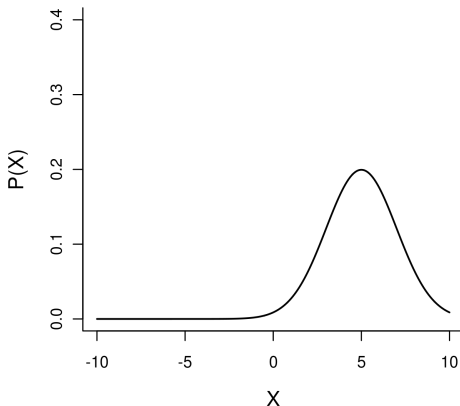
¿Cuándo se aplica?

- ¡Todo el tiempo!
- Regresión lineal, análisis de varianza ...

Distribución normal

¿Cuándo se aplica?

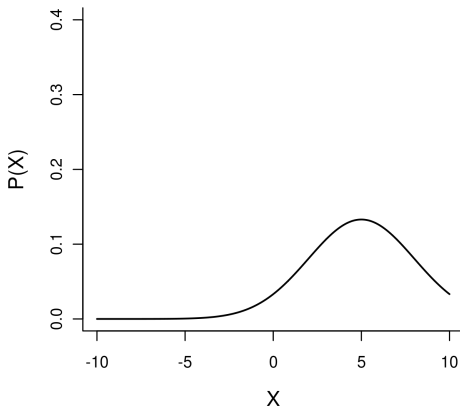
- ¡Todo el tiempo!
- Regresión lineal, análisis de varianza ...



Distribución normal

¿Cuándo se aplica?

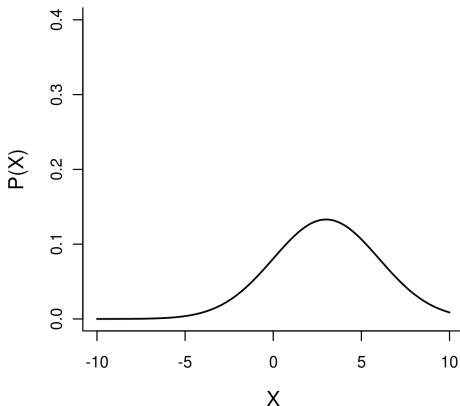
- ¡Todo el tiempo!
- Regresión lineal, análisis de varianza ...



Distribución normal

¿Cuándo se aplica?

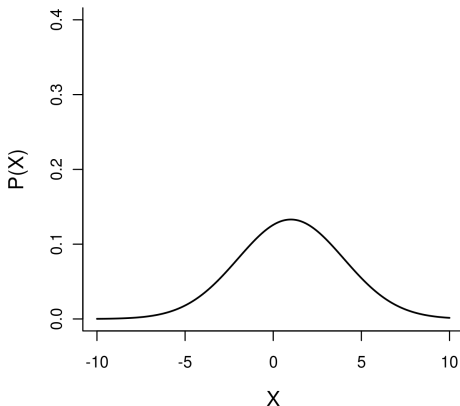
- ¡Todo el tiempo!
- Regresión lineal, análisis de varianza ...



Distribución normal

¿Cuándo se aplica?

- ¡Todo el tiempo!
- Regresión lineal, análisis de varianza ...



Distribuciones

Generalidades

Binomial

Poisson

Normal

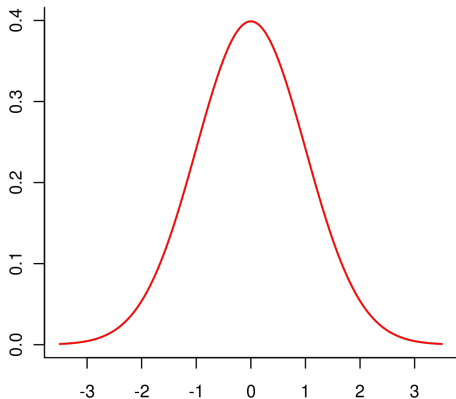
Otras

Procedimiento

¿Cuál test?

Distribución Normal Estándar

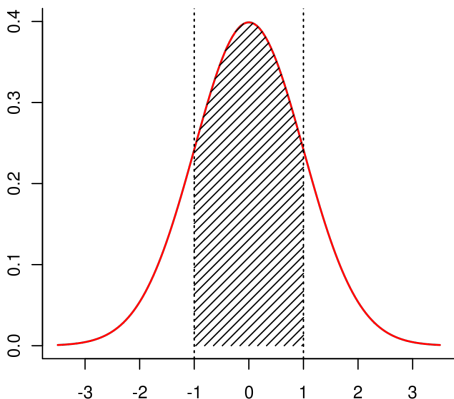
$$X \rightsquigarrow \mathcal{N}(0, 1)$$



- $\pm 1 \sigma \sim 68\%$
- $\pm 2 \sigma \sim 95\%$
- $\pm 3 \sigma \sim 99\%$

Distribución Normal Estándar

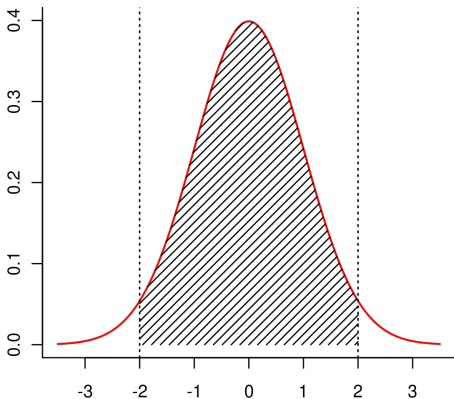
$$X \rightsquigarrow \mathcal{N}(0, 1)$$



- $\pm 1 \sigma \sim 68\%$
- $\pm 2 \sigma \sim 95\%$
- $\pm 3 \sigma \sim 99\%$

Distribución Normal Estándar

$$X \rightsquigarrow \mathcal{N}(0, 1)$$



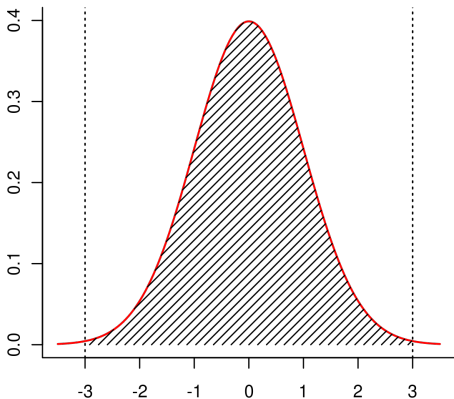
• $\pm 1 \sigma \sim 68\%$

• $\pm 2 \sigma \sim 95\%$

• $\pm 3 \sigma \sim 99\%$

Distribución Normal Estándar

$$X \rightsquigarrow \mathcal{N}(0, 1)$$



- $\pm 1 \sigma \sim 68\%$
- $\pm 2 \sigma \sim 95\%$
- $\pm 3 \sigma \sim 99\%$

Otras distribuciones de variables

- Lognormal (largo, peso . . .)
- Exponencial (Tiempo de fracaso)
- Gamma
- Distribución de Weibull
- Beta

Distribuciones de estadísticos

Distribuciones

Generalidades

Binomial

Poisson

Normal

Otras

Procedimiento

¿Cuál test?

- Distribución z
- Distribución t de Student
- Distribución del χ^2
- Distribución F de Fischer

¿Qué es un test estadístico?

Herramienta para tomar decisión

- Calcular un estadístico T_{obs} de una muestra
- Comparar T_{obs} con la distribución de T_{teo} cuando la hipótesis es verdadera
- La posición de T_{obs} informa sobre la probabilidad de que la hipótesis sea verdadera

Test estadístico: procedimiento

- 1 **Pregunta biológica:** ¿Hay cóndores en el parque?
- 2 Pregunta estadística: Hipótesis H_0
- 3 Elección del test estadístico: ¿Cuál usar?
- 4 Criterios de decisión: ¿Qué riesgo de error? ¿Qué nivel de confianza?

Test estadístico: procedimiento

Distribuciones

Procedimiento

¿Qué es un test?

Generalidades

Hipótesis

Procedimiento

Decisión

Poder

Collección de

datos

Cálculo

Valor P

Significancia

¿Cuál test?

- 1 Pregunta biológica: ¿Hay cóndores en el parque?
- 2 Pregunta estadística: Hipótesis H_0
- 3 Elección del test estadístico: ¿Cuál usar?
- 4 Criterios de decisión: ¿Qué riesgo de error? ¿Qué nivel de confianza?

Test estadístico: procedimiento

Distribuciones

Procedimiento

¿Qué es un test?

Generalidades

Hipótesis

Procedimiento

Decisión

Poder

Collección de
datos

Cálculo

Valor P

Significancia

¿Cuál test?

¿Cuál test?

- 1 Pregunta biológica: ¿Hay cóndores en el parque?
- 2 Pregunta estadística: Hipótesis H_0
- 3 Elección del test estadístico: ¿Cuál usar?
- 4 Criterios de decisión: ¿Qué riesgo de error? ¿Qué nivel de confianza?

Test estadístico: procedimiento

Distribuciones

Procedimiento

¿Qué es un test?

Generalidades

Hipótesis

Procedimiento

Decisión

Poder

Collección de
datos

Cálculo

Valor P

Significancia

¿Cuál test?

¿Cuál test?

- 1 Pregunta biológica: ¿Hay cóndores en el parque?
- 2 Pregunta estadística: Hipótesis H_0
- 3 Elección del test estadístico: ¿Cuál usar?
- 4 Criterios de decisión: ¿Qué riesgo de error? ¿Qué nivel de confianza?

Test estadístico: procedimiento

Distribuciones

Procedimiento

¿Qué es un test?

Generalidades

Hipótesis

Procedimiento

Decisión

Poder

Collección de
datos

Cálculo

Valor P

Significancia

¿Cuál test?

- 5 ¡Colección de los datos!
- 6 Cálculo de el estadístico del test
- 7 Decisión estadística: ¿Se puede rechazar H_0 o no?
- 8 Inferencia y explicación biológica

Test estadístico: procedimiento

Distribuciones

Procedimiento

¿Qué es un test?

Generalidades

Hipótesis

Procedimiento

Decisión

Poder

Colectión de
datos

Cálculo

Valor P

Significancia

¿Cuál test?

- 5 ¡Colección de los datos!
- 6 Cálculo de el estadístico del test
- 7 Decisión estadística: ¿Se puede rechazar H_0 o no?
- 8 Inferencia y explicación biológica

Test estadístico: procedimiento

Distribuciones

Procedimiento

¿Qué es un test?

Generalidades

Hipótesis

Procedimiento

Decisión

Poder

Collección de
datos

Cálculo

Valor P

Significancia

¿Cuál test?

- 5 ¡Colección de los datos!
- 6 Cálculo de el estadístico del test
- 7 Decisión estadística: ¿Se puede rechazar H_0 o no?
- 8 Inferencia y explicación biológica

Test estadístico: procedimiento

Distribuciones

Procedimiento

¿Qué es un test?

Generalidades

Hipótesis

Procedimiento

Decisión

Poder

Colectión de

datos

Cálculo

Valor P

Significancia

¿Cuál test?

- 5 ¡Colección de los datos!
- 6 Cálculo de el estadístico del test
- 7 Decisión estadística: ¿Se puede rechazar H_0 o no?
- 8 Inferencia y explicación biológica

Buenas y malas hipótesis

- Una buena hipótesis se puede rechazar/falsear

Buenas y malas hipótesis

- Una buena hipótesis se puede rechazar/falsear
- 1 Hay cóndores en el parque

Buenas y malas hipótesis

- Una buena hipótesis se puede rechazar/falsear
- 1 Hay cóndores en el parque
 - 2 No hay cóndores en el parque

Buenas y malas hipótesis

- Una buena hipótesis se puede rechazar/falsear
- ① Hay cóndores en el parque
- ② No hay cóndores en el parque
- ¡Ausencia de prueba no es prueba de ausencia!

Hipótesis nula

- “Nada está pasando”
- “Las medias de dos muestras son las mismas”
- “La pendiente de la relación es cero”

⇒ La hipótesis nula **se puede falsear**. Rechazar cuando los datos muestran que es suficientemente improbable

Hipótesis nula

- “Nada está pasando”
- “Las medias de dos muestras son las mismas”
- “La pendiente de la relación es cero”

⇒ La hipótesis nula **se puede falsear**. Rechazar cuando los datos muestran que es suficientemente improbable

Hipótesis nula

- “Nada está pasando”
- “Las medias de dos muestras son las mismas”
- “La pendiente de la relación es cero”

⇒ La hipótesis nula **se puede falsear**. Rechazar cuando los datos muestran que es suficientemente improbable

Hipótesis nula

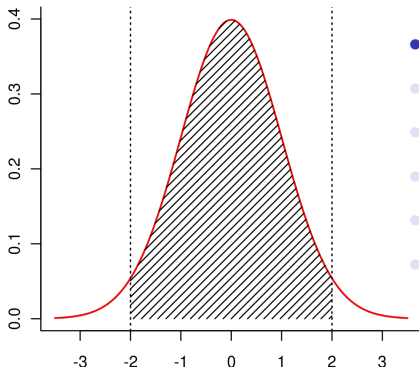
- “Nada está pasando”
- “Las medias de dos muestras son las mismas”
- “La pendiente de la relación es cero”

⇒ La hipótesis nula **se puede falsear**. Rechazar cuando los datos muestran que es suficientemente improbable

Elección del test

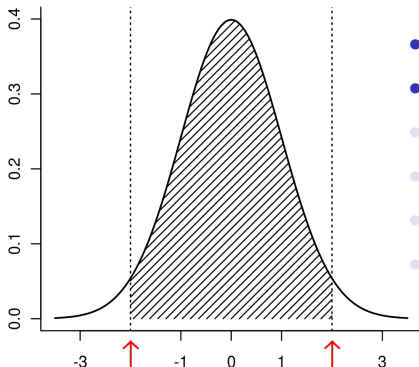
- Tipo de variables: cualitativas, cuantitativas . . .
- Número y tamaño de las muestras
- Condiciones de cada test

Criterios de decisión (1)



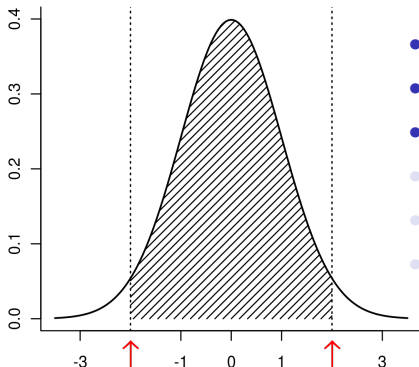
- $\pm 2 \sigma \sim 95\%$
- Valores umbrales
- Región de aceptación
- 5% menos probable
- Región de rechazo
- Riesgo α

Criterios de decisión (1)



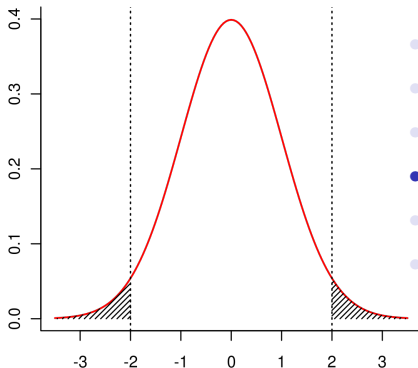
- $\pm 2 \sigma \sim 95\%$
- Valores umbrales
- Región de aceptación
- 5% menos probable
- Región de rechazo
- Riesgo α

Criterios de decisión (1)



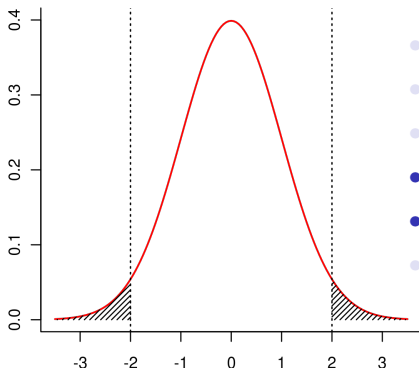
- $\pm 2 \sigma \sim 95\%$
- Valores umbrales
- Región de aceptación
- 5% menos probable
- Región de rechazo
- Riesgo α

Criterios de decisión (1)



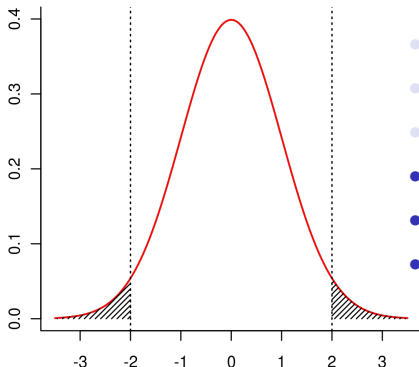
- $\pm 2 \sigma \sim 95\%$
- Valores umbrales
- Región de aceptación
- 5% menos probable
- Región de rechazo
- Riesgo α

Criterios de decisión (1)



- $\pm 2 \sigma \sim 95\%$
- Valores umbrales
- Región de aceptación
- 5% menos probable
- Región de rechazo
- Riesgo α

Criterios de decisión (1)



- $\pm 2 \sigma \sim 95\%$
- Valores umbrales
- Región de aceptación
- 5% menos probable
- Región de rechazo
- Riesgo α

Criterios de decisión (2)

- 2 errores posibles :

Tipo I : Rechazar H_0 cuando es verdadera

Tipo II : Aceptar H_0 cuando es falsa

Hipótesis nula	Situación real	
	Verdadera	Falsa
Acepta	Decisión correcta Poder $1 - \beta$	Tipo II Riesgo β
Rechaza	Tipo I Riesgo α	Decisión correcta

Criterios de decisión (2)

- 2 errores posibles :

Tipo I : Rechazar H_0 cuando es verdadera

Tipo II : Aceptar H_0 cuando es falsa

Hipótesis nula	Situación real	
	Verdadera	Falsa
Acepta	Decisión correcta Poder $1 - \beta$	Tipo II Riesgo β
Rechaza	Tipo I Riesgo α	Decisión correcta

Criterios de decisión (2)

- 2 errores posibles :

Tipo I : Rechazar H_0 cuando es verdadera

Tipo II : Aceptar H_0 cuando es falsa

Hipótesis nula	Situación real	
	Verdadera	Falsa
Acepta	Decisión correcta Poder $1 - \beta$	Tipo II Riesgo β
Rechaza	Tipo I Riesgo α	Decisión correcta

Criterios de decisión (2)

- 2 errores posibles :

Tipo I : Rechazar H_0 cuando es verdadera

Tipo II : Aceptar H_0 cuando es falsa

Hipótesis nula	Situación real	
	Verdadera	Falsa
Acepta	Decisión correcta Poder $1 - \beta$	Tipo II Riesgo β
Rechaza	Tipo I Riesgo α	Decisión correcta

Criterios de decisión (2)

- 2 errores posibles :

Tipo I : Rechazar H_0 cuando es verdadera

Tipo II : Aceptar H_0 cuando es falsa

Hipótesis nula	Situación real	
	Verdadera	Falsa
Acepta	Decisión correcta Poder $1 - \beta$	Tipo II Riesgo β
Rechaza	Tipo I Riesgo α	Decisión correcta

Hay que comprometer . . .

Poder: Probabilidad de rechazar H_0 cuando es falsa

- Error I: rechazar H_0 cuando es verdadera α
- Error II: aceptar H_0 cuando es falsa β
- Poder: $1 - \beta$
- α y β relacionados
- Cuando $\alpha \searrow \beta \nearrow$

Hay que comprometer . . .

Poder: Probabilidad de rechazar H_0 cuando es falsa

- Error I: rechazar H_0 cuando es verdadera α
- Error II: aceptar H_0 cuando es falsa β
- Poder: $1 - \beta$
- α y β relacionados
- Cuando $\alpha \searrow \beta \nearrow$

Hay que comprometer . . .

Poder: Probabilidad de rechazar H_0 cuando es falsa

- Error I: rechazar H_0 cuando es verdadera α
- Error II: aceptar H_0 cuando es falsa β
- Poder: $1 - \beta$
- α y β relacionados
- Cuando $\alpha \searrow \beta \nearrow$

Hay que comprometer . . .

Poder: Probabilidad de rechazar H_0 cuando es falsa

- Error I: rechazar H_0 cuando es verdadera α
- Error II: aceptar H_0 cuando es falsa β
- Poder: $1 - \beta$
- α y β relacionados
- Cuando $\alpha \searrow \beta \nearrow$

Hay que comprometer . . .

Poder: Probabilidad de rechazar H_0 cuando es falsa

- Error I: rechazar H_0 cuando es verdadera α
- Error II: aceptar H_0 cuando es falsa β
- Poder: $1 - \beta$
- α y β relacionados
- Cuando $\alpha \searrow \beta \nearrow$

Hay que comprometer . . .

Poder: Probabilidad de rechazar H_0 cuando es falsa

- Error I: rechazar H_0 cuando es verdadera α
- Error II: aceptar H_0 cuando es falsa β
- Poder: $1 - \beta$
- α y β relacionados
- Cuando $\alpha \searrow \beta \nearrow$

Hay que comprometer . . .

Poder: Probabilidad de rechazar H_0 cuando es falsa

- Error I: rechazar H_0 cuando es verdadera α
- Error II: aceptar H_0 cuando es falsa β
- Poder: $1 - \beta$
- α y β relacionados
- Cuando $\alpha \searrow \beta \nearrow$

¿Cuando α debe ser alto?

Ejemplo: Efectos secundarios de una droga

- Test final antes de comercializar
- Grupo A: droga | Grupo B: placebo
- H_0 : no hay diferencia entre grupos A y B
- H_1 : A tiene mayor frecuencia de anomalías que B

¿Cuándo α debe ser alto?

Aceptar riesgo α más alto para reducir riesgo β

α alto: error de tipo I

- H_0 rechazada pero verdadera
- No se comercializa
- Más estudios para determinar efecto real

¿Cuándo α debe ser alto?

Aceptar riesgo α más alto para reducir riesgo β

α alto: error de tipo I

- H_0 rechazada pero verdadera
- No se comercializa
- Más estudios para determinar efecto real

β alto: error de tipo II

- H_0 “aceptada” pero falsa
- Comercialización
- ¡Mucha gente sufre de los efectos secundarios!

Colección de los datos

¡Acuérdense!

- Aleatorización
- Replicación

Computación del estadístico del test

Ejemplo: Prevalencia de la malaria

- “La prevalencia es la misma en A y en B”
- $H_0 : \mu_A = \mu_B$
- El estadístico del test representa la diferencia de prevalencia: $T = f(\text{prev}_A - \text{prev}_B)$
- Distribución de T corresponde a H_0 verdadera

Computación del estadístico del test

Ejemplo: Prevalencia de la malaria

- “La prevalencia es la misma en A y en B”
- $H_0 : \mu_A = \mu_B$
- El estadístico del test representa la diferencia de prevalencia: $T = f(\text{prev}_A - \text{prev}_B)$
- Distribución de T corresponde a H_0 verdadera

Computación del estadístico del test

Ejemplo: Prevalencia de la malaria

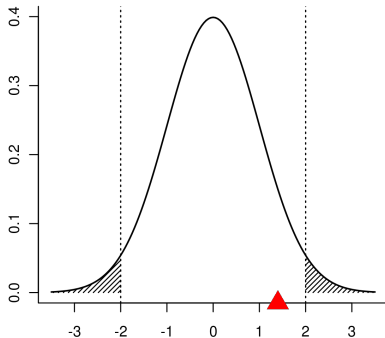
- “La prevalencia es la misma en A y en B”
- $H_0 : \mu_A = \mu_B$
- El estadístico del test representa la diferencia de prevalencia: $T = f(\text{prev}_A - \text{prev}_B)$
- Distribución de T corresponde a H_0 verdadera

Computación del estadístico del test

Ejemplo: Prevalencia de la malaria

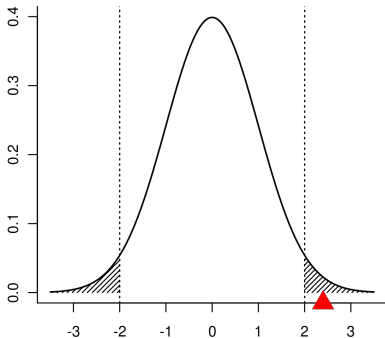
- “La prevalencia es la misma en A y en B”
- $H_0 : \mu_A = \mu_B$
- El estadístico del test representa la diferencia de prevalencia: $T = f(\text{prev}_A - \text{prev}_B)$
- Distribución de T corresponde a H_0 verdadera

Comparación de T con la distribución teórica



- T_{obs} no está en la región de rechazo
- No se puede rechazar H_0
- No es posible afirmar que hay una diferencia de prevalencia entre A y B

Comparación de T con la distribución teórica



- T_{obs} está en la región de rechazo
- Se puede rechazar H_0
- Se concluye que la prevalencia de la malaria es diferente entre A y B
- El riesgo de que esta conclusión sea falsa es $\alpha = 5\%$

Distribuciones

Procedimiento

¿Qué es un test?

Generalidades

Hipótesis

Procedimiento

Decisión

Poder

Collección de

datos

Cálculo

Valor P

Significancia

¿Cuál test?

- Medida de la credibilidad de la hipótesis nula

Valor P

- Medida de la credibilidad de la hipótesis nula

Ejemplo

- $H_0 : \mu_A = \mu_B$
- $p < 0.05$
- $p = 0.23$

Valor P

- Medida de la credibilidad de la hipótesis nula

Ejemplo

- $H_0 : \mu_A = \mu_B$
- $p < 0.05$
- $p = 0.23$

Valor P

- Medida de la credibilidad de la hipótesis nula

Ejemplo

- $H_0 : \mu_A = \mu_B$
- $p < 0.05 \Rightarrow$ improbable que H_0 sea verdadera: $\mu_A \neq \mu_B$
- $p = 0.23$

Valor P

- Medida de la credibilidad de la hipótesis nula

Ejemplo

- $H_0 : \mu_A = \mu_B$
- $p < 0.05$
- $p = 0.23$

Valor P

- Medida de la credibilidad de la hipótesis nula

Ejemplo

- $H_0 : \mu_A = \mu_B$
- $p < 0.05$
- $p = 0.23 \Rightarrow$ No hay suficiente evidencia para rechazar H_0

Significancia

- ¿Qué significa “Resultado significativo”?
- Diccionario: Que tiene sentido
- Estadística: Improbable que haya ocurrido por azar si la hipótesis nula es verdadera
- Improbable: Ocurre menos de 5% de las veces

Significancia

- ¿Qué significa “Resultado significativo”?
- Diccionario: Que tiene sentido
- Estadística: Improbable que haya ocurrido por azar si la hipótesis nula es verdadera
- Improbable: Ocurre menos de 5% de las veces

Significancia

- ¿Qué significa “Resultado significativo”?
- Diccionario: Que tiene sentido
- Estadística: Improbable que haya ocurrido por azar si la hipótesis nula es verdadera
- Improbable: Ocurre menos de 5% de las veces

Significancia

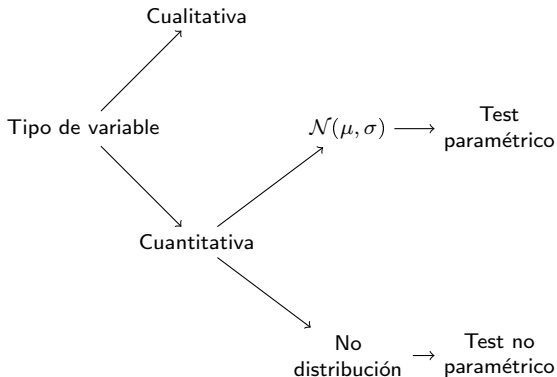
- ¿Qué significa “Resultado significativo”?
- Diccionario: Que tiene sentido
- Estadística: Improbable que haya ocurrido por azar si la hipótesis nula es verdadera
- Improbable: Ocurre menos de 5% de las veces

Significancia

- ¿Qué significa “Resultado significativo”?
- Diccionario: Que tiene sentido
- Estadística: Improbable que haya ocurrido por azar si la hipótesis nula es verdadera
- Improbable: Ocurre menos de 5% de las veces

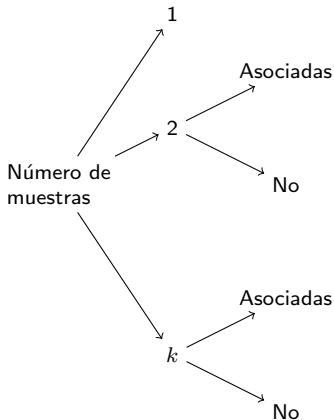
¿Como elegir el test adecuado?

Algunas directrices (1)



¿Como elegir el test adecuado?

Algunas directrices (2)



Dependencia – Asociación

Tests asociados

- Muestras asociadas: vienen del mismo grupo
- Relacionadas por correlación o por regresión
- Conexión espacial
- Conexión temporal

⇒ Usar tests específicos: e.g., “paired t-test”

Comparar una muestra con una distribución teórica

⇒ Test de conformidad

- Test t de conformidad
- Test de Wilcoxon
- Test binomial
- Test χ^2 de conformidad
- ...

Comparar dos muestras

⇒ Test de comparación (de homogeneidad)

- Test t (posiblemente “asociado”)
- Test de Mann-Whitney
- Test de Fisher
- Test χ^2
- ...

Comparar *más* de dos muestras

⇒ Test de comparación (continuación)

- Anova / Manova
- Test de Kruskal-Wallis
- Test de Friedman
- Test χ^2
- ...

Evaluar el grado de asociación entre variables

Muestras independientes

⇒ Correlación y regresión

- Correlación de Pearson / de Spearman ($n = 2$)
- Regresión simple / regresión logística ($n=2$)
- Regresión no paramétrica
- Regresión múltiple / regresión logística múltiple ($n > 2$)
- ...

Comparar un grupo con una distribución teórica

Medidas $X \sim \mathcal{N}(\mu, \sigma)$	Categoría, grado, sin distribución	Binomial
Test t 1 muestra	Test de Wilcoxon	Test χ^2 , test binomial

Comparar 2 grupos no asociados

Medidas $X \sim \mathcal{N}(\mu, \sigma)$	Categoría, grado, sin distribución	Binomial
Test t 1 muestra	Test de Wilcoxon	Test χ^2 , test binomial
Test t no asociado	Test de Mann-Whitney	Test de Fisher, test χ^2

Comparar 2 grupos asociados

Medidas $X \sim \mathcal{N}(\mu, \sigma)$	Categoría, grado, sin distribución	Binomial
Test t 1 muestra	Test de Wilcoxon	Test χ^2 , test binomial
Test t no asociado	Test de Mann-Whitney	Test de Fisher, test χ^2
Test t asociado	Test de Wilcoxon	Test de McNemar

Comparar ≥ 3 grupos no asociados

Medidas $X \sim \mathcal{N}(\mu, \sigma)$	Categoría, grado, sin distribución	Binomial
Test t 1 muestra	Test de Wilcoxon	Test χ^2 , test binomial
Test t no asociado	Test de Mann-Whitney	Test de Fisher, test χ^2
Test t asociado	Test de Wilcoxon	Test de McNemar
Anova simple	Test de Kruskal-Wallis	Test χ^2

Comparar ≥ 3 grupos asociados

Medidas $X \sim \mathcal{N}(\mu, \sigma)$	Categoría, grado, sin distribución	Binomial
Test t 1 muestra	Test de Wilcoxon	Test χ^2 , test binomial
Test t no asociado	Test de Mann-Whitney	Test de Fisher, test χ^2
Test t asociado	Test de Wilcoxon	Test de McNemar
Anova simple	Test de Kruskal-Wallis	Test χ^2
Anova con medidas repetidas	Test de Friedman	Test Q de Cochran

Cuantificar asociación entre 2 variables

Medidas $X \sim \mathcal{N}(\mu, \sigma)$	Categoría, grado, sin distribución	Binomial
Test t 1 muestra	Test de Wilcoxon	Test χ^2 , test binomial
Test t no asociado	Test de Mann-Whitney	Test de Fisher, test χ^2
Test t asociado	Test de Wilcoxon	Test de McNemar
Anova simple	Test de Kruskal-Wallis	Test χ^2
Anova con medidas repetidas	Test de Friedman	Test Q de Cochran
Correlación de Pearson	Correlación de Spearman	Coefficientes de contingencia

Predecir valor desde 1 variable

Medidas $X \sim \mathcal{N}(\mu, \sigma)$	Categoría, grado, sin distribución	Binomial
Test t 1 muestra	Test de Wilcoxon	Test χ^2 , test binomial
Test t no asociado	Test de Mann-Whitney	Test de Fisher, test χ^2
Test t asociado	Test de Wilcoxon	Test de McNemar
Anova simple	Test de Kruskal-Wallis	Test χ^2
Anova con medidas repetidas	Test de Friedman	Test Q de Cochran
Correlación de Pearson	Correlación de Spearman	Coefficientes de contingencia
Regresión (no)lineal simple	Regresión no paramétrica	Regresión logística simple

Predecir valor desde varias variables

Medidas $X \sim \mathcal{N}(\mu, \sigma)$	Categoría, grado, sin distribución	Binomial
Test t 1 muestra	Test de Wilcoxon	Test χ^2 , test binomial
Test t no asociado	Test de Mann-Whitney	Test de Fisher, test χ^2
Test t asociado	Test de Wilcoxon	Test de McNemar
Anova simple	Test de Kruskal-Wallis	Test χ^2
Anova con medidas repetidas	Test de Friedman	Test Q de Cochran
Correlación de Pearson	Correlación de Spearman	Coefficientes de contingencia
Regresión (no)lineal simple	Regresión no paramétrica	Regresión logística simple
Regresión (no)lineal multiple	_____	Regresión logística multiple

Más recursos para elegir un test

Distribuciones

Procedimiento

¿Cuál test?

Árbol de decisión

Comparación

Asociación

Más directrices

- *Handbook of Biological Statistics:*
<http://udel.edu/~mcdonald/statbigchart.html>
- *Statistics Online Computational Resources:*
www.socr.ucla.edu/Applets.dir/ChoiceOfTest.html
- *GraphPad / Intuitive Biostatistics:*
www.graphpad.com/www/Book/Choose.htm
- *Social Research Methods:*
www.socialresearchmethods.net/selstat/ssstart.htm
- *James D. Leeper, University of Alabama:*
<http://bama.ua.edu/~jleeper/627/choosestat.html>
- *S. Holttum, B. Blizzard, Canterbury Christ Church University:*
www.whichtest.info/index.html