

Intervalos de confianza y test de hipótesis para dos poblaciones normales

Estadística, Grado en Sistemas de Información

Constantino Antonio García Martínez

Universidad San Pablo Ceu

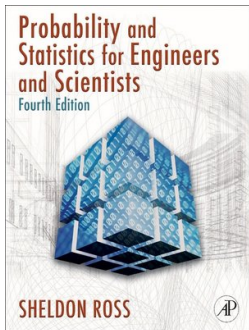
1. Comparaciones de medias en poblaciones normales

Varianzas totalmente desconocidas

Varianzas desconocidas pero iguales

Datos apareados

2. Comparaciones de varianzas



S. Ross. Introduction to Probability and Statistics for Engineers and Scientists. Chapter 6.



C.D. Barr, D.M. Diez, M. Çetinkaya-Rundel. OpenIntro Statistics. Chapters 4.

Comparaciones de medias en poblaciones normales

Comparaciones de medias en poblaciones normales

Varianzas totalmente desconocidas

Ejercicio: Diferencias por sexos

Los datos contenidos en "howell1.csv" son datos censales parciales del área !Kung San compilados a partir de entrevistas realizadas a finales de la década de 1960. ¿Depende la altura de los !Kung adultos del sexo del individuo? ($\alpha = 0.01$).

Apoya tus resultados con un gráfico y calcula el tamaño del efecto. Emplea los datos en "howell1.csv".

```
## [1] 3.098612e-51
## Cohen's d |          95% CI
## -----
## 1.95      | [1.69, 2.21]
##
## - Estimated using un-pooled SD.
```

Comparaciones de medias en poblaciones normales

Varianzas desconocidas pero iguales

Ejercicio: Varianzas desconocidas pero iguales

Repita el ejercicio relativo a los !Kung adultos si se puede asumir que la desviación típica poblacional para hombres y mujeres es la misma ($\sigma_h = \sigma_m$).

```
## [1] 4.068028e-53
## Cohen's d |      95% CI
## -----
## 1.96      | [1.70, 2.21]
##
## - Estimated using pooled SD.
```


Comparaciones de medias en poblaciones normales

Datos apareados

Ejercicio: Datos apareados

Unos científicos examinaron la función de la vesícula biliar antes y después de una cirugía para detener el reflujo. Los autores midieron la funcionalidad de la vesícula biliar calculando la fracción de eyección de la vesícula biliar (GBEF) antes y después de la operación, cuyo objetivo es aumentar la GBEF. ¿Hay evidencia para concluir que la operación aumenta el GBEF? Datos en "gbef_long.txt" (o "gbef.txt", para un reto).

```
## [1] 0.04086217  
## [1] "Cohens'd : 0.553061950304492"
```

Comparaciones de varianzas

Para comparar las varianzas de dos poblaciones normales

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \quad X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$$

intentaremos buscar un estadístico que nos permita estimar $\frac{\sigma_1^2}{\sigma_2^2}$.

Comparaciones de varianzas

Para comparar las varianzas de dos poblaciones normales

$$X_1 \sim \mathcal{N}(\mu_1, \sigma_1^2) \quad X_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$$

intentaremos buscar un estadístico que nos permita estimar $\frac{\sigma_1^2}{\sigma_2^2}$.

Teorema

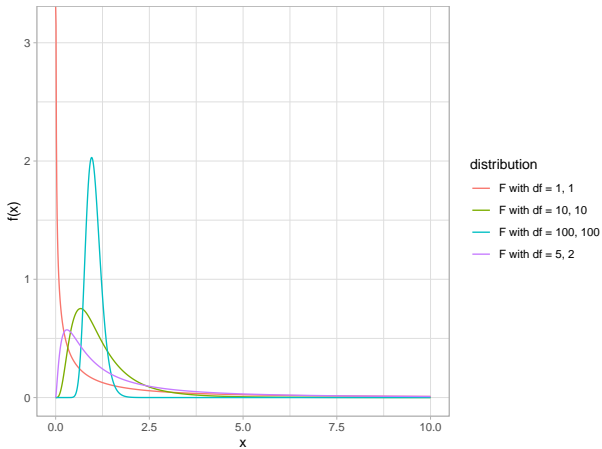
Si obtenemos dos muestras independientes de tamaño n_1 y n_2 de dos **poblaciones normales** con varianzas σ_1^2 y σ_2^2 . Entonces, el estadístico

$$\frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2}$$

tiene una **distribución F** con $n_1 - 1$, $n_2 - 1$ grados de libertad (también llamada, **F de Snedecor o de Fisher-Snedecor**).

$$\frac{\hat{S}_1^2/\sigma_1^2}{\hat{S}_2^2/\sigma_2^2} \sim \mathcal{F}_{n_1-1, n_2-1}.$$

Distribución F



Ejercicio:

En una empresa, se están comparando dos métodos de producción de cierto chip (A, mucho más barato, y B). La potencia media consumida por ambos chips es idéntica, si bien los dos métodos tienen distinta variabilidad. Se obtienen dos muestras de tamaño 16 y 10 y sus varianzas muestrales son 24 y 18 (en Watts²). Usando un nivel de confianza del 98 %, ¿qué método es preferible? Usa la función *var.test*.

```
## [1] 0.2687046 5.1930508
```

Ejercicio: Estadístico F

Resuelve el mismo ejercicio empleando haciendo los cálculos del IC y del p-valor “a mano”.

Ejercicio: !Kung y varianzas

Usa un test de ratio de varianzas para discutir si es razonable asumir igualdad de varianzas en el ejercicio de los !Kung (¿Existe evidencia de que las varianzas por sexo son distintas?)