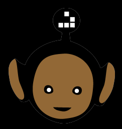
Notas tema 11 AED inferencias sobre la media de un vector

Rodrigo Castillo

7 de septiembre de 2020



1. caso univariado

para hacer una conclusion valida de la media de una muestra, si tienes p variables correlacionadas deben ser analizadas los test cubridos en este conjunto de notas son de la forma

$$H_o : \mu = \mu_o \tag{1}$$

donde $mu_{p\times 1}$ es el vector de las medias poblacionales y $\mu_{o,p\times 1}$ son valores nulos en la hipótesis
para el caso univariado...

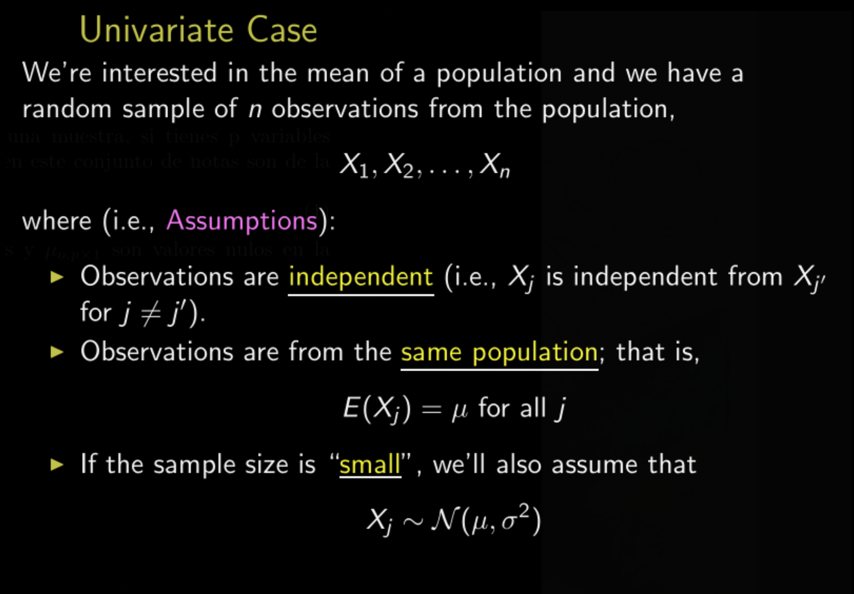


Figura 1: Caso Univariado

hipotesis y test ...

- Hypothesis:

$$H_o : \mu = \mu_o \qquad \text{versus} \qquad H_1 : \mu \neq \mu_o$$

  where μ_o is some specified value. In this case, H_1 is 2–sided alternative.
- Test Statistic:

$$t = \frac{\bar{X} - \mu_o}{s/\sqrt{n}}$$

  where $\bar{X} = (1/n)\sum_{j=1}^{n} X_j$ and $s = \sqrt{(1/(n-1))\sum_{j-1}^{n}(X_j - \bar{X})^2}$
- Sampling Distribution: If H_o and assumptions are true, then the sampling distribution of t is Student's - t distribution with $df = n - 1$.
- Decision: Reject H_o when t is “large” (i.e., small p–value).

Figura 2: Hipotesis y test

entonces la desicion es clasica para la distribucion $t - student$

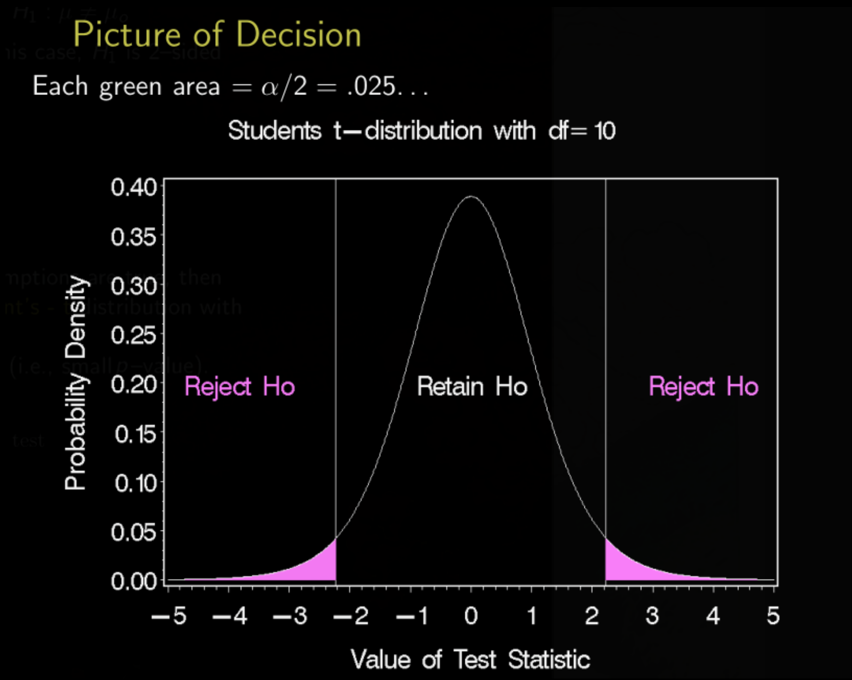


Figura 3: Desición

intervalos de confianza ...

$$\left|\frac{\bar{x}-\mu_o}{s/\sqrt{n}}\right| \leq t_{n-1,(\alpha/2)}$$

en donde $T_{n-1}(\frac{\alpha}{2})$ es la parte superior *blahblahblah*...

$$\left|\frac{\bar{x}-\mu_o}{s/\sqrt{n}}\right| \leq t_{n-1,(\alpha/2)}$$

where $t_{n-1,(\alpha/2)}$ is the upper $(\alpha/2)100\%$ percentile of Student's t-distribution with $df = n-1$. ...OR

$$\left\{\mu_o \text{ such that } \bar{x} - t_{n-1,(\alpha/2)}\frac{s}{\sqrt{n}} \leq \mu_o \leq \bar{x} + t_{n-1,(\alpha/2)}\frac{s}{\sqrt{n}}\right\}$$

A $100(1-\alpha)^{th}$ confidence interval or region for μ is

$$\left(\bar{x} - t_{n-1,(\alpha/2)}\frac{s}{\sqrt{n}}, \quad \bar{x} + t_{n-1,(\alpha/2)}\frac{s}{\sqrt{n}}\right)$$

Before for sample is selected, the ends of the interval depend on random variables $\bar{X}$'s and s; this is a random interval. $100(1-\alpha)^{th}$ percent of the time such intervals with contain the "true" mean μ

Figura 4: Test

2. caso multivariado

la extension del caso univariado, vamos a reemplazar los escalares para los vectores y matrices

$$t^2 = \frac{(\bar{x}-\mu_o)^2}{s^2/n} = n(\bar{x}-\mu_o)(s^2)^{-1}(\bar{x}-\mu_o)$$

donde t^2 es un metodo cuadrado estadistico de la distancia entre la muestra $\hat{x}$ y el valor hipotetico μ_o

se tiene que $t^2_{df} = F_{1,df}$

eso quiere decir que la muestra de distribucion de

$$t^2 = n(\bar{x}-\mu_o)(s^2)^{-1}(\bar{x}-\mu_o) \sim \mathcal{F}_{1,n-1}.$$

y se extiende así ...

hotelingm T^2

ver el ejemplo 2 y 3

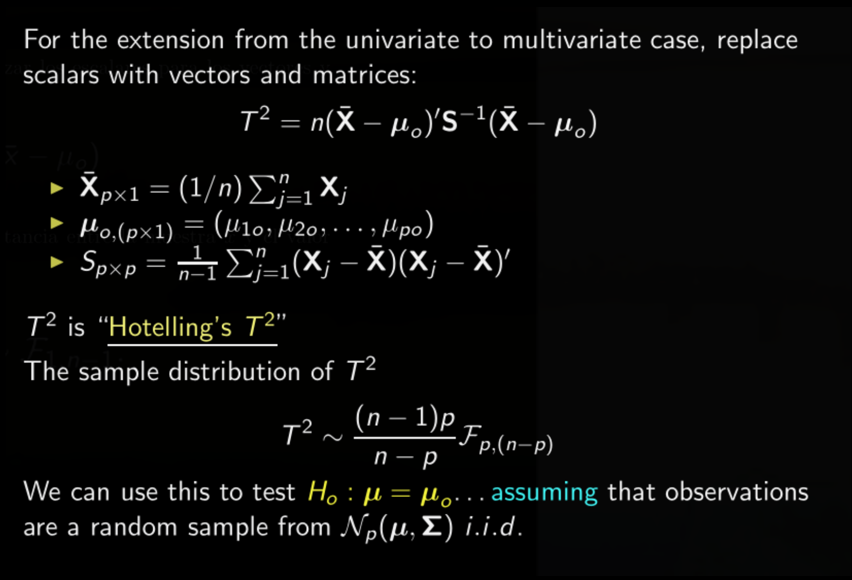


Figura 5: extension al caso multivariado

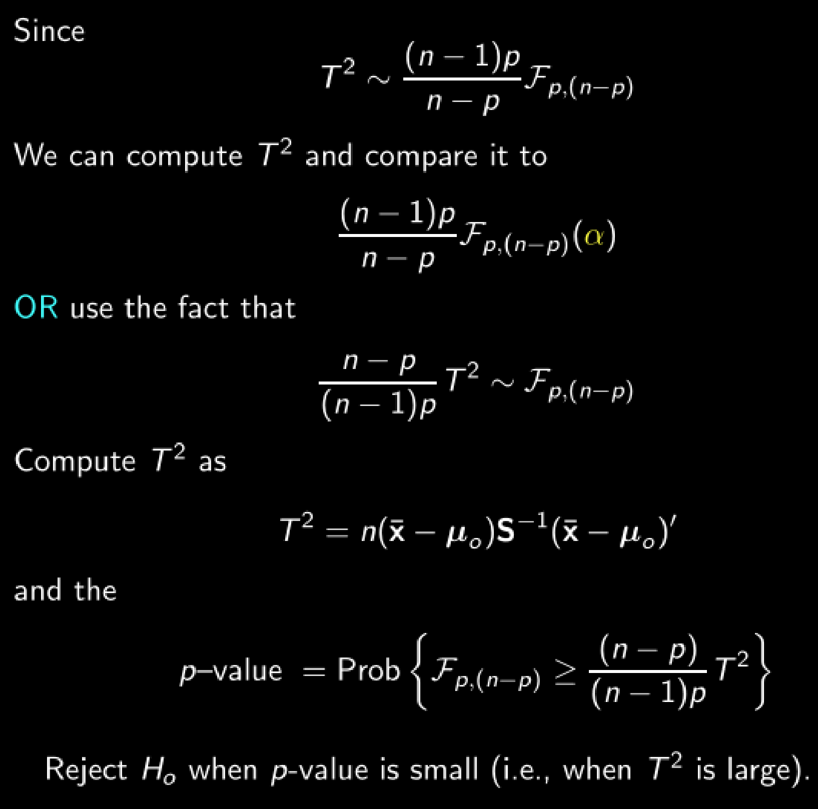


Figura 6: Hoteling

3. cosas cool

4. Likelihood ratio

T^2 esta basado en el principio de la union, que agarra un caso multivariado y lo vuelve un problema uniariado, el problema tipo $T^2 \frac{=}{5} a'(\hat{X} - \mu - 0)$ es una combinacion lineal

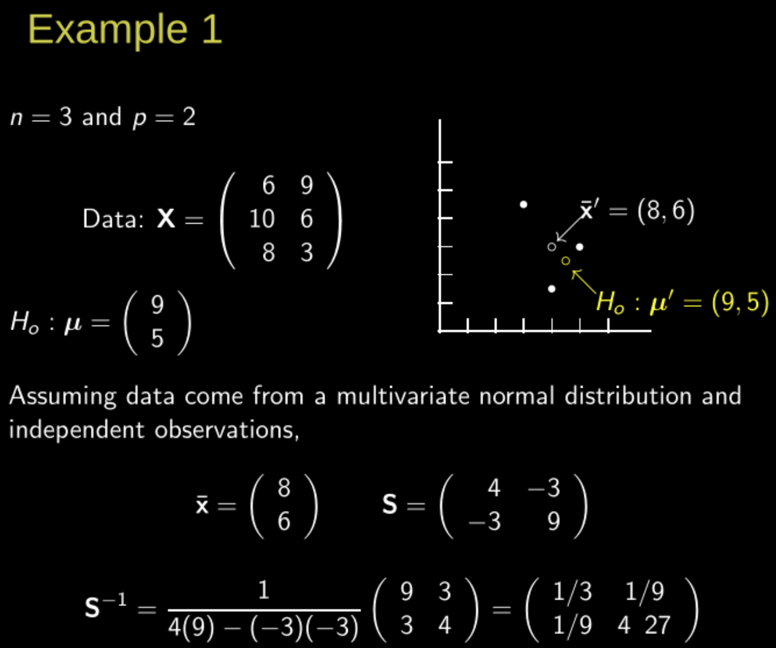


Figura 7: ejemplo1

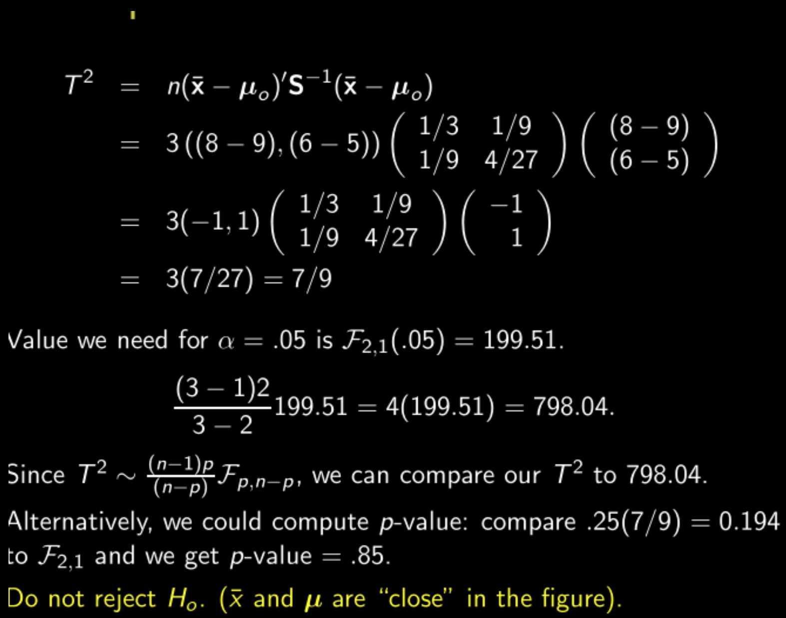


Figura 8: ejemplo2 -continuacion

selecionamos el la combinacion del vector que nos lleve al maximo posible valor de T^2 , el plan es ...

$$t^2 = \begin{pmatrix} \text{normal} \\ \text{random} \\ \text{variable} \end{pmatrix} \left(\frac{\text{chi-square random varible}}{\text{degress of freedom}} \right)^{-1} \begin{pmatrix} \text{normal} \\ \text{random} \\ \text{variable} \end{pmatrix}$$

Figura 9: Cosa cool

- Another approach to testing null hypothesis about mean vector $\boldsymbol{\mu}$ (as well as other multivariate tests in general).
- It's equivalent to Hotelling's T^2 for $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ or $H_o : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$.
- It's more general than T^2 in that it can be used to test other hypotheses (e.g., those regarding $\boldsymbol{\Sigma}$) and in different circumstances.
- Foreshadow: When testing more than 1 or 2 mean vectors, there are lots of different test statistics (about 5 common ones).
- T^2 and likelihood ratio tests are based on different underlying principles.

Figura 10: Likelihood Ratio

- Θ_o = a set of unknown parameters under H_o (e.g., $\boldsymbol{\Sigma}$).
- Θ = the set of unknown parameters under the alternative hypothesis (model), which is more general (e.g., $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$).
- $\mathcal{L}(\cdot)$ is the likelihood function. It is a function of parameters that indicates "how likely Θ (or Θ_o) is given the data".
- $\mathcal{L}(\Theta) \geq \mathcal{L}(\Theta_o)$.
  - The more general model/hypothesis is always more (or equally) likely than the more restrictive model/hypothesis.

The Likelihood Ratio Statistic is

$$\Lambda = \frac{\max \mathcal{L}(\Theta_o)}{\max \mathcal{L}(\Theta)} \quad \rightarrow \quad \begin{array}{ll} \bar{\mathbf{X}} = \hat{\boldsymbol{\mu}} & \text{MLE of mean} \\ \mathbf{S}_n = \hat{\boldsymbol{\Sigma}} & \text{MLE of covariance matrix} \end{array}$$

If Λ is "small", then the data are not likely to have occurred under $H_o \longrightarrow$ Reject H_o.

If Λ is "large", then the data are likely to have occurred under $H_o \longrightarrow$ Retain H_o.

Figura 11: Basic Idea

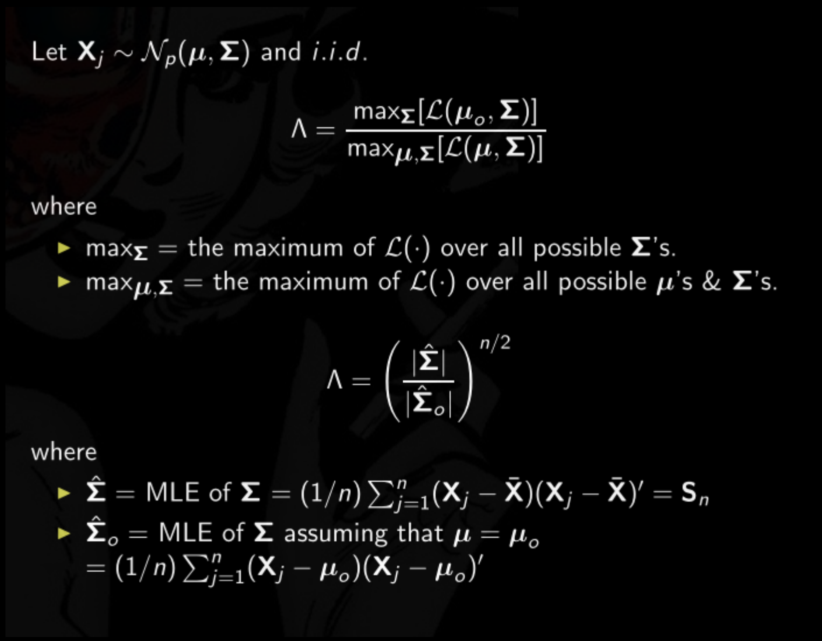


Figura 12: Vector De Medias

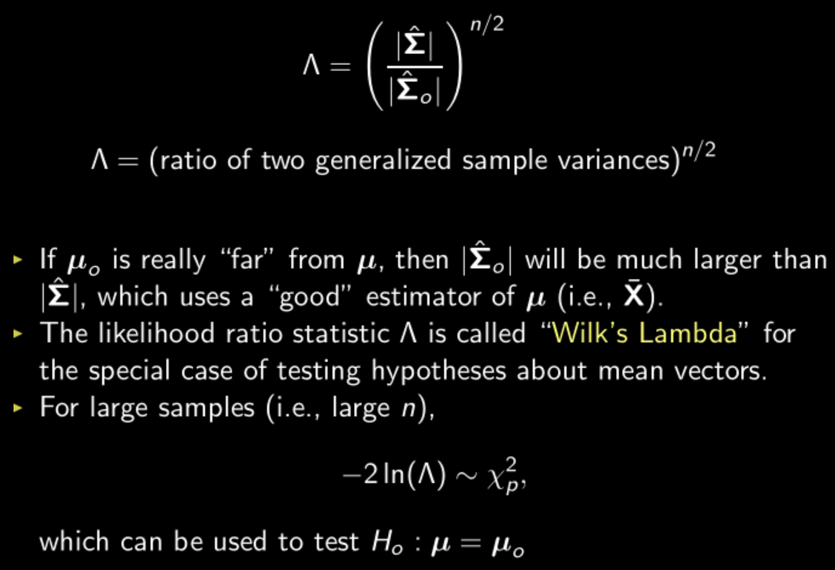


Figura 13: Vector de medias 2

We need to consider the number of parameter estimates under each hypothesis:
The alternative hypothesis ("full model"),

$$\Theta = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\} \longrightarrow p \text{ means} \quad + \quad \frac{p(p-1)}{2} \text{ covariances}$$

The null hypothesis,

$$\Theta_o = \{\boldsymbol{\Sigma}\} \longrightarrow \frac{p(p-1)}{2} \text{ covariances}$$

$$\begin{aligned} \text{degrees of freedom} = df \quad &= \quad \text{difference between number of parameters} \\ &\quad\;\; \text{estimated under each hypothesis} \\ &= \quad p \end{aligned}$$

If the H_o is true and all assumptions valid, then for large samples, $-2\ln(\Lambda) \sim \chi^2_p$.

Figura 14: Grados de libertad