

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

$$H_o : \mu = \mu_o \quad (1)$$

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...

Univariate Case

We're interested in the mean of a population and we have a random sample of n observations from the population,

una muestra, si tienes p variables
este conjunto de notas son de la X_1, X_2, \dots, X_n

where (i.e., **Assumptions**):

- Observations are **independent** (i.e., X_j is independent from $X_{j'}$ for $j \neq j'$).
- Observations are from the **same population**; that is,

$$E(X_j) = \mu \text{ for all } j$$

- If the sample size is "**small**", we'll also assume that

$$X_j \sim \mathcal{N}(\mu, \sigma^2)$$

Figura 1: Caso Univariado

hipotesis y test ...

► **Hypothesis:**

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

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 decision es clasica para la distribucion $t - student$

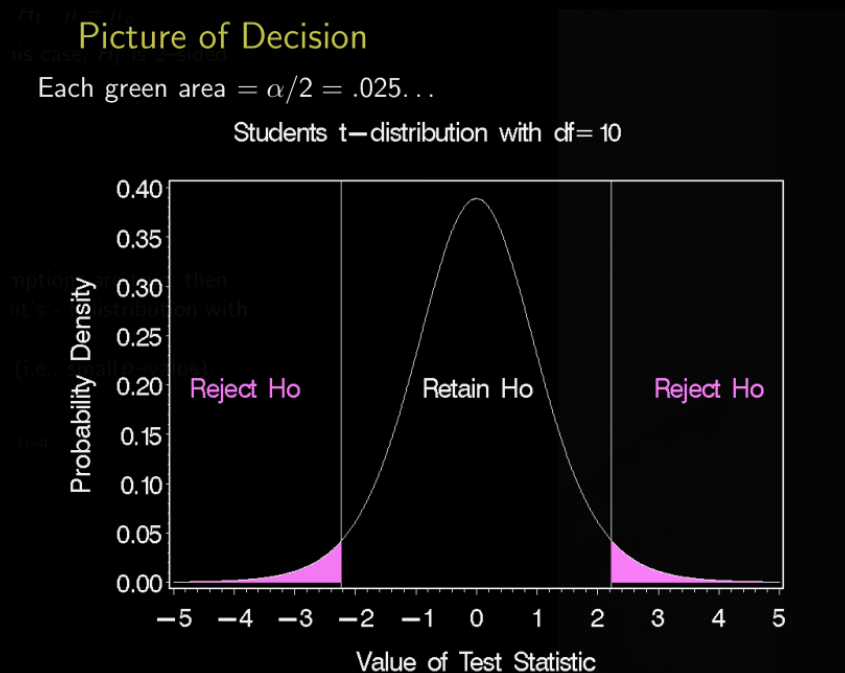


Figura 3: Desición

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 $t_{df}^2 = F_{1, df}$

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

For the extension from the univariate to multivariate case, replace scalars with vectors and matrices:

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_o)' \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_o)$$

$\bar{\mathbf{X}} = (\bar{\mathbf{x}})$

► $\bar{\mathbf{X}}_{p \times 1} = (1/n) \sum_{j=1}^n \mathbf{X}_j$

(mean) ► $\boldsymbol{\mu}_{o, (p \times 1)} = (\mu_{1o}, \mu_{2o}, \dots, \mu_{po})$

► $\mathbf{S}_{p \times p} = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})'$

T^2 is "Hotelling's T^2 "

The sample distribution of T^2

$$T^2 \sim \frac{(n-1)p}{n-p} \mathcal{F}_{p, (n-p)}$$

We can use this to test $H_o: \boldsymbol{\mu} = \boldsymbol{\mu}_o \dots$ assuming that observations are a random sample from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ i.i.d.

Figura 5: extension al caso multivariado

Since

$$T^2 \sim \frac{(n-1)p}{n-p} \mathcal{F}_{p, (n-p)}$$

We can compute T^2 and compare it to

$$\frac{(n-1)p}{n-p} \mathcal{F}_{p, (n-p)}(\alpha)$$

OR use the fact that

$$\frac{n-p}{(n-1)p} T^2 \sim \mathcal{F}_{p, (n-p)}$$

Compute T^2 as

$$T^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_o)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_o)'$$

and the

$$p\text{-value} = \text{Prob} \left\{ \mathcal{F}_{p, (n-p)} \geq \frac{(n-p)}{(n-1)p} T^2 \right\}$$

Reject H_o when p -value is small (i.e., when T^2 is large).

Figura 6: Hoteling

3. cosas cool

4. Likelihood ratio

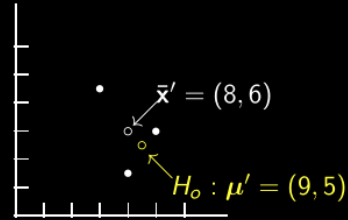
T^2 esta basado en el principio de la union, que agarra un caso multivariado y lo vuelve un problema uniariado, el problema tipo $T^2 \stackrel{\text{def}}{=} \mathbf{a}'(\hat{X} - \boldsymbol{\mu} - 0)$ es una combinacion lineal

Example 1

$n = 3$ and $p = 2$

$$\text{Data: } \mathbf{X} = \begin{pmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{pmatrix}$$

$$H_o : \boldsymbol{\mu} = \begin{pmatrix} 9 \\ 5 \end{pmatrix}$$



Assuming data come from a multivariate normal distribution and independent observations,

$$\bar{\mathbf{x}} = \begin{pmatrix} 8 \\ 6 \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} 4 & -3 \\ -3 & 9 \end{pmatrix}$$

$$\mathbf{S}^{-1} = \frac{1}{4(9) - (-3)(-3)} \begin{pmatrix} 9 & 3 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix}$$

Figura 7: ejemplo1

$$\begin{aligned} T^2 &= n(\bar{\mathbf{x}} - \boldsymbol{\mu}_o)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}_o) \\ &= 3((8-9), (6-5)) \begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix} \begin{pmatrix} (8-9) \\ (6-5) \end{pmatrix} \\ &= 3(-1, 1) \begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix} \\ &= 3(7/27) = 7/9 \end{aligned}$$

Value we need for $\alpha = .05$ is $\mathcal{F}_{2,1}(.05) = 199.51$.

$$\frac{(3-1)2}{3-2} 199.51 = 4(199.51) = 798.04.$$

Since $T^2 \sim \frac{(n-1)p}{(n-p)} \mathcal{F}_{p, n-p}$, we can compare our T^2 to 798.04.

Alternatively, we could compute p -value: compare $.25(7/9) = 0.194$ to $\mathcal{F}_{2,1}$ and we get p -value = .85.

Do not reject H_o . ($\bar{\mathbf{x}}$ and $\boldsymbol{\mu}$ are "close" in the figure).

Figura 8: ejemplo2 -continuacion

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

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

Figura 9: Cosa cool

- ▶ Another approach to testing null hypothesis about mean vector μ (as well as other multivariate tests in general).
- ▶ It's equivalent to Hotelling's T^2 for $H_o : \mu = \mu_o$ or $H_o : \mu_1 = \mu_2$.
- ▶ It's more general than T^2 in that it can be used to test other hypotheses (e.g., those regarding Σ) 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., Σ).
- ▶ Θ = the set of unknown parameters under the alternative hypothesis (model), which is more general (e.g., μ and Σ).
- ▶ $\mathcal{L}(\cdot)$ 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

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

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

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

Figura 11: Basic Idea

Let $\mathbf{X}_j \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and *i.i.d.*

$$\Lambda = \frac{\max_{\boldsymbol{\Sigma}} [\mathcal{L}(\boldsymbol{\mu}_o, \boldsymbol{\Sigma})]}{\max_{\boldsymbol{\mu}, \boldsymbol{\Sigma}} [\mathcal{L}(\boldsymbol{\mu}, \boldsymbol{\Sigma})]}$$

where

- $\max_{\boldsymbol{\Sigma}}$ = the maximum of $\mathcal{L}(\cdot)$ over all possible $\boldsymbol{\Sigma}$'s.
- $\max_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}$ = the maximum of $\mathcal{L}(\cdot)$ over all possible $\boldsymbol{\mu}$'s & $\boldsymbol{\Sigma}$'s.

$$\Lambda = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_o|} \right)^{n/2}$$

where

- $\hat{\boldsymbol{\Sigma}} = \text{MLE of } \boldsymbol{\Sigma} = (1/n) \sum_{j=1}^n (\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})' = \mathbf{S}_n$
- $\hat{\boldsymbol{\Sigma}}_o = \text{MLE of } \boldsymbol{\Sigma} \text{ assuming that } \boldsymbol{\mu} = \boldsymbol{\mu}_o$
 $= (1/n) \sum_{j=1}^n (\mathbf{X}_j - \boldsymbol{\mu}_o)(\mathbf{X}_j - \boldsymbol{\mu}_o)'$

Figura 12: Vector De Medias

$$\Lambda = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_o|} \right)^{n/2}$$

$$\Lambda = (\text{ratio of two generalized sample variances})^{n/2}$$

- If $\boldsymbol{\mu}_o$ is really “far” from $\boldsymbol{\mu}$, then $|\hat{\boldsymbol{\Sigma}}_o|$ will be much larger than $|\hat{\boldsymbol{\Sigma}}|$, which uses a “good” estimator of $\boldsymbol{\mu}$ (i.e., $\bar{\mathbf{X}}$).
- The likelihood ratio statistic Λ is called “**Wilk's Lambda**” for the special case of testing hypotheses about mean vectors.
- For large samples (i.e., large n),

$$-2\ln(\Lambda) \sim \chi_p^2,$$

which can be used to test $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$

Figura 13: Vector de medias 2

We need to consider the number of parameter estimates under each hypothesis:

The alternative hypothesis ("full model") ,

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

The null hypothesis,

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

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

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

Figura 14: Grados de libertad