Tema 3. Contrastos d'hipòtesis paramètriques

3.1. Definició d'Espai Paramètric

CONCEPTES A REVISAR: (Tema 1)

- Població vs. Mostra
- Funció de quantia/densitat vs. Funció de versemblança
- Espai Mostral: Conjunt de totes les possibles mostres de mida n.
- Representativitat, Selecció probabilística (MAS).

MODEL ESTADÍSTIC: Model que pretén explicar i/o predir el comportament d'un fenomen aleatori. Està format per:

A) MODEL PROBABILÍSTIC:

$$X \approx f(X, \theta)$$

que depèn de:

- $f(\cdot)$ • X • θ

En general $\begin{cases} -\infty < X < \infty \\ \theta \in \Theta \end{cases}$

ESPAI PARAMÈTRIC

B) MODEL MOSTRAL

Constituït per la MOSTRA, que depèn de:

- Mida
- Forma d'extracció

És el NEXE d'UNIÓ entre:

MOD. PROBABILÍSTIC $\Longleftrightarrow$ REALITAT

Tema 3. Contrastos d'hipòtesis paramètriques

3.2. Partició de l'espai paramètric: Hipòtesis simples, hipòtesis compostes. Hipòtesi nul·la i Hipòtesi alternativa.

- Suposem una funció de probabilitat d'una variable aleatòria Φ.
  f(X;θ) on:
  - f(·) pot ser discreta o continua,
  - depèn de θ Є Θ,
  - X és la mostra.

- HIPÒTESI ESTADÍSTICA: És una conjectura sobre el valor concret que θ tingui en la realitat.
- Per establir una hipòtesi hem de partir l'Espai paramètric en dues parts:
  - θ_0 Conjunt de valors que compleixen la hipòtesi.
  - θ_1 Conjunt de valors que NO compleixen la hipòtesi.
  - Aquests dos conjunts de valors han de ser complementaris $\theta = \theta_0 \cup \theta_1$.

- Definim HIPÒTESI NUL·LA a la hipòtesi que implica el valor existent del paràmetre. La que suposem MÉS ESTABLE. $\longrightarrow$ $H_0[\theta_0 \in \Theta_0]$

- Definim HIPÒTESI ALTERNATIVA a l'altra. $\longrightarrow$ $H_1[\theta_1 \in \Theta_1]$

Necessitem SUFICIENT EVIDÈNCIA PER A REFUSAR-LA.

Tema 3. Contrastos d'hipòtesis paramètriques

• Si θ_0 i θ_1 es composen només d'un únic element diem que són HIPÒTESIS SIMPLES. Si es composen de més d'un element diem que són HIPÒTESIS COMPOSTES.

Exemples $\Rightarrow$ $\begin{cases} H_0 : \mu = a \\ H_1 : \mu = b \end{cases}$ $\begin{cases} H_0 : \mu = a \\ H_1 : \mu > a \end{cases}$ $\begin{cases} H_0 : \mu = a \\ H_1 : \mu < a \end{cases}$ $\begin{cases} H_0 : \mu = a \\ H_1 : \mu \neq a \end{cases}$

NOTA: Amb una hipòtesi simple la distribució de probabilitat està ben definida. En cas d'hipòtesis compostes, no.

3.3. Decisions estadístiques i tipus d'errors. Probabilitats dels errors.

CONTRASTOS D'HIPÒTESIS: És la regla de decisió mitjançant la qual optem per una de les hipòtesis plantejades. Per això:

- Es fa una PARTICIÓ DE L'ESPAI MOSTRAL en dos CONJUNTS DISJUNTS C i C*.
- Si el PUNT MOSTRA (o estadístic) pertany a la REGIÓ CRÍTICA (C) es refusa H_0.
- Si el PUNT MOSTRA (o estadístic) pertany a la REGIÓ d'ACCEPTACIÓ (C*) NO es refusa H0. (NO DIEM QUE S'ACCEPTA H_1).

ECET 3.1

NO REFUSAR (REFUSAR) → Que la mostra ha (o no ha) proporcionat "suficient" evidència per a que sigui raonable prendre aquesta decisió. (NO ES UN CONCEPTE ABSOLUT).

Tema 3. Contrastos d'hipòtesis paramètriques

Què pot succeir quan prenem aquesta decisió?? ⇨

		Decisió	
		No Refusar H_0	Refusar H_0
Realitat	H_0	√	Error Tipus I
	H_1	Error Tipus II	√

El que es pretendrà és establir un tipus de control sobre aquest esquema mitjançant el coneixement de les probabilitats de cometre aquests errors. Així:

$$* P(Error\ de\ Tipus\ I) = P(\text{refusar } H_0 \text{ quan és certa}) = P\left(X \in C / H_0\right) = \alpha$$

Rep el nom de Nivell de significació

$$* P(Error\ de\ Tipus\ II) = P(\text{NO refusar } H_0 \text{ quan és falsa}) = P\left(X \notin C^* / H_1\right) = \beta$$

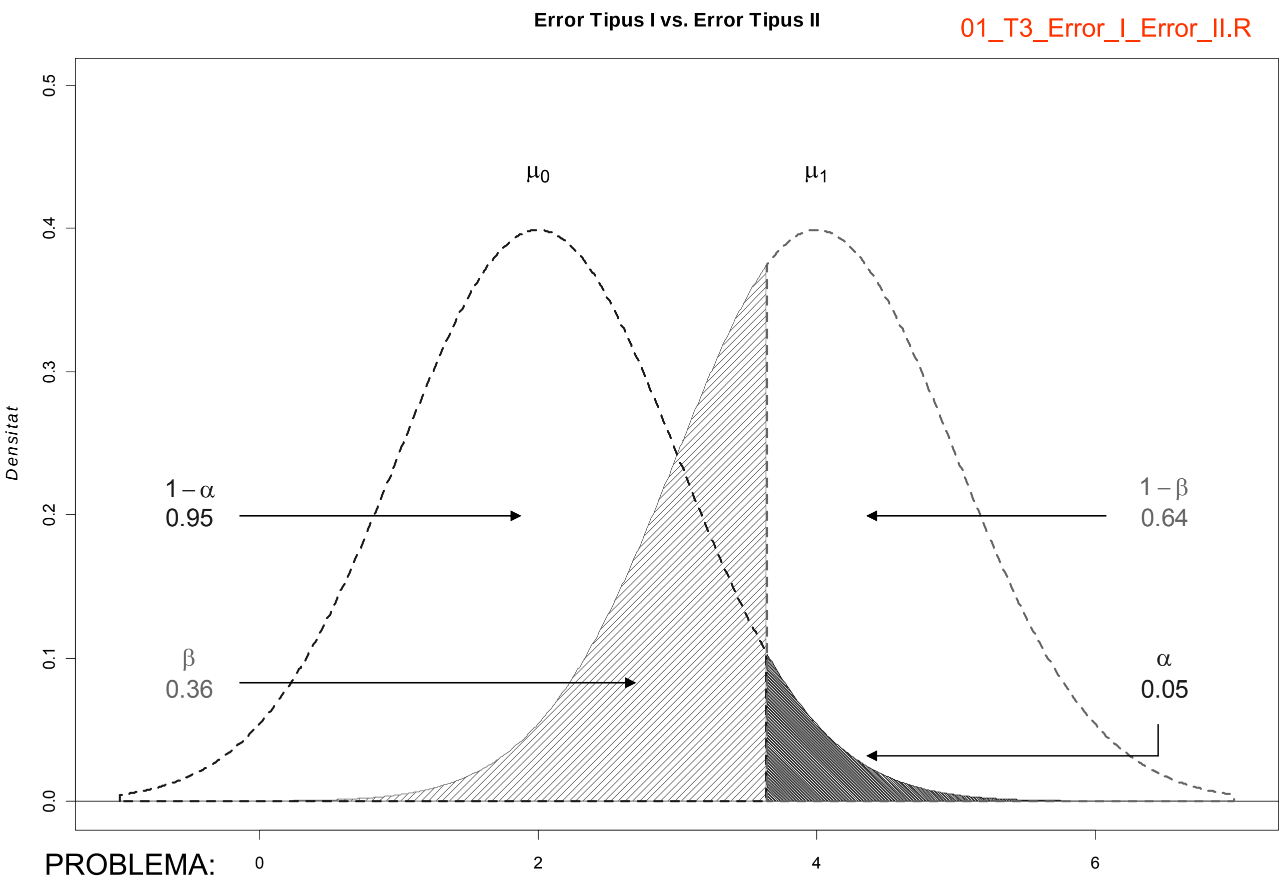
No té nom específic

$$* \quad P(\text{NO refusar } H_0 \text{ quan és certa}) = P\left(X \in C^* / H_0\right) = 1 - \alpha$$

Rep el nom de Nivell de confiança

$$* \quad P(\text{refusar } H_0 \text{ quan és falsa}) = P\left(X \in C / H_1\right) = 1 - \beta$$

Rep el nom de Potència



PROBLEMA:
No sabem en quina de les quatre posicions ens trobem: Com solucionem el problema?

Tema 3. Contrastos d'hipòtesis paramètriques

Davant aquest problema, la solució passa per fixar el nivell de significació "α" el més baix possible, atès que és la probabilitat de cometre un error i al mateix temps maximitzar la potència "1-β", atès que és la probabilitat d'un encert.

PROBLEMA: Joc de la Corda. No es pot fixar "α" i "1-β" simultàniament.
SOLUCIÓ:

- "És més greu cometre un error de Tipus I que un error de tipus II".
- De la mateixa forma que considerem que és més greu condemnar a un innocent que absoldre a un culpable en un judici.

H_0 → "Innocent"
H_1 → "Culpable"

Per tant:

1. Es fixa el NIVELL DE SIGNIFICACIÓ "α" en valors reduïts → 0.10, 0.05, 0.01. escollits arbitràriament.
2. Atès que per fer el contrast hem de partir l'espai mostral en 2 regions, el nombre de contrastos possibles es tan elevat com el nombre de particions possibles.
3. Escollim de totes les particions aquella que maximitzi la potència "1-β".
4. Així, davant un mateix nivell de significació "α" si un contrast presenta més potència que qualsevol altre, direm que és el CONTRAST UNIFORMEMENT MÉS POTENT. El problema es que no sempre existeix

   Això condueix al LEMA DE NEYMAN-PEARSON (RAÓ DE VERSEMBLANÇA

ECET 3.2

ECET 3.3

Tema 3. Contrastos d'hipòtesis paramètriques

NEYMAN-PEARSON (CONTRAST DE RAÓ DE VERSEMBLANÇA)

1. Sobre el paràmetre θ d'una funció de quantia o densitat f(x,θ) s'estableixen dues hipòtesis SIMPLES. → $\begin{cases} H_0: & \theta=\theta_0 \\ H_1: & \theta=\theta_1 \end{cases}$
2. Es pren una mostra aleatòria simple de mida n i es construeix la funció de versemblança $L(X,\theta)$
3. Es particularitza la funció per a cada hipòtesi → $\begin{cases} L(X,\theta_0) \\ L(X,\theta_1) \end{cases}$
4. Dividim l'espai mostral en dos subconjunts disjunts C i C*
5. Quan el punt mostra (estadístic) X Є C (regió crítica). Refusem H0 → $\dfrac{L(X,\theta_0)}{L(X,\theta_1)} \le k$
6. Quan el punt mostra (estadístic) X Є C* (regió d'acceptació) No refusem H0 → $\dfrac{L(X,\theta_0)}{L(X,\theta_1)} > k$
7. En aquestes condicions, si el contrast es realitza amb un NIVELL DE SIGNIFICACIÓ "α", la REGIÓ CRÍTICA presenta una major o igual POTENCIA "1-β que qualsevol altre regió de la mateixa mida. → $P\left(\dfrac{L(X,\theta_0)}{L(X,\theta_1)} \le k \middle/ H_0\right) = \alpha$

• No vol dir que sigui ÚNICA, sinó que és una de les varies que poden existir amb la mateixa potència i és indiferent fer servir una u altre.

• No existeixen proves uniformement més potents per H_1 bilaterals.

ECET 3.4

Tema 3. Contrastos d'hipòtesis paramètriques

3.4. Contrastos d'hipòtesis sobre la mitjana, la variància i la proporció poblacionals.

SOBRE LA MITJANA POBLACIONAL (1 mostra)

SUPÒSIT 1: Sigui $X_1, X_2, ..., X_n$ una mostra aleatòria d'una distribució normal amb mitjana μ DESCONEGUDA i variància σ^2 CONEGUDA.

CAS 1 $\begin{cases} H_0: \mu=\mu_0 \\ H_1: \mu\neq\mu_0 \end{cases}$

1º.- El millor estadístic per a contrastar μ és $\bar{X}$

2º.- Sabem que sota H_0 $\bar{X} \approx N\left(\mu_0, \sigma/\sqrt{n}\right)$

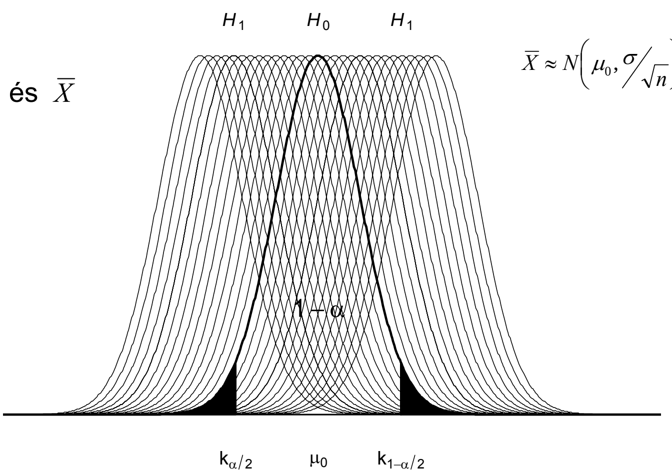
3º.- Refusem H_0 si $\bar{X}$ és SUFICIENTMENT DIFERENT de μ_0

3.1.- Preseleccionem α

3.2.- Obtenim la regió crítica BILATERAL

3.3.- Refusem H0 si $\begin{cases} P\left(\bar{X} \ge k_{1-\alpha/2}\right) = \alpha/2 \\ P\left(\bar{X} \le k_{\alpha/2}\right) = \alpha/2 \end{cases}$

$\bar{X} \approx N\left(\mu_0, \sigma/\sqrt{n}\right)$



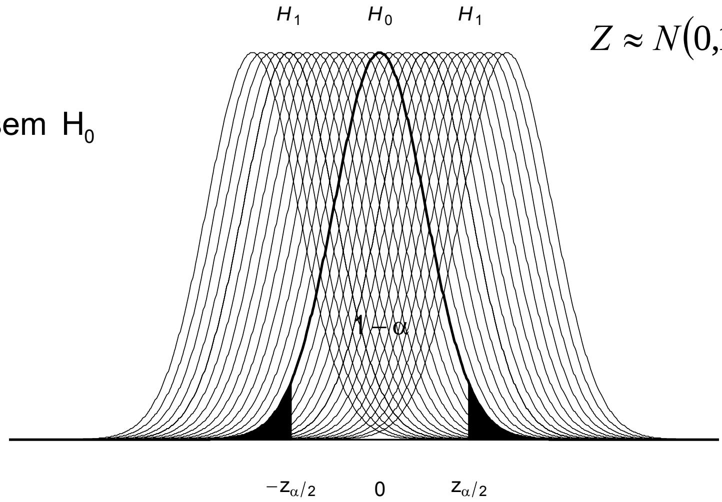
$$P\left(\overline{X} \geq k_{1-\alpha/2}\right) = P\left(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \geq \frac{k_{1-\alpha/2}-\mu_0}{\sigma/\sqrt{n}}\right) = P\left(Z \geq z_{\alpha/2}\right) = \alpha/2$$

$$P\left(\overline{X} \leq k_{\alpha/2}\right) = P\left(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \leq \frac{k_{\alpha/2}-\mu_0}{\sigma/\sqrt{n}}\right) = P\left(Z \leq -z_{\alpha/2}\right) = \alpha/2$$

i, per tant només amb

$$\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right| \geq z_{\alpha/2} \quad \Rightarrow \quad \text{Refusem } H_0$$

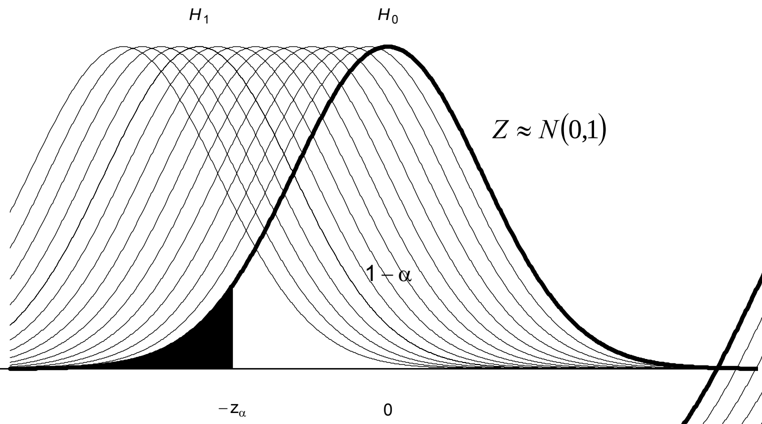
$Z \approx N(0,1)$



CAS 2 $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu < \mu_0 \end{cases}$

IDEM

$$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{N}} \leq -Z_\alpha \quad \Rightarrow \quad \text{Refusem } H_0$$

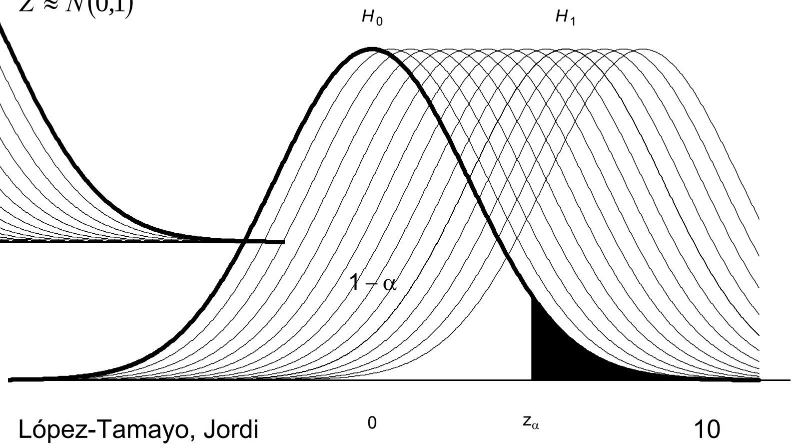


CAS 3 $\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu > \mu_0 \end{cases}$

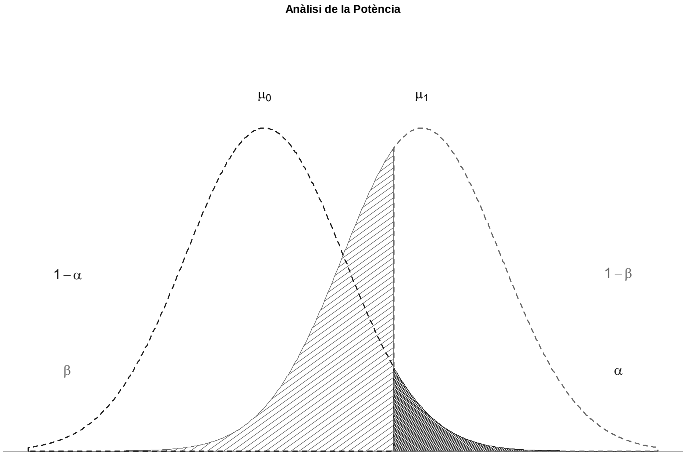
IDEM

$$\frac{\overline{X}-\mu_0}{\sigma/\sqrt{N}} \geq Z_\alpha \quad \Rightarrow \quad \text{Refusem } H_0$$

$Z \approx N(0,1)$



Tema 3. Contrastos d’hipòtesis paramètriques



$$\alpha = P\left(\overline{X} \geq k_\alpha\right) = P\left(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \geq \frac{k_\alpha - \mu_0}{\sigma/\sqrt{n}}\right) =$$

$$= P\left(\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \geq z_\alpha\right) = P\left(\overline{X} \geq \mu_0 + z_\alpha \sigma/\sqrt{n}\right)$$

$$1-\beta = P\left(\overline{X} \geq \mu_0 + z_\alpha \sigma/\sqrt{n} \Big/ H_1 : N\left(\mu_1, \sigma/\sqrt{n}\right)\right) =$$

$$= P\left(\frac{\overline{X}-\mu_1}{\sigma/\sqrt{n}} \geq \frac{\left(\mu_0 + z_\alpha \sigma/\sqrt{n}\right) - \mu_1}{\sigma/\sqrt{n}}\right)$$

<u>LA POTÈNCIA DEPÈN DE:</u>

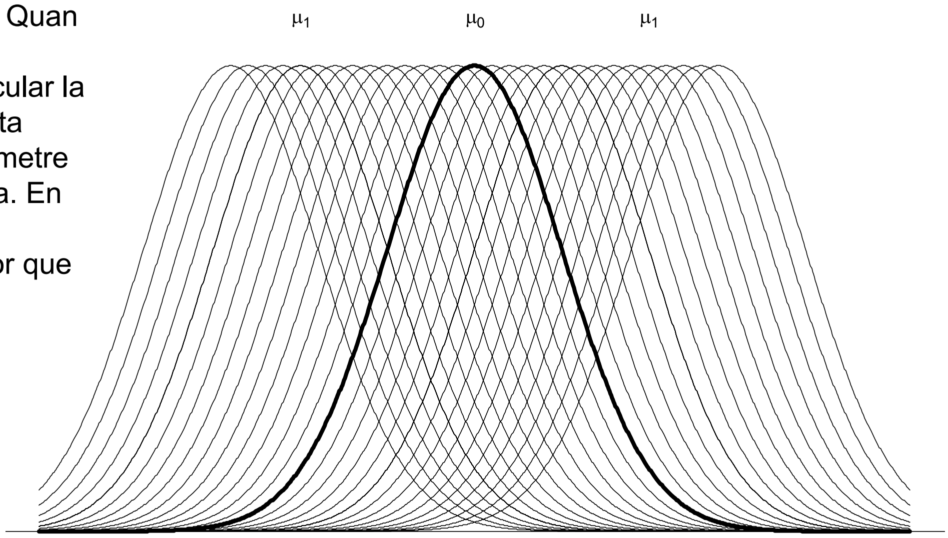
1. El nivell de significació → α
2. La mida de la mostra → n
3. De la diferència de mitjanes → μ_0-μ_1
4. De la desviació estàndard de la població → σ
5. Per valors fixes d'n, $\mu_0 - \mu_1$ i σ:

$$\begin{cases} \Delta\alpha \Rightarrow \Delta 1\text{-}\beta \Rightarrow \nabla\beta \\ \nabla\alpha \Rightarrow \nabla 1\text{-}\beta \Rightarrow \Delta\beta \end{cases}$$

ECET 3.5 **ECET 3.6**

Tema 3. Contrastos d’hipòtesis paramètriques

<u>FUNCIÓ DE POTÈNCIA.</u> Quan la hipòtesi alternativa és composta, no podem calcular la potència, atès que aquesta depèn de quin és el paràmetre sota la hipòtesi alternativa. En aquest cas, tot l’aparell matemàtic depèn del valor que finalment prengui μ_1.



$$1-\beta = \begin{cases} \displaystyle\int_{-\infty}^{k} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(k-\mu_1)^2}{2\sigma^2}} dx & \text{Si } \mu_0 < \mu_1 \\ \displaystyle\int_{k}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(k-\mu_1)^2}{2\sigma^2}} dx & \text{Si } \mu_0 > \mu_1 \end{cases}$$

Per tant, tenim tantes probabilitats com possibles paràmetres Així, si grafiquem totes aquestes possibles probabilitats tenim la funció de potència

Tema 3. Contrastos d'hipòtesis paramètriques

<u>SOBRE LA MITJANA POBLACIONAL (1 mostra)</u>

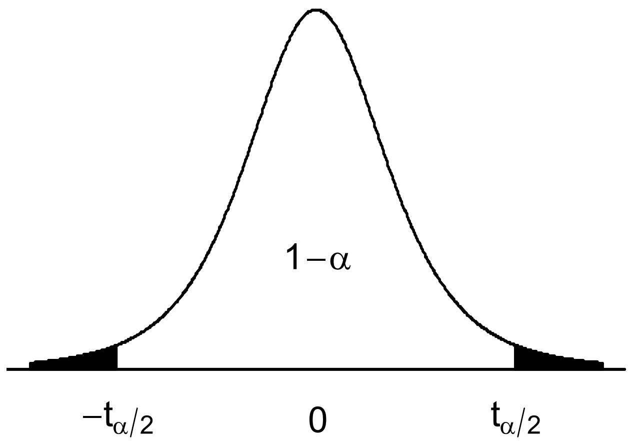
<u>SUPÒSIT 2</u>: Sigui X_1, X_2,....,X_n una mostra aleatòria d'una distribució normal amb mitjana μ DESCONEGUDA i variància σ^2 DESCONEGUDA.

En aquest cas és exactament igual tret que al no conéixer la variància poblacional, hem de proposar un estimador d'aquesta a partir de la mostra. En aquest cas, quan es procedeix a la tipificació de la variable, aquesta ja no es distribueix segons una N(01,), sinó com una T-STUDENT d'n-1 graus de llibertat.

$$\overline{X} \approx N\left(\mu, \sigma/\sqrt{n}\right) \Rightarrow \frac{\overline{X}-\mu}{S/\sqrt{n}} \approx t_{n-1}$$

CAS 1 $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$

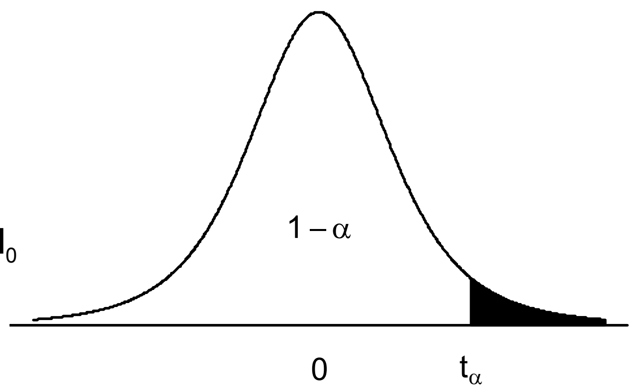
$$\left|\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\right| \geq t_{\alpha/2,n-1} \Rightarrow \text{Refusem } H_0$$



Tema 3. Contrastos d'hipòtesis paramètriques

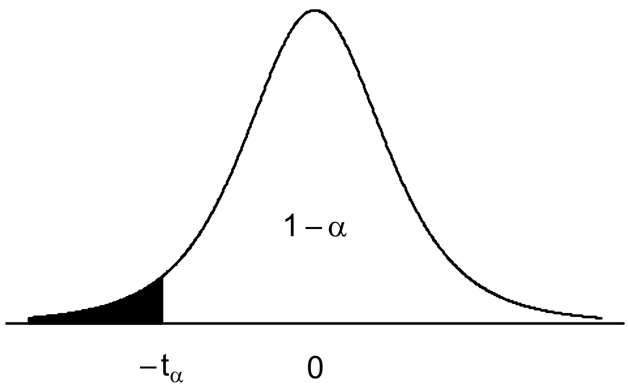
CAS 2 $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$

$$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \geq t_{\alpha,n-1} \Rightarrow \text{Refusem } H_0$$



CAS 3 $\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$

$$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \leq -t_{\alpha,n-1} \Rightarrow \text{Refusem } H_0$$



ECET 3.7

Tema 3. Contrastos d'hipòtesis paramètriques

SOBRE LA VARIÀNCIA POBLACIONAL (1 mostra)

Sigui $X_1, X_2, \ldots, X_n$ una mostra aleatòria d'una distribució normal amb mitjana μ DESCONEGUDA i variància σ^2 CONEGUDA.

CAS 1 $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 \neq \sigma_0^2 \end{cases}$

1º.- El millor estadístic per a contrastar σ^2 és S_*^2

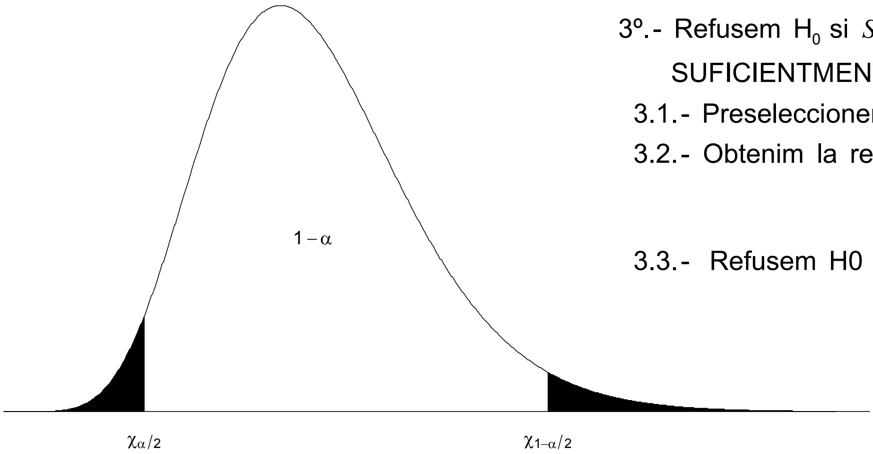
2º.- Sabem que sota H_0 $\frac{(n-1)S_*^2}{\sigma_0^2} \approx \chi_{n-1}^2$

3º.- Refusem H_0 si S_*^2 és SUFICIENTMENT DIFERENT de σ_0^2

3.1.- Preseleccionem α

3.2.- Obtenim la regió crítica BILATERAL

3.3.- Refusem H0 si $\begin{cases} P\left(\frac{(n-1)S_*^2}{\sigma_0^2} \geq \chi_{1-\alpha/2,n-1}\right) = \alpha/2 \\ P\left(\frac{(n-1)S_*^2}{\sigma_0^2} \leq \chi_{\alpha/2,n-1}\right) = \alpha/2 \end{cases}$

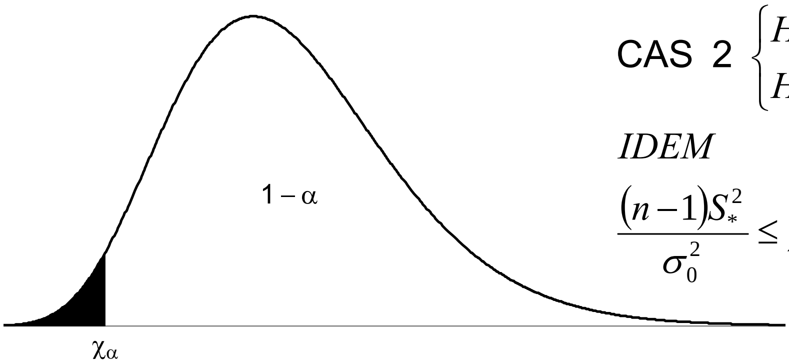


Tema 3. Contrastos d'hipòtesis paramètriques

CAS 2 $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases}$

IDEM

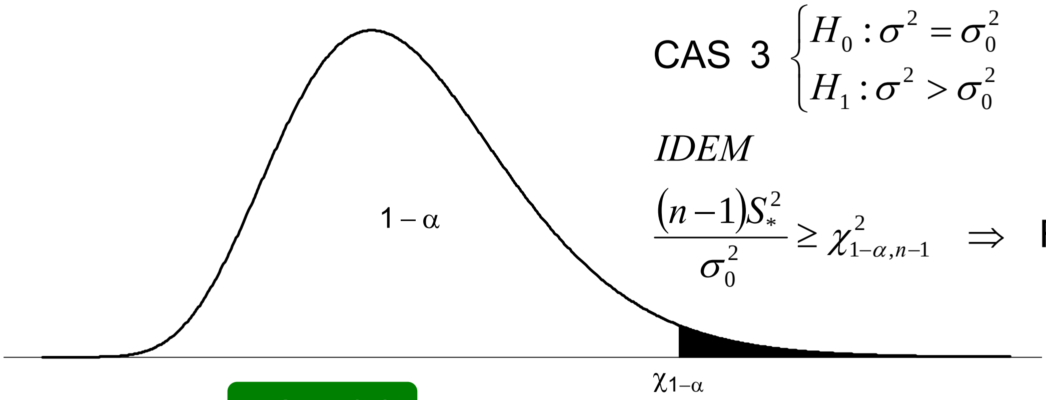
$$\frac{(n-1)S_*^2}{\sigma_0^2} \leq \chi_{\alpha,n-1}^2 \quad \Rightarrow \quad \text{Refusem } H_0$$



CAS 3 $\begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases}$

IDEM

$$\frac{(n-1)S_*^2}{\sigma_0^2} \geq \chi_{1-\alpha,n-1}^2 \quad \Rightarrow \quad \text{Refusem } H_0$$



ECET 3.8

Tema 3. Contrastos d'hipòtesis paramètriques

SOBRE LA PROPORCIÓ POBLACIONAL (1 mostra)

En alguns casos s'ha de contrastar una proporció. En principi sabem que la distribució no és normal, però també sabem que sota determinades circumstàncies (TCL), es pot aproximar per una normal, atès que la distribució en el mostreig és assimptòticament normal.

CAS 1 $\begin{cases} H_0 : \Pi = \Pi_0 \\ H_1 : \Pi \neq \Pi_0 \end{cases}$

1º.- El millor estadístic per a contrastar Π és p

2º.- Sabem que sota H_0, si n és suficientment gran

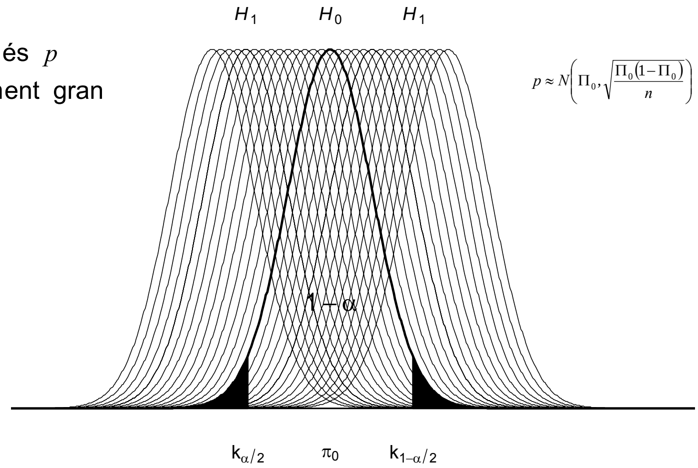
i $np(1-p) > 5$ $\quad p \approx N\left(\Pi_0, \sqrt{\dfrac{\Pi_0(1-\Pi_0)}{n}}\right)$

3º.- Refusem H_0 si p és SUFICIENTMENT DIFERENT de Π_0

3.1.- Preseleccionem α

3.2.- Obtenim la regió crítica BILATERAL

3.3.- Refusem H0 si $\begin{cases} P\left(p \geq k_{1-\alpha/2}\right) = \alpha/2 \\ P\left(p \leq k_{\alpha/2}\right) = \alpha/2 \end{cases}$



$p \approx N\left(\Pi_0, \sqrt{\dfrac{\Pi_0(1-\Pi_0)}{n}}\right)$

Tema 3. Contrastos d'hipòtesis paramètriques

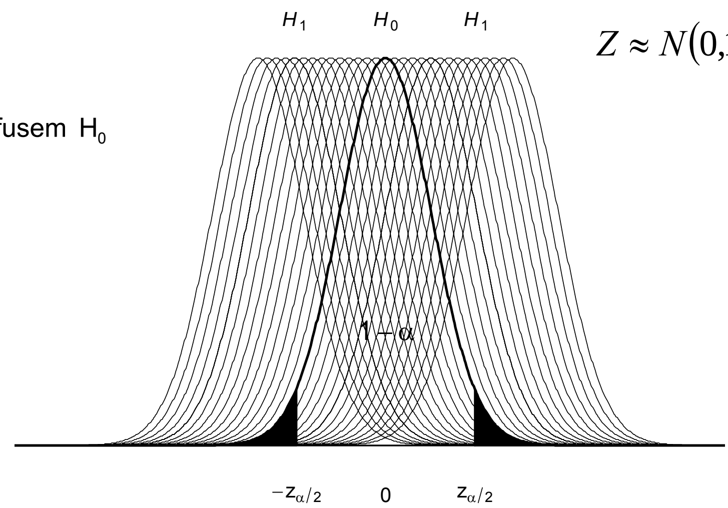
$$P\left(p \geq k_{1-\alpha/2}\right) = P\left(\frac{p-\Pi_0}{\sqrt{\dfrac{\Pi_0(1-\Pi_0)}{n}}} \geq \frac{k_{1-\alpha/2} - \Pi_0}{\sqrt{\dfrac{\Pi_0(1-\Pi_0)}{n}}}\right) = P\left(Z \geq z_{\alpha/2}\right) = \alpha/2$$

$$P\left(p \leq k_{\alpha/2}\right) = P\left(\frac{p-\Pi_0}{\sqrt{\dfrac{\Pi_0(1-\Pi_0)}{n}}} \leq \frac{k_{1-\alpha/2} - \Pi_0}{\sqrt{\dfrac{\Pi_0(1-\Pi_0)}{n}}}\right) = P\left(Z \leq -z_{\alpha/2}\right) = \alpha/2$$

i, per tant només amb

$$\left|\frac{p-\Pi_0}{\sqrt{\dfrac{\Pi_0(1-\Pi_0)}{n}}}\right| \geq z_{\alpha/2} \quad \Rightarrow \text{ Refusem } H_0$$

$Z \approx N(0,1)$



CAS 2 $\begin{cases} H_0 : \Pi = \Pi_0 \\ H_1 : \Pi < \Pi_0 \end{cases}$

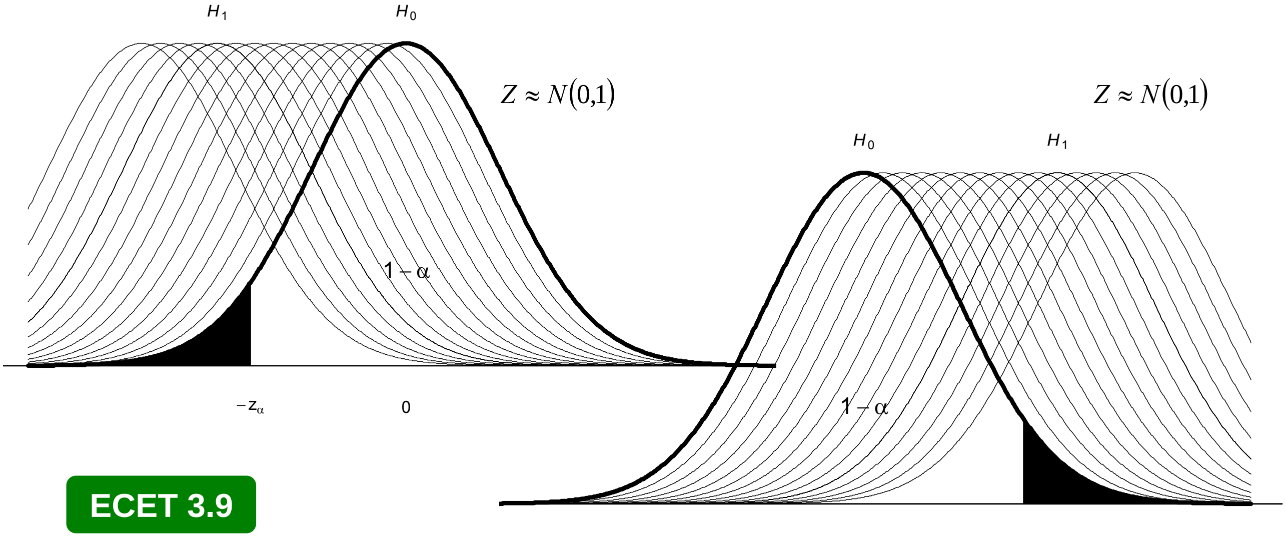
IDEM

$$\frac{p - \Pi_0}{\sqrt{\frac{\Pi_0(1-\Pi_0)}{n}}} \le -Z_\alpha \quad \Rightarrow \quad \text{Refusem } H_0$$

CAS 2 $\begin{cases} H_0 : \Pi = \Pi_0 \\ H_1 : \Pi > \Pi_0 \end{cases}$

IDEM

$$\frac{p - \Pi_0}{\sqrt{\frac{\Pi_0(1-\Pi_0)}{n}}} \ge Z_\alpha \quad \Rightarrow \quad \text{Refusem } H_0$$



ECET 3.9

SOBRE LA MITJANA POBLACIONAL (2 mostres)

SUPÒSIT 1: Sigui $X_1, X_2, \ldots, X_n$ i $Y_1, Y_2, \ldots, Y_m$ dues mostres aletòries de dos distribucions normals independents amb mitjanes μ_x i μ_y DESCONEGUDES i variàncies σ_x^2 i σ_y^2 CONEGUDES.

CAS 1 $\begin{cases} H_0 : \mu_x - \mu_y = \delta_0 \\ H_1 : \mu_x - \mu_y \ne \delta_0 \end{cases}$

$$\overline{X} - \overline{Y} \approx N\left((\mu_{x0} - \mu_{y0}), \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right)$$

1º.- El millor estadístic per a contrastar $\mu_x - \mu_y$ és $\overline{X} - \overline{Y}$

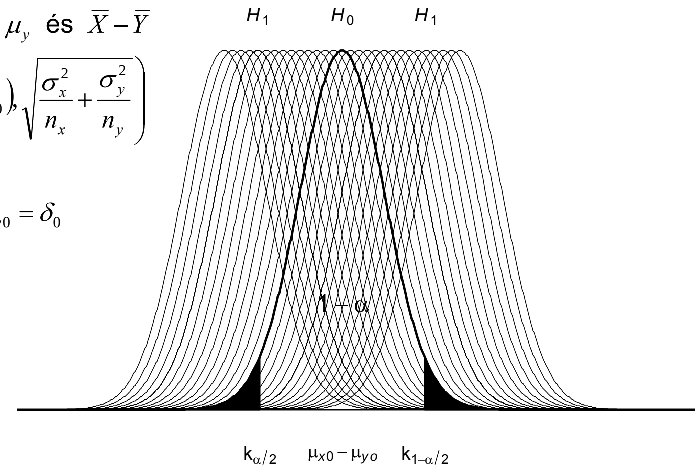
2º.- Sabem que sota H_0 $\overline{X} - \overline{Y} \approx N\left((\mu_{x0} - \mu_{y0}), \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}} \right)$

3º.- Refusem H_0 si $\overline{X} - \overline{Y} = d$ és SUFICIENTMENT DIFERENT de $\mu_{x0} - \mu_{y0} = \delta_0$

3.1.- Preseleccionem α

3.2.- Obtenim la regió crítica BILATERAL

3.3.- Refusem H0 si $\begin{cases} P\left(d \ge k_{1-\alpha/2}\right) = \alpha/2 \\ P\left(d \le k_{\alpha/2}\right) = \alpha/2 \end{cases}$



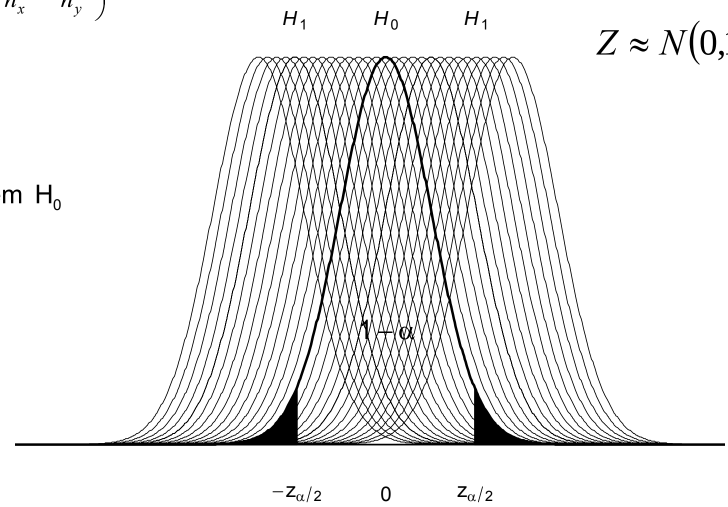
Tema 3. Contrastos d'hipòtesis paramètriques

$$P\left(d \geq k_{1-\alpha/2}\right) = P\left(\frac{d-\delta_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} \geq \frac{k_{1-\alpha/2}-\delta_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}}\right) = P\left(Z \geq z_{\alpha/2}\right) = \alpha/2$$

$$P\left(d \leq k_{\alpha/2}\right) = P\left(\frac{d-\delta_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} \leq \frac{k_{\alpha/2}-\delta_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}}\right) = P\left(Z \leq -z_{\alpha/2}\right) = \alpha/2$$

i, per tant només amb

$$\left|\frac{d-\delta_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}}\right| \geq z_{\alpha/2} \quad \Rightarrow \quad \text{Refusem } H_0$$



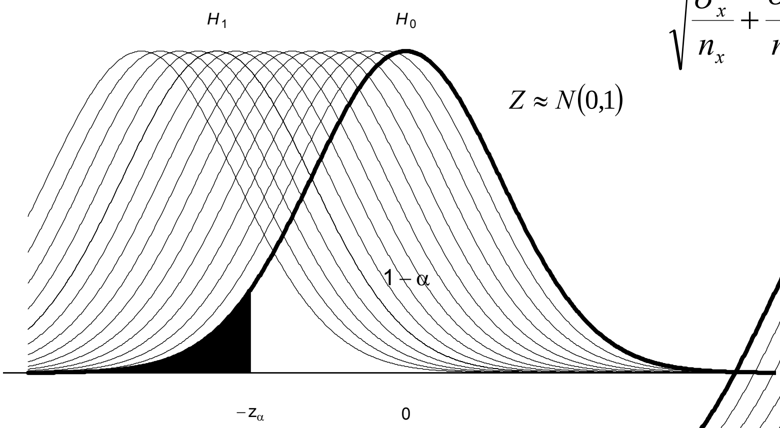
$Z \approx N(0,1)$

Tema 3. Contrastos d'hipòtesis paramètriques

CAS 2 $\begin{cases} H_0: \mu_x - \mu_y = \delta_0 \\ H_1: \mu_x - \mu_y < \delta_0 \end{cases}$

IDEM

$$\frac{d-\delta_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} \leq -Z_\alpha \quad \Rightarrow \quad \text{Refusem } H_0$$



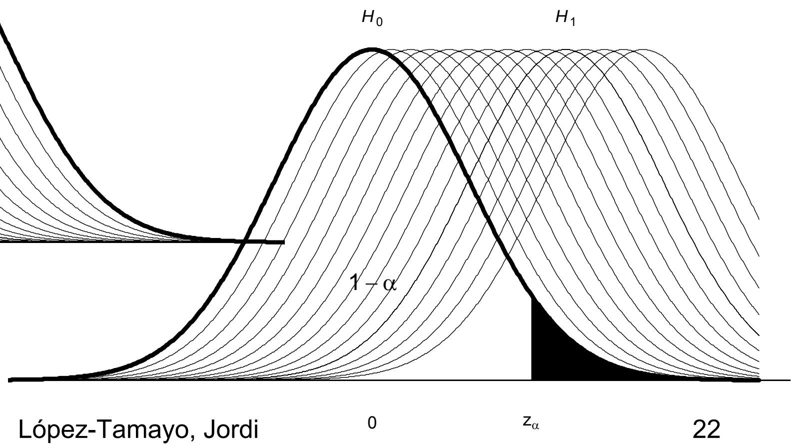
$Z \approx N(0,1)$

CAS 2 $\begin{cases} H_0: \mu_x - \mu_y = \delta_0 \\ H_1: \mu_x - \mu_y > \delta_0 \end{cases}$

IDEM

$$\frac{d-\delta_0}{\sqrt{\frac{\sigma_x^2}{n_x}+\frac{\sigma_y^2}{n_y}}} \geq Z_\alpha \quad \Rightarrow \quad \text{Refusem } H_0$$

$Z \approx N(0,1)$



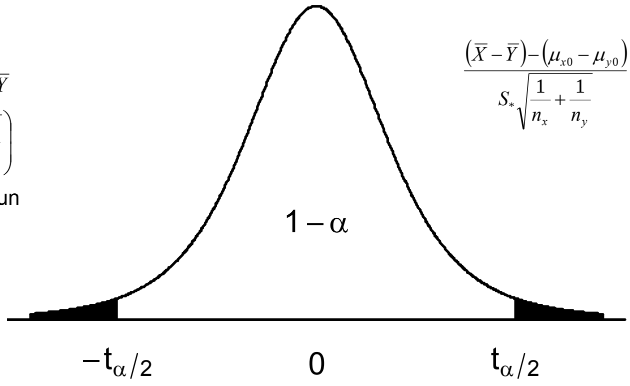
ECET 3.10

Tema 3. Contrastos d’hipòtesis paramètriques

SOBRE LA MITJANA POBLACIONAL (2 mostra)

SUPÒSIT 2: Sigui $X_1, X_2, \ldots, X_{nx}$ i $Y_1, Y_2, \ldots, Y_{ny}$ dues mostres aletòries de dues distribucions normals independents amb mitjanes μ_x i μ_y DESCONEGUDES i variàncies DESCONEGUDES però IGUALS $\sigma_x^2 = \sigma_y^2 = \sigma^2$.

CAS 1 $\begin{cases} H_0 : \mu_x - \mu_y = \delta_0 \\ H_1 : \mu_x - \mu_y \neq \delta_0 \end{cases}$



$$\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_{x0}-\mu_{y0}\right)}{S_*\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}} \approx t_{n_x+n_y-2}$$

1º.- El millor estadístic per a contrastar $\mu_x - \mu_y$ és $\bar{X}-\bar{Y}$

2º.- Sabem que sota H_0 $\bar{X}-\bar{Y} \approx N\left(\left(\mu_{x0}-\mu_{y0}\right), \sigma\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}\right)$

3º.- Ara be, com σ^2 és desconeguda, hem de proposar un estimador $\hat{\sigma}^2$ a partir de la mostra. Aquest és

$$\hat{\sigma}^2 = S_*^2 = \frac{\left(n_x-1\right)S_{*x}^2+\left(n_y-1\right)S_{*y}^2}{n_x+n_y\text{-}2}$$ i en conseqüència

$$\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_{x0}-\mu_{y0}\right)}{S_*\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}} \not\approx N(0,1),$$ sinó $t_{n_x+n_y-2}$

3º.- Refusem H_0 si $\bar{X}-\bar{Y}=d$ és SUFICIENTMENT DIFERENT de $\mu_{x0}-\mu_{y0}=\delta_0$

3.1.- Preseleccionem α

3.2.- Obtenim la regió crítica BILATERAL

3.3.- Refusem H0 si $\begin{cases} P\left(d \geq t_{\alpha/2,\, n_x+n_y-2}\right) = \alpha/2 \\ P\left(d \leq -t_{\alpha/2,\, n_x+n_y-2}\right) = \alpha/2 \end{cases}$

CAS 2 $\begin{cases} H_0 : \mu_x - \mu_y = \delta_0 \\ H_1 : \mu_x - \mu_y < \delta_0 \end{cases}$

$$\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_{x0}-\mu_{y0}\right)}{S_*\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}} \leq -t_{\alpha, n_x+n_y-2}$$

$\Rightarrow$ Refusem H_0

CAS 3 $\begin{cases} H_0 : \mu_x - \mu_y = \delta_0 \\ H_1 : \mu_x - \mu_y > \delta_0 \end{cases}$

$$\frac{\left(\bar{X}-\bar{Y}\right)-\left(\mu_{x0}-\mu_{y0}\right)}{S_*\sqrt{\frac{1}{n_x}+\frac{1}{n_y}}} \geq t_{\alpha, n_x+n_y-2}$$

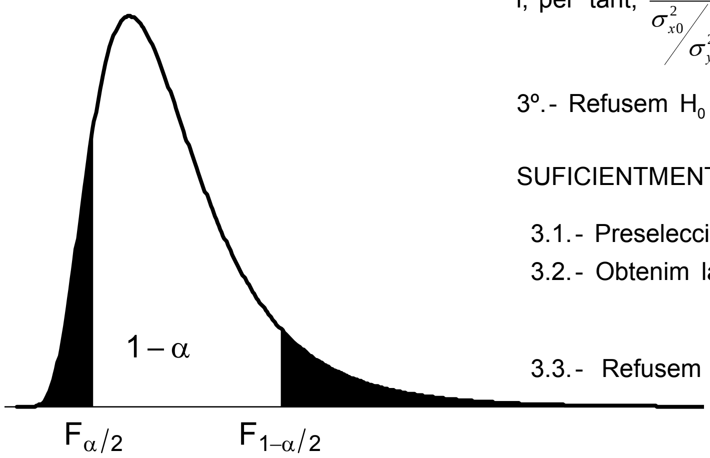
$\Rightarrow$ Refusem H_0

ECET 3.11

Tema 3. Contrastos d’hipòtesis paramètriques

SOBRE LA VARIÀNCIA POBLACIONAL (2 mostra)

SUPÒSIT 2: Sigui $X_1, X_2, \ldots, X_{nx}$ i $Y_1, Y_2, \ldots, Y_{ny}$ dues mostres aletòries de dos distribucions normals independents amb mitjanes μ_x i μ_y DESCONEGUDES i variàncies DESCONEGUDES σ_x^2 i σ_y^2.



CAS 1 $\begin{cases} H_0 : \sigma_x^2 / \sigma_y^2 = \varepsilon_0 \\ H_1 : \sigma_x^2 / \sigma_y^2 \neq \varepsilon_0 \end{cases}$

1º.- El millor estadístic per a contrastar σ^2 és S_*^2

2º.- Sabem que sota H_0 $\frac{\left(n_x-1\right)S_{x*}^2}{\sigma_{x0}^2} \approx \chi^2_{n_x-1}$ i $\frac{\left(n_y-1\right)S_{y*}^2}{\sigma_{y0}^2} \approx \chi^2_{n_y-1}$

i, per tant, $$\frac{S_x^2 / S_y^2}{\sigma_{x0}^2 / \sigma_{y0}^2} \approx F_{n_x-1, n_y-1}$$

3º.- Refusem H_0 si $S_x^2 / S_y^2 = e$ és

SUFICIENTMENT DIFERENT de $\sigma_{x0}^2 / \sigma_{y0}^2 = \varepsilon_0$

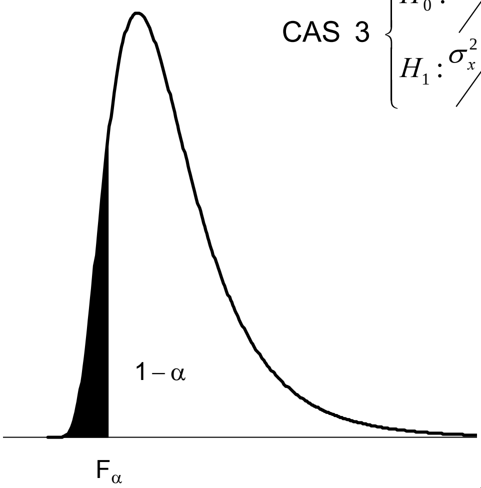
3.1.- Preseleccionem α

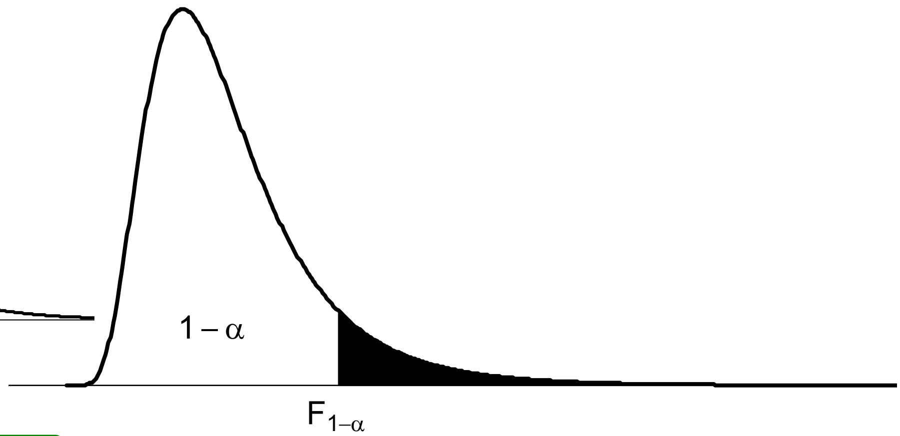
3.2.- Obtenim la regió crítica BILATERAL

3.3.- Refusem H_0 si $\begin{cases} P\left(\frac{e}{\varepsilon_0} \geq F_{1-\alpha/2,\, n_x-1, n_y-1}\right) = \alpha/2 \\ P\left(\frac{e}{\varepsilon_0} \leq F_{\alpha/2,\, n_x-1, n_y-1}\right) = \alpha/2 \end{cases}$

CAS 2 $\begin{cases} H_0 : \sigma_x^2 / \sigma_y^2 = \varepsilon_0 \\ H_1 : \sigma_x^2 / \sigma_y^2 < \varepsilon_0 \end{cases} \Rightarrow IDEM \Rightarrow \dfrac{S_x^2 / S_y^2}{\sigma_{x0}^2 / \sigma_{y0}^2} \le F_{\alpha, n_x - 1, n_y - 1} \Rightarrow$ Refusem H_0

CAS 3 $\begin{cases} H_0 : \sigma_x^2 / \sigma_y^2 = \varepsilon_0 \\ H_1 : \sigma_x^2 / \sigma_y^2 > \varepsilon_0 \end{cases} \Rightarrow IDEM \Rightarrow \dfrac{S_x^2 / S_y^2}{\sigma_{x0}^2 / \sigma_{y0}^2} \ge F_{1-\alpha, n_x - 1, n_y - 1} \Rightarrow$ Refusem H_0





ECET 3.12

SOBRE LA PROPORCIÓ POBLACIONAL (2 mostres)

Suposem que es pretén comparar dues proporcions poblacionals, a partir de dues mostres aleatòries independents.

CAS 1 $\begin{cases} H_0 : \Pi_x - \Pi_y = \gamma_0 \\ H_1 : \Pi_x - \Pi_y \ne \gamma_0 \end{cases}$

$$p_x - p_y \approx N\left(\Pi_{x0} - \Pi_{y0}, \sqrt{\frac{\Pi_{x0}(1-\Pi_{x0})}{n_x} + \frac{\Pi_{y0}(1-\Pi_{y0})}{n_y}} \right)$$

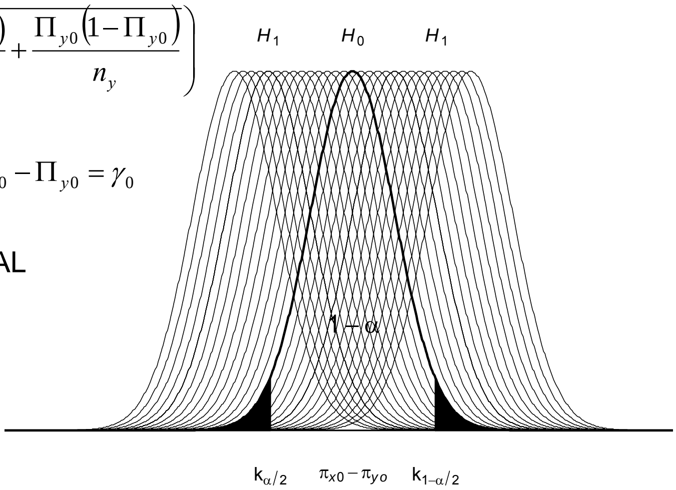
1º./2º. $p_x - p_y \approx N\left(\Pi_{x0} - \Pi_{y0}, \sqrt{\dfrac{\Pi_{x0}(1-\Pi_{x0})}{n_x} + \dfrac{\Pi_{y0}(1-\Pi_{y0})}{n_y}} \right)$

3º.- Refusem H_0 si $p_x - p_y = g$ és SUFICIENTMENT DIFERENT de $\Pi_{x0} - \Pi_{y0} = \gamma_0$

3.1.- Preseleccionem α

3.2.- Obtenim la regió crítica BILATERAL

3.3.- Refusem H_0 si $\begin{cases} P\left(g \ge k_{1-\alpha/2} \right) = \alpha/2 \\ P\left(g \le k_{\alpha/2} \right) = \alpha/2 \end{cases}$



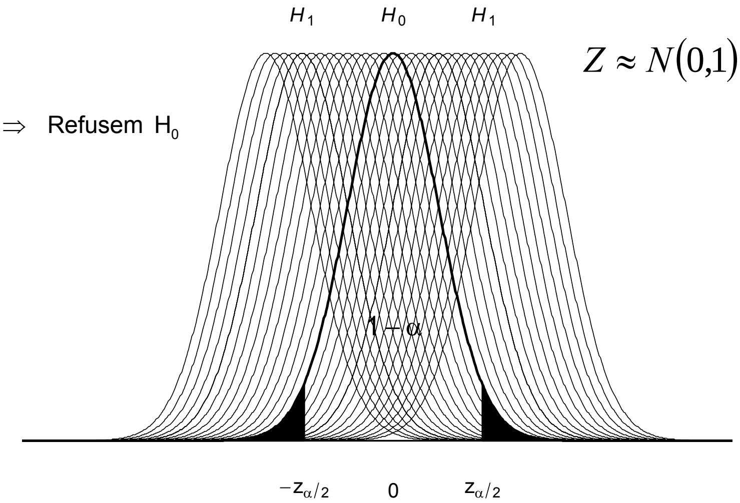
Tema 3. Contrastos d'hipòtesis paramètriques

$$P\left(p_x - p_y \geq k_{1-\alpha/2}\right) = P\left(\frac{(p_x - p) - (\Pi_{x0} - \Pi_{y0})}{\sqrt{\frac{\Pi_{x0}(1-\Pi_{x0})}{n_x} + \frac{\Pi_{y0}(1-\Pi_{y0})}{n_y}}} \geq \frac{k_{1-\alpha/2} - (\Pi_{x0} - \Pi_{y0})}{\sqrt{\frac{\Pi_{x0}(1-\Pi_{x0})}{n_x} + \frac{\Pi_{y0}(1-\Pi_{y0})}{n_y}}}\right) = P\left(Z \geq z_{\alpha/2}\right) = \alpha/2$$

$$P\left(p \leq k_{\alpha/2}\right) = P\left(\frac{(p_x - p) - (\Pi_{x0} - \Pi_{y0})}{\sqrt{\frac{\Pi_{x0}(1-\Pi_{x0})}{n_x} + \frac{\Pi_{y0}(1-\Pi_{y0})}{n_y}}} \leq \frac{k_{\alpha/2} - (\Pi_{x0} - \Pi_{y0})}{\sqrt{\frac{\Pi_{x0}(1-\Pi_{x0})}{n_x} + \frac{\Pi_{y0}(1-\Pi_{y0})}{n_y}}}\right) = P\left(Z \leq -z_{\alpha/2}\right) = \alpha/2$$

i, per tant només amb

$$\left|\frac{(p_x - p) - (\Pi_{x0} - \Pi_{y0})}{\sqrt{\frac{\Pi_{x0}(1-\Pi_{x0})}{n_x} + \frac{\Pi_{y0}(1-\Pi_{y0})}{n_y}}}\right| \geq z_{\alpha/2} \quad \Rightarrow \quad \text{Refusem } H_0$$



Tema 3. Contrastos d'hipòtesis paramètriques

CAS 2 $\begin{cases} H_0 : \Pi_x - \Pi_y = \gamma_0 \\ H_1 : \Pi_x - \Pi_y < \gamma_0 \end{cases}$

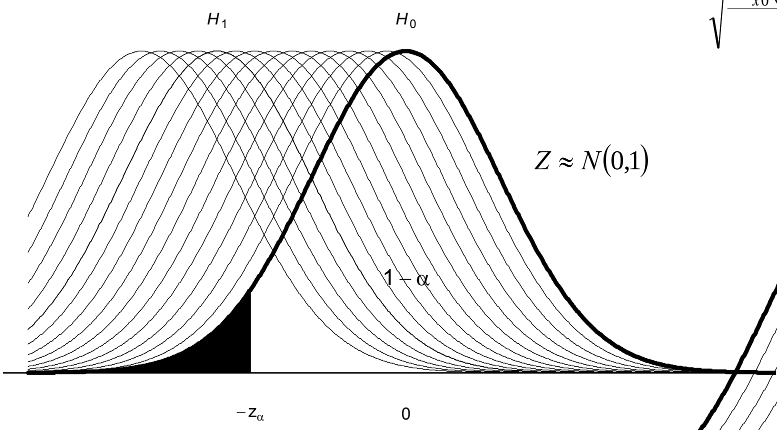
IDEM

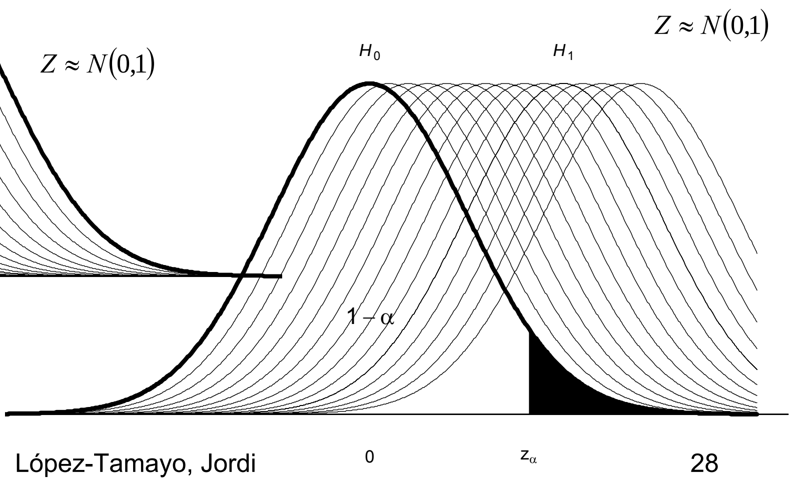
$$\frac{(p_x - p) - (\Pi_{x0} - \Pi_{y0})}{\sqrt{\frac{\Pi_{x0}(1-\Pi_{x0})}{n_x} + \frac{\Pi_{y0}(1-\Pi_{y0})}{n_y}}} \leq -Z_\alpha \quad \text{Refusem } H_0$$

CAS 2 $\begin{cases} H_0 : \Pi_x - \Pi_y = \gamma_0 \\ H_1 : \Pi_x - \Pi_y > \gamma_0 \end{cases}$

IDEM

$$\frac{(p_x - p) - (\Pi_{x0} - \Pi_{y0})}{\sqrt{\frac{\Pi_{x0}(1-\Pi_{x0})}{n_x} + \frac{\Pi_{y0}(1-\Pi_{y0})}{n_y}}} \geq Z_\alpha \quad \text{Refusem } H_0$$





ECET 3.13

Tema 3. Contrastos d'hipòtesis paramètriques

3.5. ANOVA. Anàlisi de la Variància UN FACTOR.

L'anàlisi de la variància permet contrastar si les mitjanes de tres o més poblacions són iguals. És a dir, si $\mu_1=\mu_2=\ldots\ldots\ldots\ldots=\mu_n$. Tot i ser una anàlisi sobre la mitjana es denomina ANOVA perquè per a realitzar el contrast s'utilitzen dos estimadors de la variància poblacional. Variància intragrups i variància entregrups. Suposem que disposem de la següent informació:

Població	Població	p_1	p_2	$\cdots$	p_j	$\cdots$	p_k
	Mitjana	μ_1	μ_2	$\cdots$	μ_j	$\cdots$	μ_k
	Variància	σ_1^2	σ_2^2	$\cdots$	σ_j^2	$\cdots$	σ_k^2
Mostra	Elements Mostra	X_{11}	X_{12}	$\cdots$	X_{1j}	$\cdots$	X_{1k}
		X_{21}	X_{22}	$\cdots$	X_{2j}	$\cdots$	X_{2k}
		$\vdots$	$\vdots$	$\cdots$	$\vdots$	$\cdots$	$\vdots$
		$X_{n_1 1}$	$X_{n_2 2}$	$\cdots$	$X_{n_j j}$	$\cdots$	$X_{n_k k}$
	Mida	n_1	n_2	$\cdots$	n_j	$\cdots$	n_k
	Mitjana	$\bar{X}_1$	$\bar{X}_2$	$\cdots$	$\bar{X}_j$	$\cdots$	$\bar{X}_k$
	Variància*	S_{*1}^2	S_{*2}^2	$\cdots$	S_{*j}^2	$\cdots$	S_{*k}^2

Tema 3. Contrastos d'hipòtesis paramètriques

On de CADA MOSTRA :

$X_{ij} \rightarrow$ Observació i-éssima de la mostra j-éssima

$X_{n_i j} \rightarrow$ Última observació de la mostra j-éssima

$n_k \rightarrow$ Mida mostral de la mostra j-éssima

$\bar{X}_j \rightarrow$ Mitjana mostral de la mostra j-éssima

$$\frac{\sum_{i=1}^{n_j} X_{ij}}{n_j}$$

$S_{*j}^2 \rightarrow$ Variància mostral de la mostra j-éssima

$$\frac{\sum_{i=1}^{n_j}\left(X_{ij}-\bar{X}_j\right)^2}{n_j-1}$$

On de TOTES LES MOSTRES :

$\bar{X} \rightarrow$ Mitjana de totes les observacions de totes les mostres

$$\frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j} X_{ij}}{\sum_{j=1}^{k} n_j}$$

$S_*^2 \rightarrow$ Variància mostral de tots els elements

$$\frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(X_{ij}-\bar{X}\right)^2}{\sum_{j=1}^{k} n_j-1}$$

CONTRAST :

$$\begin{cases} H_0: & \mu_1=\mu_2=\cdots=\mu_k=\mu \\ H_1: & H_0 \text{ no és certa} \end{cases}$$

Tema 3. Contrastos d'hipòtesis paramètriques

Restriccions per a poder fer el contrast

1. Les k mostres han de ser aleatòries e independents.
2. Han de ser poblacions normals.
3. Les variàncies de les poblacions han de ser iguals. És a dir: $\sigma_1^2=\sigma_2^2=\ldots=\sigma_j^2=\ldots=\sigma_k^2=\sigma$

El contrast es construeix basant-se en dos tipus de variabilitat:

- La variabilitat dintre de cada mostra (INTRA)
- La variabilitat entre les diferents mostres (ENTRE)

De la matriu de dades anterior es poden definir DOS EFECTES:

- La diferència de la mitjana de la població "j" respecte a la total $\quad v_j = \mu_j - \mu$
- La diferència de cada observació respecte a la mitjana poblacional de la població a la qual pertany. $\quad \varepsilon_{ij} = X_{ij} - \mu_j$

Aleshores es poden fer els següents canvis:

$$\begin{cases} \mu_j = v_j + \mu \\ \mu_j = X_{ij} - \varepsilon_{ij} \end{cases} \Rightarrow v_j + \mu = X_{ij} - \varepsilon_{ij} \quad \Rightarrow X_{ij} = v_j + \mu + \varepsilon_{ij}$$

És a dir. Cada observació és igual a la mitjana poblacional total més la desviació que pot existir de la mitjana poblacional de la població a la que pertany més la desviació que hi entre l'element i mitjana poblacional de la població a la qual pertany.

Tema 3. Contrastos d'hipòtesis paramètriques

És a dir: $\rightarrow \quad X_{ij} - \mu = (\mu_j - \mu) + (X_{ij} - \mu_j)$

Quan ens situarem sota la H_0? Quan tinguem SUFICIENT EVIDÈNCIA SOBRE $\mu_j - \mu = 0$

Per a determinar l'estadístic de prova substituïm pels corresponents estimadors:

$$X_{ij} - \overline{X} = (\overline{X}_j - \overline{X}) + (X_{ij} - \overline{X}_j)$$

Prenem quadrats a tots dos costats i sumem per a tot ij

$$\sum_{j=1}^{k}\sum_{i=1}^{n_i}(X_{ij} - \overline{X})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_i}(\overline{X}_j - \overline{X})^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_i}(X_{ij} - \overline{X}_j)^2 + \cancel{2\sum_{j=1}^{k}\sum_{i=1}^{n_i}(\overline{X}_j - \overline{X})(X_{ij} - \overline{X})}$$

En conseqüència:

$$\underbrace{\sum_{j=1}^{k}\sum_{i=1}^{n_i}(X_{ij} - \overline{X})^2}_{\substack{SQT \\ \text{Suma Quadrats Totals}}} = \underbrace{\sum_{j=1}^{k} n_j(\overline{X}_j - \overline{X})^2}_{\substack{SQE \\ \text{Suma Quadrats Entregrups}}} + \underbrace{\sum_{j=1}^{k}\sum_{i=1}^{n_i}(X_{ij} - \overline{X}_j)^2}_{\substack{SQI \\ \text{Suma Quadrats Intragrups}}}$$

Sabem que una suma dividida pel nombre de graus de llibertat dona lloc al quadrat mitjà. Així

$$QME = \frac{SQE}{k-1} \approx \chi^2_{k-1} \quad \text{i} \quad QMI = \frac{SQI}{n-k} \approx \chi^2_{n-k}$$

Tema 3. Contrastos d'hipòtesis paramètriques

Sota la hipòtesi nul·la $\mu_1 = \mu_2 = \cdots = \mu_k = \mu$ o el que és el mateix $v_j = \mu_j - \mu = 0$ per $\forall$j

Es pot demostrar que:

$$E(QME) = \sigma^2 + \frac{\sum_{j=1}^{k} n_j v_j^2}{k-1} \qquad \text{i} \qquad E(QMI) = \sigma^2$$

Per tant si H_0 és certa $v_j = 0$ i per tant $E(QME) = \sigma^2$

Aleshores, si podem demostrar que no existeixen diferències entre QME i QMI haurem de concloure que NO ES POT REFUSAR H_0

Per tant si $\frac{SQE/k-1}{SQI/n-k} \approx F_{k\text{-}1,\, n\text{-}k}$, refusarem H0 si $P\left(\frac{SQE/k-1}{SQI/n-k} \geq F_{1-\alpha,k-1,n-k}\right) = \alpha$

ECET 3.14