FIB

Departament d'Estadística
i Investigació Operativa
UNIVERSITAT POLITÈCNICA DE CATALUNYA

UPC

Disseny d'experiments

Bloc 5 – Probabilitat i Estadística

Novembre 2020

Índex

Inferència estadística. Guió

Guió de la part d'Estadística de PE:

- B4: Tècnica general de la inferència [estadística]
  - estimar un paràmetre (*Intervals de Confiança*)
  - refutar un paràmetre (*Proves d'Hipòtesis*)
- **B5: Aplicació (I): Avaluació de millores**
  - ***Disseny d'experiments:*** **comparació de dues poblacions.**
- B6: Aplicació (II): Predicció
  - *Previsió* d'una var. resposta, en funció d'una var. explicativa.

Introducció al disseny d'experiments

Distribució F de Fisher-Snedecor

- **Definició**: Siguin $X_1 \sim \chi^2_n \quad i \quad X_2 \sim \chi^2_m$. Llavors:

$$Y = \frac{X_1/n}{X_2/m} \sim F_{n,m}$$

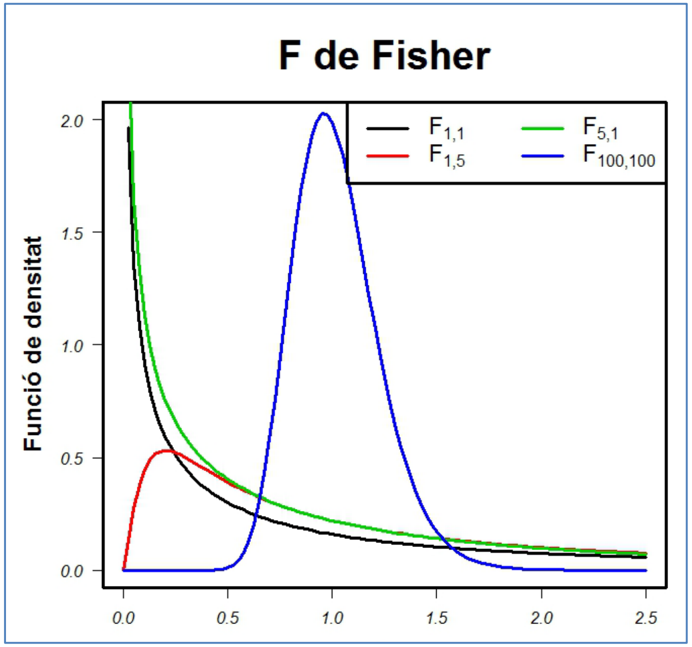
$1/Y \sim F_{m,n}$

- **Notació**: $F \sim F_{n,m}$
- **Paràmetres**: *n* (graus de llibertat numerador)

  m (graus de llibertat denominador)

- **Funció de probabilitat i distribució**:

  [F distribution at Wikipedia]



NOTA: La distribució F de Fisher, la farem servir per comparar variàncies de 2 poblacions

R: df, pf, qf

Script per veure que el quocient de χ^2 dividits pels g.ll és una F

```
M=500 ; n=5; m=7
samplechi2n = rchisq(M,n)
samplechi2m = rchisq(M,m)
F = (samplechi2n/n) / (samplechi2m/m)
hist(F, breaks="Scott", freq=FALSE)
curve(df(x, n, m),add=TRUE,col=2,lwd=2)
quantile(F, c(0.25, 0.50, 0.75))
qf(c(0.25, 0.50, 0.75), n, m)
```

Distribució F de Fisher-Snedecor. Càlcul de probabilitats

- Sigui $Y_1 \sim F_{1,\infty}$

  $P(Y_1 > 3.84) = 0.05$ [Taula F(0.95): 1ª columna, última fila]

- Sigui $Y_2 \sim F_{10,5}$

  $P(Y_2 > 10.05) = 0.01$ [Taula F(0.99): 10a columna, 5a fila]

  ```
  R: qf(0.99,10,5) = 10.05
  R: 1-pf(10.05,10,5) = 0.01
  ```

- Sigui $Y_3 \sim F_{5,10}$ (Recordeu: $Y_3 = 1/Y_2$)

  $P(Y_3 > 1/10.05) = P(1/Y_2 > 1/10.05) = P(Y_2 < 10.05) = 0.99$

  ```
  R: 1-pf(1/10.05,5,10) = 0.99
  R: pf(10.05,10,5) = 0.99
  ```

  [Cas general: Si $Y_1 \sim F_{n,m}$ i $Y_2 \sim F_{m,n}$ llavors $P(Y_1>y)=P(Y_2<1/y)$]

Tipus de variables

- **Objectiu**: estimar l'efecte (causal)
  - d'una **intervenció (X)**
  - en una **resposta (Y)**
  - donades unes **condicions (Z)**
- La resposta Y ha de mesurar el nostre objectiu [Ex: temps d'execució d'un algoritme]
- La intervenció X és el nostre potencial per canviar el futur [Ex: algoritme A o B]
- Les condicions Z 'predeterminen' el futur (resposta Y) [Ex: CPU emprada]
  - Haurien de ser idèntiques pels diferents valors de X
  - Veurem que baixen σ_Y^2 i permeten augmentar la potència
  - Veurem al B6 com utilitzar-les per anticipar Y

Observar versus assignar

- **Objectiu**: Observar els canvis a la distribució de **Y** quan **X** canvia [Ex: estudiar un canvi en un paràmetre de Y com la esperança **E(Y) =** $\boldsymbol{\mu_Y}$ si es passa de $\mathbf{X_A}$ a $\mathbf{X_B}$]

- **Observar**: Sobre les altres condicions Z no tenim control i no podem evitar que canviïn conjuntament amb X i, en conseqüència, no sabrem si els canvis a Y es deuen a X o a alguna altra Z. [Ex: nombre d'errors de sintaxi en un codi R segons el SO que cada alumne ha escollit lliurament. Els que empren Linux són millor programadors i faran menys errors → En Linux es fan menys errors?]

- **Assignar**: si assignem les X, podrem controlar les Z per que siguin independents de X, i no puguin ser una explicació de les variacions a la resposta Y. [Ex: nombre d'errors de sintaxi en un codi R segons el SO que cada alumne ha emprat segons assignació. En teoria, no s'observaran diferències]

Tècniques estadístiques

[Els problemes es resoldran d'acord amb les tècniques del B4]

- **IC**: permet estimar la magnitud de l'efecte de X [Ex: si comparem $\bar{y}_A$ i $\bar{y}_B$ podem trobar una estimació de la diferència de mitjanes poblacionals $\mu_A - \mu_B$]
  - L'IC és el resultat definitiu d'un estudi experimental: és directe i simple d'interpretar
- **PH**: permet rebutjar valors rellevants de l'efecte [Ex: si comparem $\bar{y}_A$ i $\bar{y}_B$ podem decidir si la diferència de mitjanes poblacionals $\mu_A - \mu_B$ és nul·la]
  - habitualment, la hipòtesi nul·la (H_0) equival a un "punt mort" o a l'estat actual (que es voldria rebutjar)
  - la hipòtesi alternativa seria un nou estat, superior a l'actual [Ex: H_0: $\mu_A = \mu_B$; H_1: $\mu_A > \mu_B$ o, mes neutre, H_1: $\mu_A \neq \mu_B$]
  - en funció de l'evidència disponible amb la mostra, una decisió o l'altre condueix a posar en marxa diferents accions

Atenció: Les decisions poden ser equivocades, cal balancejar el risc dels errors i el cost de les seves conseqüències (es veurà més endavant)

Tipus de dissenys

- **Segons la resposta**. Per comparar **2 opcions X** estudiarem proves segons la **resposta Y** sigui numèrica o dicotòmica:

Numèrica: **comparem mitjanes**	***Dicotòmica:*** **comparem proporcions**
Cas: Comparació de 2 mitjanes	Cas: Comparació de 2 proporcions
Pregunta: E(Y) depèn de la opció *X*?	Pregunta: π(Y) depèn de la opció *X*?
Hipòtesi formal: H_0: $\mu_A = \mu_B$	Hipòtesi formal: H_0: $\pi_A = \pi_B$
IC95% de l'efecte diferencial: $\mu_A - \mu_B$	IC95% de l'efecte diferencial: $\pi_A - \pi_B$
Ex: Comparar temps de 2 algoritmes	Ex: Comparar proporció d'aprovats en 2 cursos

- **Segons la recollida de dades**:
  - *Mostres independents:* cada cas és una **mesura independent** [Ex: Per comparar els temps de càrrega de 2 navegadors, els hi faig carregar 30 pàgines web diferents a cadascun]
  - *Mostres aparellades:* cada cas dóna lloc a dues mesures, **parells de mesures** [Ex: Per comparar els temps de càrrega de 2 navegadors, faig carregar les mateixes 30 pàgines web a cadascun]

Nota: Sempre que es pugui, s'ha de fer un disseny amb dades aparellades ja que és més eficient

Mostres Independents

Comparació de mitjanes ($\mu_1=\mu_2$). Mostres independents

- La prova formal que volem posar a prova és:

$$\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases} \Leftrightarrow \begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 \neq 0 \end{cases}$$

- Els indicadors (Esperança i Variància) de la diferència són:

$$E(\bar{y}_1 - \bar{y}_2) = E(\bar{y}_1) - E(\bar{y}_2) = \mu_1 - \mu_2$$

$$\mathrm{V}(\bar{y}_1 - \bar{y}_2) = V(\bar{y}_1) + V(\bar{y}_2) - 2Cov(\bar{y}_1, \bar{y}_2) \xrightarrow[indep.]{} \mathrm{V}(\bar{y}_1 - \bar{y}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

- Per tant, sota H_0 es compleix:

$$\hat{z} = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \xrightarrow[si\ \ H_0\ \text{és}\ \mu_1=\mu_2]{} \hat{z} = \frac{(\bar{y}_1 - \bar{y}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

[Premisses: $Y_1 \sim N(\mu_1, \sigma_1^2)$ i $Y_2 \sim N(\mu_2, \sigma_2^2)$ i m.a.s. independents]

Nota: Normalment, σ_1 i σ_2 seran **desconegudes**, però les podem suposar iguals (malgrat ser desconegudes). El que farem serà combinar les estimacions s_1 i s_2 donant lloc a una única estimació "pooled" (veure següent diapositiva)

Comparació de mitjanes ($\mu_1=\mu_2$). Mostres independents

- Si podem assumir igualtat de variàncies: $\sigma_1^2 = \sigma_2^2 = \sigma^2$, llavors, s_1^2 i s_2^2 seran estimadors del mateix paràmetre σ^2. En aquesta condició, una bona estimació de σ^2 s'obté ponderant s_1^2 i s_2^2 amb els seus respectius graus de llibertat donant lloc a la **s "pooled"**:

$$\hat{\sigma}^2 = s^2 = \frac{(n_1-1)\cdot s_1^2 + (n_2-1)\cdot s_2^2}{(n_1-1)+(n_2-1)} = \frac{\sum_{i=1}^{n_1}(y_{1i}-\bar{y}_1)^2 + \sum_{i=1}^{n_2}(y_{2i}-\bar{y}_2)^2}{n_1+n_2-2}$$

- Al substituir **σ** pel seu estimador ***s*** *pasem a tenir una t-Student:*

$$\hat{z} = \frac{(\bar{y}_1-\bar{y}_2)-(\mu_1-\mu_2)}{\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim N(0,1) \rightarrow \hat{t} = \frac{(\bar{y}_1-\bar{y}_2)-(\mu_1-\mu_2)}{s\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

Nota: Aquest estadístic representa la "senyal" proporcionada per la distància entre mitjanes relativa al "soroll" aleatori que porti aquesta senyal

Comparació de mitjanes ind. (μ_1=μ_2). Exemple

Els temps mitjans d'execució de dos programes provats en diferents bancs de dades independents (n_1=50 i n_2=100) han sigut: $\bar{y}_1 = 24 \; i \; \bar{y}_2 = 21$ amb s_1= 8 i s_2= 6. Suposem $\sigma^2{}_1 = \sigma^2{}_2$. Es desitja decidir quin programa es posa al mercat. Tenen rendiments diferents?

1. **Variables**: Y_1 i Y_2 de temps d'execució
2. **Estadístic**: $\hat{t} = \frac{(\bar{y}_1 - \bar{y}_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ amb $s^2 = \frac{(n_1-1)\cdot s_1^2+(n_2-1)\cdot s_2^2}{(n_1-1)+(n_2-1)}$
3. **Hipòtesi**: $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$ (bilateral)
4. **Distribució estadístic**: t_{148} **Premisses**: m.a.s. i Y_1 i Y_2 Normals amb $\sigma_1 = \sigma_2 = \sigma$
5. **Càlculs estadístic**: $s^2 = \frac{(50-1)\cdot s_1^2+(100-1)\cdot s_2^2}{(50-1)+(100-1)} = 45.27 \rightarrow \hat{t} = \frac{(24-21)}{6.72\sqrt{\frac{1}{50}+\frac{1}{100}}} = 2.58$
6. **P-valor** = $P(|t_{148}| > |2.58|) = 0.011$

[R: pt(-2.58,148)+(1-pt(2.58,148)] [(Taules ($t_{148} \approx Z$): 2·(1-0.9951)]

```
R: t.test(Y_A, Y_B, var.equal=TRUE)
```

Comparació de mitjanes ind. (μ_1=μ_2). Exemple (cont)

7. **Conclusió**:

   – *P-valor = 0.011 < 0.05 → Rebutgem H_0 (descartem que μ_1=μ_2) amb un p-valor de 0.011*

   – *$\hat{t}$ = 2.58 > 1.976 = $t_{148,0.975}$ → Rebutgem H_0 (descartem que μ_1=μ_2) amb risc α = 0.05*

   Conclusió pràctica:

   – *(PH): Aquestes dades (o més extremes) són poc probables (P=0.011) si els rendiments fossin iguals: Rebutgem H_0*

8. **Inferència amb Interval de Confiança**:

$$IC(\boldsymbol{\mu_1} - \boldsymbol{\mu_2}, 95\%) = (\bar{y}_1 - \bar{y}_2) \mp t_{n_1+n_2-2,1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s^2}{n_1} + \frac{s^2}{n_2}} = 3 \mp 1.976 \cdot 1.165$$

$$= [0.70\,, 5.30]$$

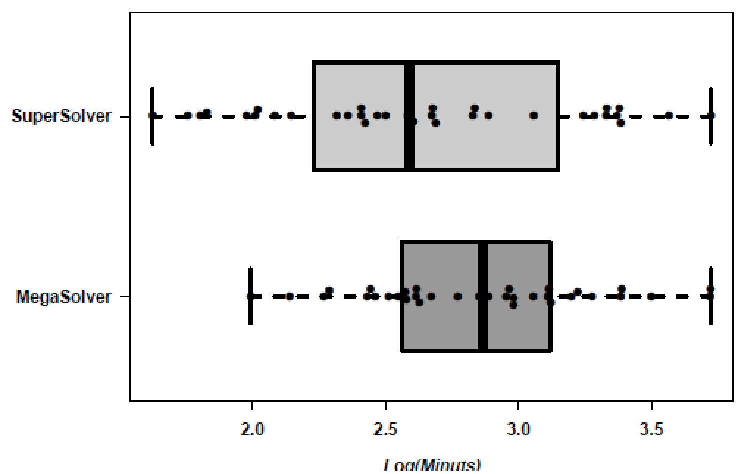
   – *IC95%: El programa "2" triga en mitjana entre 0.7 i 5.3 segons menys amb una confiança del 95%*

Comparació de mitjanes ind. (μ_1=μ_2). Exercici

Per comparar la velocitat amb la qual resolen dos servidors diferents, *SuperSolver* i *MegaSolver*, problemes d'optimització s'envia un total de 70 problemes de maximització diferents als dos servidors, 35 a cadascun. Pel fet que el temps que triguen els servidors per resoldre els problemes, és asimètrica cap a la dreta, treballem a continuació amb els logaritmes dels temps. Siguin X el logaritme del temps que triga el *SuperSolver* i Y el del *MegaSolver* .

Els valors descriptius a cada mostra són els següents i a més a més es mostra una representació gràfica:

	Mitjana	Mediana	Desv. est.	Mínim	Màxim
SuperSolver	2,63	2,59	0,57	1,63	3,72
MegaSolver	2,85	2,86	0,44	1,99	3,72



Comparació de mitjanes ind. ($\mu_1=\mu_2$). Exercici

1. **Variables**: Y_1 i Y_2 de temps d'execució
2. **Estadístic**: $\hat{t} = \frac{(\bar{y}_1-\bar{y}_2)}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$ amb $s^2 = \frac{(n_1-1)\cdot s_1^2+(n_2-1)\cdot s_2^2}{(n_1-1)+(n_2-1)}$
3. **Hipòtesi**: $\begin{cases} H_0: \mu_1 = \mu_2 \\ H_1: \mu_1 \neq \mu_2 \end{cases}$ (bilateral)
4. **Distribució estadístic**: t_{68} **Premisses**: m.a.s. i Y_1 i Y_2 Normals amb $\sigma_1 = \sigma_2 = \sigma$
5. **Càlculs estadístic**: $s^2 = \frac{(35-1)\cdot s_1^2+(35-1)\cdot s_2^2}{(35-1)+(35-1)} = 0.26 \rightarrow \hat{t} = \frac{(2.85-2.63)}{0.51\sqrt{\frac{1}{35}+\frac{1}{35}}} = 1.80$
6. **P-valor** = $P(|t_{68}| > |1.80|) = 0.076$
7. **IC95%**

$$(\mu_1-\mu_2, 95\%) = (\bar{y}_1-\bar{y}_2) \mp t_{n_1+n_2-2,1-\frac{\alpha}{2}} \cdot \sqrt{\frac{s^2}{n_1}+\frac{s^2}{n_2}} = 0.22 \mp 2.00 \cdot 0.51 \cdot \sqrt{\frac{1}{35}+\frac{1}{35}} =$$
$$[-0.02\, , 0.46]$$

Conclusió: No hi ha evidència per a dir que triguin diferent

Comparació de var ($\sigma_1^2 = \sigma_2^2$). Mostres independents

- La prova formal que volem posar a prova és:

$$\begin{cases} H_0: \sigma_1 = \sigma_2 \\ H_1: \sigma_1 \neq \sigma_2 \end{cases} \Leftrightarrow \begin{cases} H_0: \sigma_1/\sigma_2 = 1 \\ H_1: \sigma_1/\sigma_2 \neq 1 \end{cases}$$

- Sota H_0, tenim l'estadístic: amb la següent distribució

$\hat{F} = \frac{S_1^2}{S_2^2} \sim F_{n_1-1,n_2-1}$ [Premisses: m.a.s independents i $Y_1 \sim N(\mu_1, \sigma_1^2)$ i $Y_2 \sim N(\mu_2, \sigma_2^2)$]

- Es rebutjarà H_0 sempre que:

$\hat{F} < F_{n1-1,n2-1,\alpha/2}$ o bé que $\hat{F} > F_{n1-1,n2-1,1-\alpha/2}$

- Si posem la S més gran (S_M) al numerador, amb l'estadístic: $\hat{F} = \frac{S_M^2}{S_m^2} \sim F_{n_M-1,n_m-1} > 1$ només cal mirar la cua dreta a les taules. Llavors, es rebutjarà H_0 sempre que

$$\hat{F} > F_{n_M-1,n_m-1,1-\alpha/2}$$

- No obstant, si la prova és **unilateral**, tot el risc α s'acumula al costat adient (depenent del sentit de H_1 i de quin grup presenta la major variància mostral)

Nota tècnica: Pel càlcul del p-valor, si la prova és bilateral, sempre s'haurà d'agafar la cua més petita (esquerra o dreta) i multiplicar-la per 2

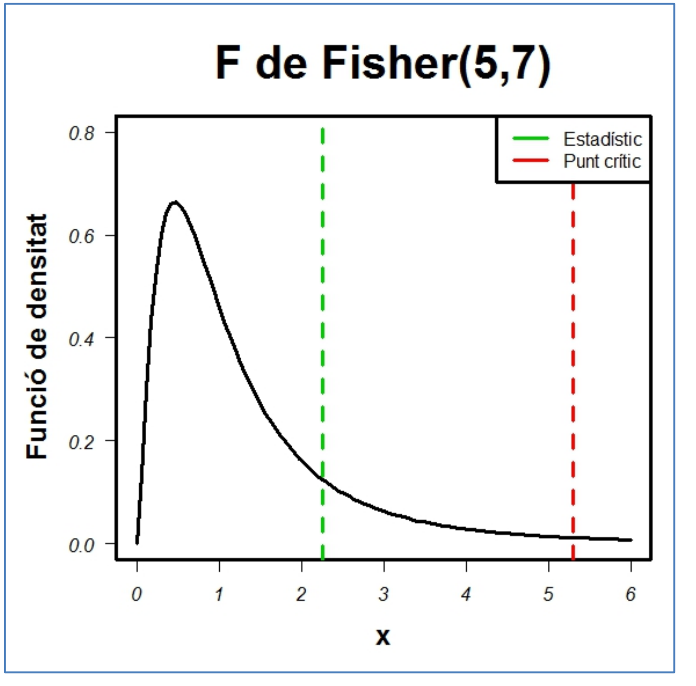
Comprar. de var ($\sigma_1^2 = \sigma_2^2$). Mostres independents. Ex.

Estem interessats en comparar la duració dels recanvis dels cartutxos de tinta de dos marques: "ORIGINAL, S.A." (**A**) y "YO_TAMBIÉN_LO_HAGO, S.L." (**B**). Hem comprovat que la marca B té una mitjana major de duració que la marca A. Però sospitem que la variabilitat pot ser diferent (provarem si **són iguals o no**). En dues mostres, hem obtingut els següents resultats:

A	$\bar{y}_A = 363$	S_A^2=64	$n_A = 8$
B	$\bar{y}_B = 407$	S_B^2=144	$n_B = 6$

1. **Variable**: Y_A i Y_B (duracions)
2. **Estadístic**: $S_{Major}^2 / S_{Menor}^2$
3. **Hipòtesis**: $H_0: \sigma_B^2 = \sigma_A^2$ vs $H_1: \sigma_B^2 \neq \sigma_A^2$
4. **Distribució estadístic**: $F_{5,7}$
5. **Càlculs**: $S_B^2 / S_A^2 = 144/64 = 2.25$
6. **P-valor**: $2 \cdot P\left(F_{5,7} > 2.25\right) = 0.32$
7. **Conclusió**: Com $2.25 < 5.29 = F_{5,7,0.975}$ i p-valor>0.05

llavors no podem rebutjar H_0. No hem trobat evidència per contradir que les variàncies poblacionals siguin iguals

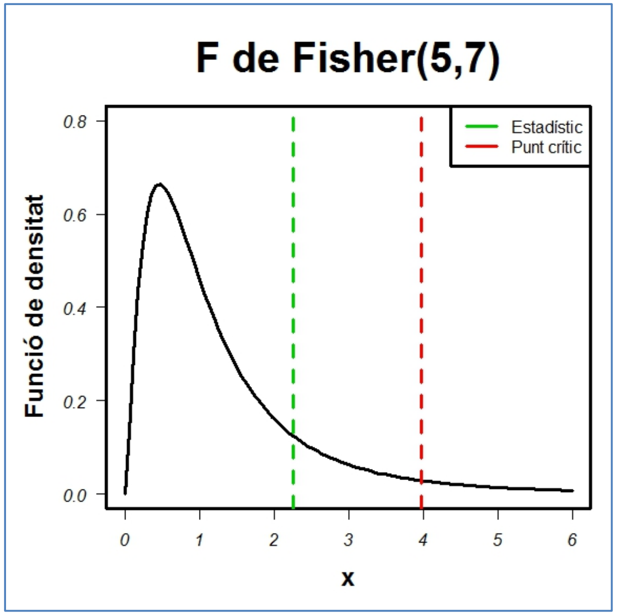


Compar. de var ($\sigma_1^2 = \sigma_2^2$). Mostres independents. Ex.

I si posem a prova si la variabilitat **és igual versus superior** en B?

R: `var.test(`Y_A`, `Y_B`)`

1. **Variable**: Y_A i Y_B (duracions)
2. **Estadístic**: S^2_{Major}/S^2_{Menor}
3. **Hipòtesis**: $H_0: \sigma_B^2 = \sigma_A^2$ vs $H_1: \sigma_B^2 > \sigma_A^2$
4. **Distribució estadístic sota H_0**: $F_{5,7}$
5. **Càlculs**: $S_B^2/S_A^2 = 144/64 = 2.25$
6. **P-valor:** $P(F > 2.25) = 0.16$
7. **Conclusió**: No podem rebutjar H_0



Ja que $2.25 < 3.97 = F_{5,7,0.95}$ [`R: qf(0.95,5,7)= 3.97`]

O bé *P*-valor = 0.16 > 0.05 = α [`R: 1-pf(2.25,5,7)= 0.16`]

No hem trobat evidència per contradir que les variàncies poblacionals siguin iguals.

Nota: si la prova és bilateral, posem al numerador el valor més gran segons les dades; però si és unilateral, segons H_1

Compar. de var ($\sigma_1^2 = \sigma_2^2$). Mostres independents. Exer.

Posa a prova la igualtat de variàncies en l'exercici dels optimitzadors

	Mitjana	Mediana	Desv. est.	Mínim	Màxim
SuperSolver	2,63	2,59	0,57	1,63	3,72
MegaSolver	2,85	2,86	0,44	1,99	3,72

1. **Variable**: Y_A i Y_B (temps)
2. **Estadístic**: S^2_{Major}/S^2_{Menor}
3. **Hipòtesis**: $H_0: \sigma_B^2 = \sigma_A^2$ vs $H_1: \sigma_B^2 \neq \sigma_A^2$
4. **Distribució estadístic sota H_0**:
5. **Càlculs**: $S_B^2/S_A^2 = 0.57^2/0.44^2 = 1.68$
6. **P-valor:** $2 \cdot P(F > 1.68) = 0.14$
7. **Conclusió**: No podem rebutjar H_0

Comparació de prop ($\pi_1 = \pi_2$). Mostres independents

- Ara desitgem posar a prova: $H_0 : \pi_1 = \pi_2$ vs $H_1 : \pi_1 \neq \pi_2$

[Ex: existeix la mateixa proporció d'alumnes que utilitzin PC (vs MAC) en la Facultat d'Informàtica i en l'Escola de Telecomunicacions?]

- Si el disseny de l'estudi, comporta **2 mostres**, una en la FIB i un altre en ETSETB, el plantejament és "**homogeneïtat**" entre 2 poblacions:

$$¿P(PC|FIB) = \pi_1 = \pi_2 = P(PC|ETSETB)?$$

- Mentre que si es tracta d' **1 única mostra**, en la que s'han recollit ambdues variables, el plantejament és "**independència**" de 2 variables aleatòries:

$$¿P(PC \cap FIB) = P(PC) \cdot P(FIB)?$$

- Tant l'objectiu d' "**homogeneïtat**" com el de "**independència**" es resolen de la mateixa manera. Per tant, no farem distinció entre un i altre.

Comparac. de prop ($\pi_1 = \pi_2$). Mostres independents. Ex

- L'exemple dels PC's podem presentar-ho en forma de taula 2x2. Suposem que hem obtingut 2 m.a.s. de 100 casos en cada facultat [situarem el recompte o freqüència f_{ij} de casos observats en la fila "i", columna "j"]

freqüència observada: f_{ij}	FIB	Telecos	Total
Utilitzen PC	77	63	140
Utilitzen MAC	23	37	60
Total	100	100	N = 200

- Quines serien les freqüències que es podria esperar (e_{ij}) de ser certa H_0? (Si fos certa H_0, llavors el ús de PC o MAC seria independent de la facultat i passaria, per exemple, que $P(PC \cap FIB) = P(PC) \cdot P(FIB)$
- Per calcular els e_{ij}, primer estimem les distribucions marginals:

  P(FIB)= 100/200 ; P(Telecos)= 100/200 ; P(PC)= 140/200 ; P(MAC)= 60/200

- Llavors, es poden calcular les freqüències esperades (e_{ij}) sota H_0. P.ex:

  $\mathbf{e}_{FIB,PC} = N \cdot P(PC \cap FIB) = N \cdot P(PC) \cdot P(FIB) = 200 \cdot (140/200) \cdot (100/200) = 70$

Comparac. de prop ($\pi_1 = \pi_2$). Mostres independents. Ex

freqüència esperada: e_{ij}	FIB	Telecos	Total
Utilitzen PC	70	70	140
Utilitzen MAC	30	30	60
Total	100	100	N = 200

Noteu que:

(1) $e_{ij} = \frac{e_{i.} \cdot e_{.j}}{e_{..}} = \frac{(total\ de\ fila\ i) \cdot (total\ de\ columna\ j)}{(total\ de\ totals)}$

(2) calculat el primer (**e_{11}**) , en una taula 2x2, els altres s'obtenen per diferència (e_{12}= 140-70 = 70)

- Ara tenim el resultat empíric ($\boldsymbol{f_{ij}}$) i l'esperat sota H_0 ($\boldsymbol{e_{ij}}$) i podem estudiar les seves divergències a través de l'estadístic de Pearson (chi-quadrat):

$$\hat{X}^2 = \sum_{\forall ij} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

- **Sota H_0 → $\hat{X}^2 \sim \chi^2_{(I-1)(J-1)}$** on I,J són el número de files i columnes, respectivament.
- **Premissa**: $e_{ij} \geq 5$ $\forall$i,j
- En el nostre cas, I=J=2, i la distribució χ^2 de referència té 1 g.d.l ja que (I-1)·(J-1) = 1·1 = 1

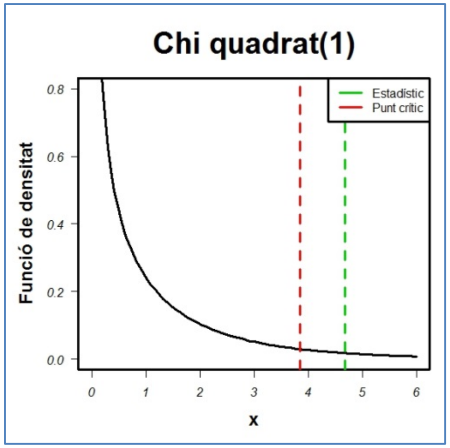
Comp. de prop ($\pi_1 = \pi_2$) en mostres indep. Exemple

$(f_{ij} - e_{ij})^2/e_{ij}$	FIB	Telecos	Total
Utilitzen PC	0.7	0.7	
Utilitzen MAC	1.633	1.633	
Total			4.667

1. **Variable**: PC o MAC
2. **Estadístic**: $\hat{X}^2 = \sum_{\forall ij} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$
3. **Hipòtesis H_0** : P(PC|FIB) = P(PC|ETSETB) ≡ $\pi_1 = \pi_2$ ≡ Independència escola i ordinador
4. **Distribució estadístic**: χ_1^2
5. **Càlculs**: $\hat{X}^2 = 4.667$
6. ***P*-valor**= P(χ_1^2 > 4.667) = 0.0308 (punt crític = $\chi_{1,0.95}^2$ = 3.841)
7. **Conclusió**: rebutgem H_0 (*P*-valor<0.05 o que 4.667>3.841)

   Aquestes dades (o més extremes) son poc probables si fos certa la independència

   Conclusió pràctica: Els Fibers prefereixen més el PC



R: chisq.test

NOTA: l'estadístic de Pearson ($\hat{X}^2$) tendirà a cero quan s'apropin f_{ij} i e_{ij} (és a dir, quan les dades millor reprodueixin H_0) per la qual cosa la regió crítica se situa només a la dreta, per l'efecte del "quadrat" (regió crítica **intrínsecament unilateral**).

Comp. de prop ($\pi_1 = \pi_2$) en mostres indep. Exercici

1. **Variable**: Aprovat/suspès vs nois/noies
2. **Estadístic**: $\hat{X}^2 = \sum_{\forall ij} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$
3. **Hipòtesi H_0**:

   P(A|noi) = P(A|noia) ≡ $\pi_1 = \pi_2$ ≡ Homogeneïtat d'aprovats entre nois i noies
4. **Distribució estadístic**: χ^2_1
5. **Càlculs**: $\hat{X}^2 = 0.601$
6. ***P*-valor**: $P(\chi^2_1 > 0.601) = 0.438$ [R:1-pchisq(0.601,1

   Punt crític: $\chi^2_{1,0.95} = 3.841$ [R: qchisq(0.95,1
7. **Conclusió**: NO rebutgem H_0 (*P*-valor>0.05 o 0.601<3.841)

   Les dades són 'plausibles' sota H_0 (P=0.438)

   Conclusió pràctica: No hi ha diferències d'aprovats per gènere

f_{ij}	noi	noia	Total
aprova	68	73	141
suspèn	32	27	59
Total	100	100	200

e_{ij}	noi	noia	Total
aprova	70.5	70.5	141
suspèn	29.5	29.5	59
Total	100	100	200

$(f_{ij} - e_{ij})^2/e_{ij}$	noi	noia	Total
aprova	0.089	0.089	
suspèn	0.212	0.212	
Total			0.601

Comp. de prop ($\pi_1 = \pi_2$). Mostres indep. i grans

- Si tenim mostres grans, podem utilitzar un estadístic alternatiu amb una altra distribució sota H_0. Sigui H_0: $\pi_1 = \pi_2$, llavors l'estadístic serà:

$$\hat{z} = \frac{(P_1 - P_2) - (\pi_1 - \pi_2)}{\sqrt{\frac{P(1-P)}{n_1} + \frac{P(1-P)}{n_2}}} \sim N(0,1) \quad \text{on} \quad P = \frac{n_1 \cdot P_1 + n_2 \cdot P_2}{n_1 + n_2} \; ; \; P_1 = \frac{X_1}{n_1} \; ; \; P_2 = \frac{X_2}{n_2}$$

[Aquest test, a diferència del de l'estadístic $\hat{X}^2$, sí que pot ser bilateral i permet calcular ICs]

Exemple: Volem comparar la proporció d'aprovats segons el gènere (cont.). Disposem de dues mostres de 100 individus (prou grans)

$P_1 = 0.68$ i $P_2 = 0.73$ → $P = 0.705$ (P representa la proporció comuna)

$SE^2 = P(1-P) \cdot (1/100 + 1/100) = 0.00416$

$\hat{\boldsymbol{z}} = (0.68 - 0.73)/\sqrt{0.00416} = -0.775$

R: `prop.test`

Com és bilateral → ***P-valor*** = P(|Z| > |$\hat{z}$|) = 2 · 0.219 = 0.438 [amb la χ^2 donava el mateix]

$\boldsymbol{IC(\pi_1 - \pi_2, 95\%)} = (0.68 - 0.73) \mp 1.96 \cdot \sqrt{(0.68 \cdot 0.32 + 0.73 \cdot 0.27)/100} = [-0.18, 0.08]$

NOTA: El resultat és el mateix que si ho haguèssim resolt amb X^2, amb idèntica conclusió: no hi ha evidència per dir que el gènere porta diferències en quant a la proporció d'aprovats.

Formulari de comparació de proporcions

Hipòtesi	Estadístic	Premisses	Distrib.(H_0)	Decisió α=0.05
H_0: $\pi_1 = \pi_2 = \pi$ H_1: $\pi_1 \neq \pi_2$	$\hat{z} = \frac{(P_1 - P_2)}{\sqrt{\frac{P(1-P)}{n_1} + \frac{P(1-P)}{n_2}}}$ $P = \frac{n_1 \cdot P_1 + n_2 \cdot P_2}{n_1 + n_2}$	n_1, n_2 grans m.a.s. indep	$\hat{z} \to N(0,1)$	Rebutjar si $\lvert\hat{z}\rvert > 1.96$
	$\hat{X}^2 = \sum_{\forall ij} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$	$e_{ij} \geq 5\ \forall\ ij$ m.a.s indep	$\hat{X}^2 \to \chi^2_{(I-1)(J-1)}$	Rebutjar si $\hat{X}^2 > \chi^2_{(I-1)(J-1),0.95}$
Les corresponents proves unilaterals es fan acumulant el risc α a un costat.				

Mostres Aparellades

Comp. de mitjanes ($\mu_1 = \mu_2$). Mostres aparellades

- Suposem que, per comparar 2 programes, A i B, les **dades o unitats** experimentals en que els provem **són les mateixes**.
- Ara, cada unitat ens proporciona informació sobre la diferència del rendiment de tots dos programes. Per això, definir una nova variable diferència **D** entre els dos rendiments que permet fer la **comparació "dins" de cada unitat**.
- Ara l'estudi de si els dos programes són iguals és pot fer a partir de la variable D:

$$H_0: \mu_A - \mu_B = 0 \ \rightarrow H_0: \mu_D = 0$$

- Per tant, l'estadístic és com el de la PH de μ per una mostra:

$$\hat{t} = \frac{\overline{D} - \mu_D}{S_D/\sqrt{n}} = (\text{si } \ \mu_D = 0) = \frac{\overline{D}}{S_D/\sqrt{n}} \sim t_{n-1}$$

  - $\overline{D}$ serà la mitjana de la variable diferencia D
  - S_D serà la desviació típica de D
  - μ_D serà el valor a contrastar: 0 per la igualtat de les 2 opcions

Comp. de mitjanes ($\mu_1 = \mu_2$). Mostres aparellades. Ex

En 6 bancs de dades s'ha obtingut els temps de 2 programes. Es desitja saber si B millora A; o decidir si al mercat canvien

							Mean	Variances	Var. "pooled"
A	23.05	39.06	21.72	24.47	28.56	27.58	**27.406**	**39.428**	**42.009**
B	20.91	37.21	19.29	19.95	25.32	24.07	**24.460**	**44.591**	

SOLUCIÓ INCORRECTA: (tractar com a dades de mostres independents H_0: $\mu_1=\mu_2$ bilateral):

$$\hat{t} = \frac{(\bar{y}_1 - \bar{y}_2)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{(27.406 - 24.460)}{6.48\sqrt{\frac{1}{6} + \frac{1}{6}}} = 0.787$$

***P*-valor** = $P(|t_{10}| > 0.79) = 0.4494$ [R: pt(-0.79,10)+(1-pt(0.79,10)] [Taules(t_{10}): 2*(1-0.9931)]

Punt crític: $t_{10,0.975} = 2.228$ [R: qt(0.975,10)] [Taules(t_{10}): 2.228]

Conclusió: donat *P-valor*=0.4494 > 0.05 o bé, donat t=0.787 < 2.228, res s'oposa a acceptar H_0 (triguen el mateix). Sota H_0 aquest resultat es 'esperable'

Comp. de mitjanes ($\mu_1 = \mu_2$). Mostres aparellades. Ex

SOLUCIÓ CORRECTA: (com a dades de mostres aparellades). El mètode per dades dependents comença per calcular la diferencia A-B per cada parella:

							Mean	Variances
D = A-B	2.13	1.85	2.43	4.51	3.24	3.51	**2.946**	**0.996**

1. **Variable**: D diferència de temps
2. **Estadístic**: $\hat{t} = \frac{(\bar{D}-\mu_D)}{S_D/\sqrt{n}} = \frac{\bar{D}}{S_D/\sqrt{n}}$
3. **Hipòtesis**: $\{H_0: \mu_D = 0 \; vs. H_1: \mu_D \neq 0$
4. **Distribució estadístic**: $\hat{t} \sim t_{n-1}$ **Premissa**: m.a. aparellades D~N
5. **Càlculs**: $\hat{t} = \frac{(\bar{D}-\mu_D)}{S_D/\sqrt{n}} = \frac{2{,}946-0}{0.998/\sqrt{6}} = 7.229$
6. **P-valor** = P(t_5> 7.229) = 0.0008 (punt crític $t_{0.975,5}$= 2.571)
7. **Conclusió**: rebutgem H_0 ja que *P*-valor < 0.05 o que 7.229 > 2.571)

   Aquestes dades no són probables sota la igualtat. Resultats com aquest son molt poc esperables sota H_0
8. $\mathbf{IC(\mu_D, 0.95)} = 2.946 \mp 2{,}571 \cdot 0.41 = [1.89{,}4.0]$

```
R: t.test(YA, YB, paired=TRUE)
```

Comp. de mitjanes ($\mu_1 = \mu_2$). Mostres aparellades. Ex

Comparació entre el resultat aparellat (correcte) i per mostres independents (incorrecte):

- Numerador: La **mitjana de les diferencies** (2.946) coincideix amb la **diferència de les mitjanes**. Per tant, el numerador és el mateix en les 2 opcions.

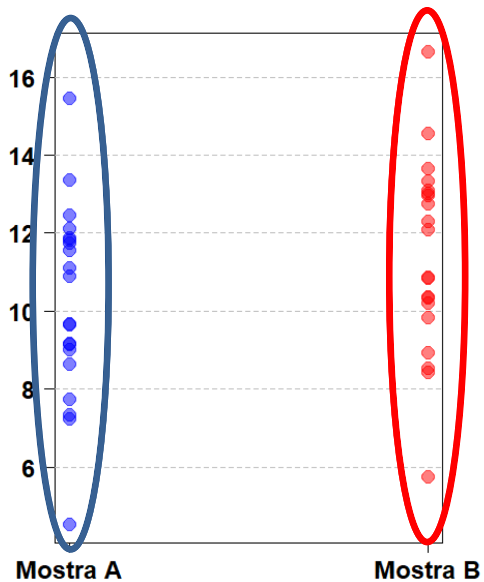
- Denominador: La **variància de les diferències** (0.996) és molt inferior a la "pooled" (42.009), donat que les diferències entre les unitats (s'espera que alguns bancs de dades siguin més 'durs' que altres) han desaparegut al comparar el rendiments dels 2 programes "dins" de cada banc de dades.

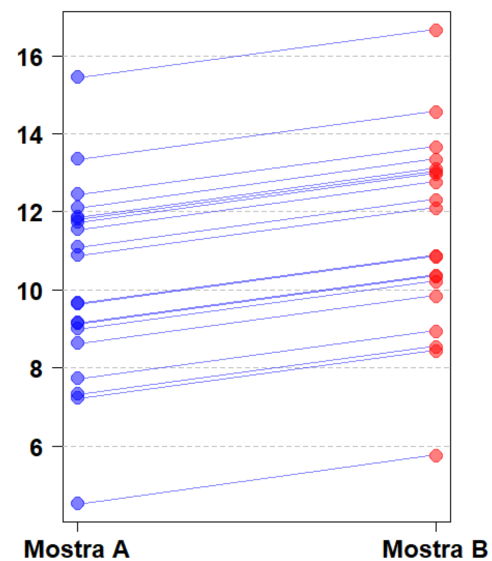
- Encara que el numerador (senyal) és el mateix, el **denominador** (soroll) **és molt inferior en la segona**. Això és el que provoca arribar a diferents conclusions en els 2 casos.

- Recorda: **el control de les condicions** (en aquest cas, la variabilitat dels bancs de dades) **augmenta l'eficiència** per trobar diferències.

Comp. de mitjanes ($\mu_1 = \mu_2$). Mostres aparellades

- És molt important no fer un anàlisi de dades aparellades com a dades independents
- Aparentment, en el gràfic de l'esquerra (independents) no veiem que pugui existir diferències en mitjana entre les dues poblacions
- En el gràfic de la dreta (aparellat) es veu clarament que la mitjana és superior en la mostra B.



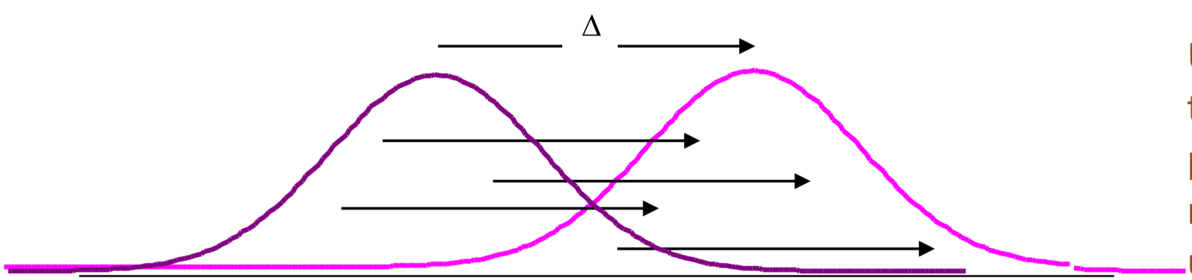


- Per aprendre més sobre mostres aparellades consulta aquesta app.

Premisses

Efecte additiu constant i Normalitat de D

- Hi ha 2 premisses comprovables: **efecte additiu constant** i **Normalitat de D**
- En la prova de comparació de mitjanes (tant aparellades com independents) s'assumeix un efecte additiu (aproximadament) constant.
- En mostres independents, no podem avaluar aquesta premissa però si en dades aparellades, ja que per a cada unitat, tenim el valor de la resposta de cada observació en les 2 mostres.
- Si l'efecte és lineal, **additiu**, cada cas (individu) té el mateix desplaçament (Δ)



Un algoritme que redueix en 2 segons el temps d'execució té un efecte **additiu**; però un algorisme que redueix a la meitat els temps d'execució no ho té (té un efecte **multiplicatiu**)

Premisses
Efecte additiu constant i Normalitat de D

- Un **efecte additiu constant** implica:
  - Idèntica forma de la distribució, i per tant: $\sigma_1^2 = \sigma_2^2$ [“***Homoscedasticitat***”]
  - Diferència que pot resumir-se en que $(\bar{y}_1 - \bar{y}_2)$ estima l‘efecte en **cada individu**
- Si s'observa un efecte multiplicatiu (en comptes d'additiu), es pot solucionar traient logaritmes de la variable resposta i fent l'anàlisi sobre la diferencia de logs:

$$\mathbf{D = \log(y_a) - \log(y_b) = \log(y_a/y_b)}$$

- Per interpretabilitat, pot interessar més conèixer un IC95% pel rati de les respostes, que no pas pel seu logaritme:
  - Es pot calcular el IC95% pel quocient fent l'exponencial (operació contraria al logaritme) de l'interval obtingut pels logaritmes.
  - NOTA: Aquest interval obtingut al desfer la transformació ja no representa un IC95% per la mitjana μ sinó per a una mitjana geomètrica. Infra-estimen la veritable mitjana poblacional.
- En les mostres aparellades, també es requereix la **premissa de Normalitat** de la variable diferència ($D = y_a - y_b$)

Premisses
Efecte additiu constant i Normalitat de D

- El gràfic de Bland-Altman (BA) permet avaluar la **premissa d'efecte additiu** constant representant les diferencies de les respostes per cada individu en funció de les seves mitjanes.
- Siguin Y1 i Y2, les respostes obtingudes en els mateixos casos (dades aparellades) el següent codi en R permet fer l'anàlisi.

R – Gràfic de BA per comprovar efecte additiu

```
library(PairedData)
Y1 <- datos$y1
Y2 <- datos$y2
p <- paired(Y1,Y2)
plot(p,type='BA')
```

R – QQnorm per comprovar premissa de Normalitat

```
qqnorm(Y1-Y2)
qqline(Y1-Y2,col=2)
```

R – Anàlisi Numèrica:

```
t.test(Y1,Y2,paired=TRUE,var.equal=TRUE)
```

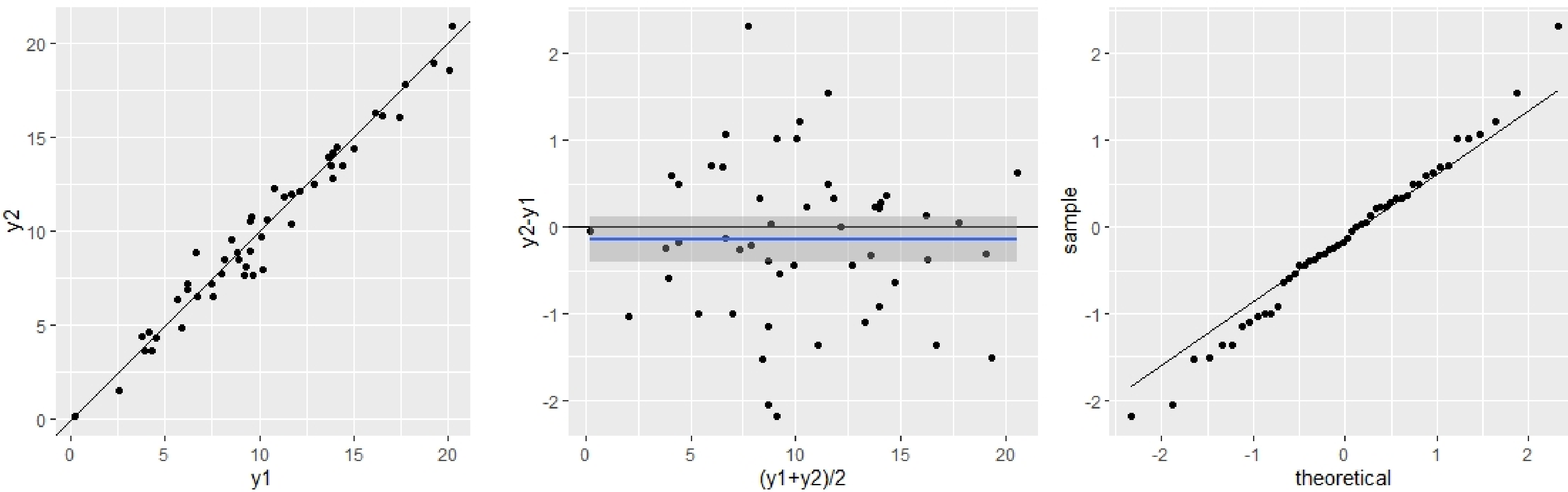
- **Veurem 3 situacions especials:**

  Cas 1: sense efecte lineal (additiu)

  Cas 2: amb efecte multiplicatiu

  Cas 3: amb efecte multiplicatiu i transformació logarítmica: `(lnY1, lnY2)`

Anàlisi de dades aparellades. No efecte



La dif. de mitjanes puntual estimada és -0.14

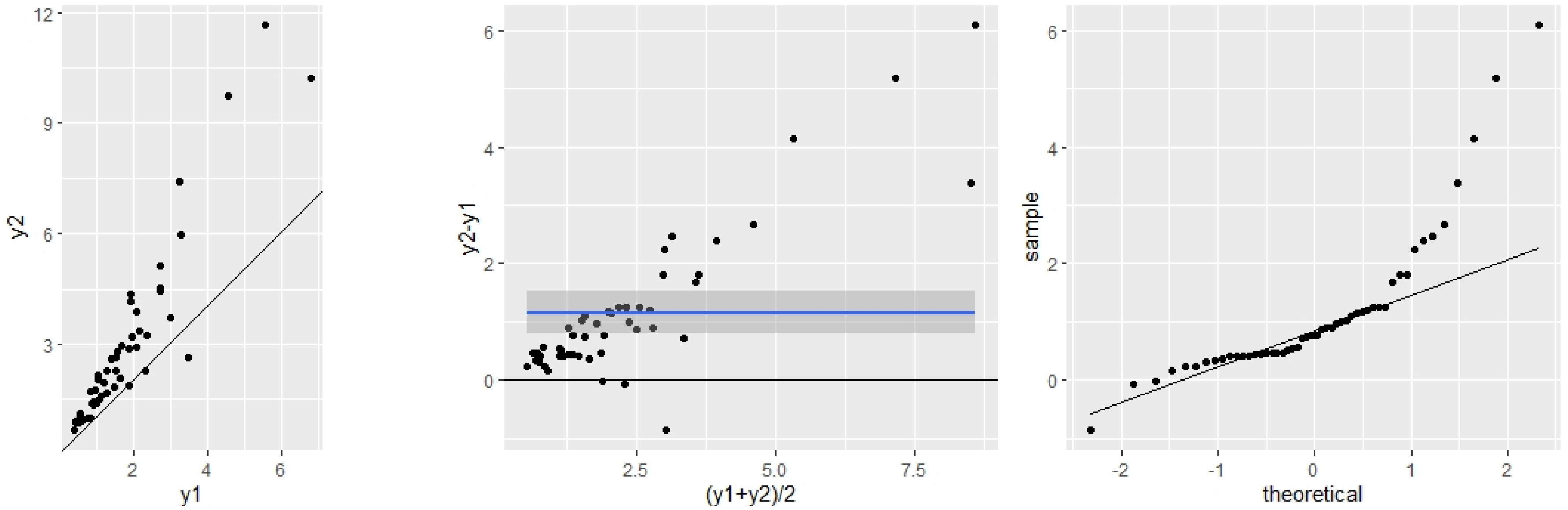
La dif. de mitjanes per interval: $IC_{95\%}(\mu_{Y2}-\mu_{Y1})$ = [-1.70 a 1.99]

És a dir, $Y_2 = Y_1$ + [-1.70 a 1.99]

Per tant, **no hi ha evidència de què ambdues mitjanes siguin diferents**.

Es pot assumir Normalitat de la variable diferència ja que tots els quantils observats s'ajusten força bé als quantils teòrics de la Normal.

Anàlisi de dades aparellades. Efecte multiplicatiu

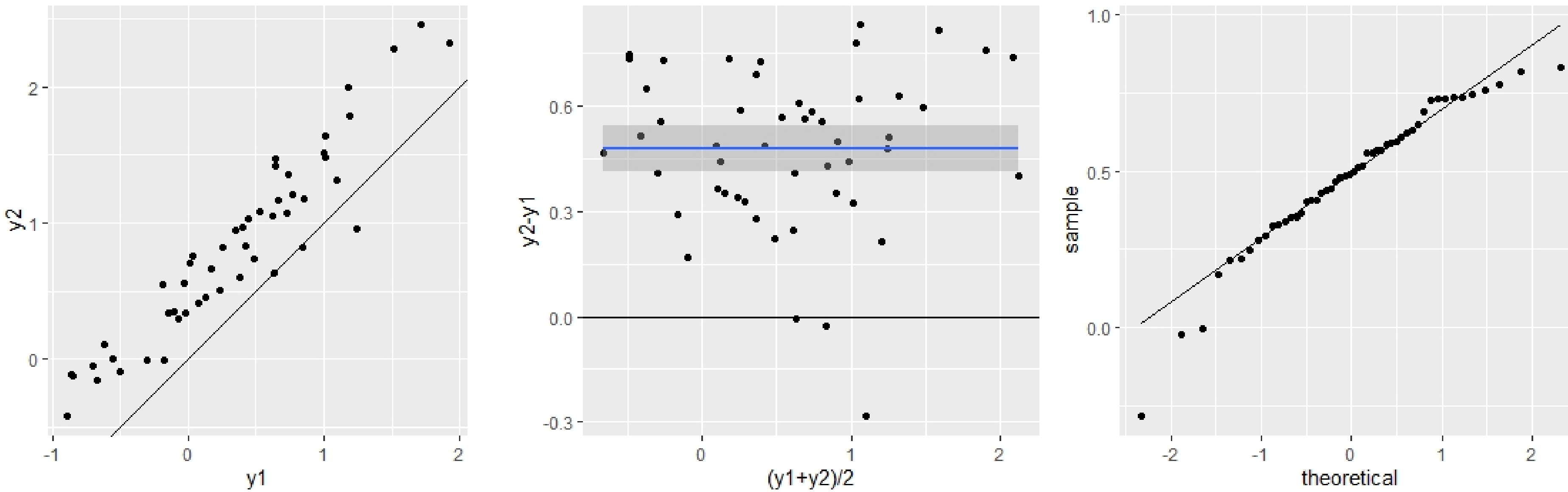


La diferència de mitjanes estimada és 1.16 amb un $IC_{95\%}(\mu_{Y2}-\mu_{Y1})$ = [0.38 a 1.92], però aquest valor no ens informa bé, ja que **l'efecte no és constant**. **Per valors grans, l'efecte és més gran** i té mes variabilitat. Si alguna transformació sobre les variables soluciona aquests problemes, la interpretació pot ser més fàcil.

La distribució de les diferències NO és Normal

Provarem solucionar-ho fent la transformació logarítmica (natural).

Anàlisi de dades aparellades. Treure logaritmes



La diferència mitjana estimada dels logaritmes és 0.48 amb un $IC_{95\%}(\mu_{Y2}-\mu_{Y1})$ de 0.21 a 0.75. Si Y1,Y2 són les variables originals i Y1',Y2' són les log-transformades:

$$\left.\begin{aligned} Y2' &= \ln(Y2) \\ Y1' &= \ln(Y1) \end{aligned}\right\} \rightarrow Y2' = Y1' + 0.48 \rightarrow \ln(Y2) = \ln(Y1) + 0.48 \rightarrow Y2 = e^{\ln(Y1)+0.48} \rightarrow Y2 = e^{\ln(Y1)} \cdot e^{0.48} = 1.62 \cdot Y1$$

Interpretació: Y2 és en mitjana 1.62 ($IC_{95\%}$ de 1.23 a 2.12) vegades més gran que Y1. **Nota**: al desfer la transformació logarítmica, les estimacions obtingudes infra-estimen el valor real.

Es pot assumir Normalitat de la variable diferència de logaritmes (= logaritme del quocient) ja que tots els quantils observats s'ajusten prou bé als quantils teòrics de la Normal.

Formulari. Proves de μ i σ en 2 mostres

Paràmetre	Hipòtesis	Estadístic	Premisses	Distrib. sota H_0	Decisió (Risc α)
μ	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$\hat{z} = \dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$	[Y_1 Y_2 $\rightarrow$N ò n_1 i n_2 $\geq$30] m.a.s ind. σ_1 , σ_2 coneg	$\hat{z} \rightarrow N(0,1)$	Rebutjar si $\|\hat{z}\| > Z_{1-\alpha/2}$ ($\|\hat{z}\|$>1.96 amb α=5%)
μ	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$\hat{t} = \dfrac{\bar{y}_1 - \bar{y}_2}{S\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}}$ $s^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$	Y_1 , Y_2 $\rightarrow$ N $\sigma_1 = \sigma_2$ m.a.s indep.	$\hat{t} \rightarrow t_{n_1+n_2-2}$	Rebutjar si $\|\hat{t}\| > t_{n1+n2-2,\ 1-\alpha/2}$ ($\|\hat{t}\|$> $t_{n1+n2-2,0.975}$ amb α=5%)
μ	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$\hat{z} = \dfrac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}}}$	n_1 , n_2 $\geq$ 100 m.a.s indep	$\hat{z} \rightarrow N(0,1)$	Rebutjar si $\|\hat{z}\| > Z_{1-\alpha/2}$
μ	$H_0 : \mu_1 = \mu_2$ $H_1 : \mu_1 \neq \mu_2$	$\hat{t} = \dfrac{\bar{D} - \mu_0}{S_D\sqrt{\dfrac{1}{n_D}}}$	D $\rightarrow$ N m.a aparellada	$\hat{t} \rightarrow t_{n-1}$	Rebutjar si $\|\hat{t}\| > t_{n-1,\ 1-\alpha/2}$
σ	$H_0 : \sigma^2_1 = \sigma^2_2$ $H_1 : \sigma^2_1 \neq \sigma^2_2$	$\hat{F} = \dfrac{S_1^2}{S_2^2} \quad (S_1^2 \geq S_2^2)$	Y_1 , Y_2 $\rightarrow$ N m.a.s indep	$\hat{F} \rightarrow F_{n_1-1,n_2-1}$	Rebutjar si $\hat{F} > F_{n1-1,n2-1,\ 1-\alpha/2}$

Les corresponents proves unilaterals es fan acumulant el risc α a un costat

Tipus d'errors

Tipus d'errors en una prova d'hipòtesi. Tipus I

- Als exemples anteriors, hem expressat les conclusions com "rebutgem H_0" o "no rebutgem H_0".
- Si l'objectiu és prendre una decisió, un criteri simple seria definir a priori un llindar α per sota del qual el P-valor és vist com "petit".
- Ara, si repetim la decisió 'n' vegades, com en un repetit procés de control de qualitat, α ens donarà la freqüència d'errors determinada:

En un 100α % dels casos que rebutgem H_0, aquesta és certa.

- **Error de tipus I.** Quan utilitzem dades mostrals per posar a prova una hipòtesi sobre els paràmetres poblacionals, es pot cometre l'error de actuar com si la hipòtesi fos falsa quan no ho és realment. La probabilitat d'aquest error és podria expressar com:

$$\alpha = P(\text{concloure } H_1 \mid H_0 \text{ certa})$$

[Per procediment, aquesta prob està fixada igual a α]

Tipus d'errors en una prova d'hipòtesi. Tipus II

- La situació complementària també es pot produir.
- **Error de tipus II**. En la mateixa situació, es pot cometre l'error de no trobar evidència en contra de la hipòtesi quan realment és falsa. És a dir, no rebutjar una hipòtesi que no és certa. La probabilitat d'aquest error és pot expressar com:

$$\beta = P(\text{concloure } H_0 \mid H_1 \text{ certa}),$$

- Aquest valor, en general, no és controlable i normalment no es pot saber quant val perquè depèn del valor real del paràmetre testejat (que és desconegut).

Tipus d'error (risc)		**Decisió o Acció**	
		A_0	A_1
Realitat	H_0	**Decisió correcta**	**Error Tipus I (risc α)**
	H_1	**Error Tipus II (risc β)**	**Decisió correcta**

Tipus d'errors en una prova d'hipòtesi. Exercici

- Control de qualitat. Un processador ha de funcionar a certa velocitat μ_0 però el sistema de fabricació pot desestabilitzar-se i baixar-la a μ_1. Estudiades les conseqüències, l'equip directiu demana a l'estadístic que dissenyi un estudi al que sotmetre cada nou processador abans de instal·lar-lo i vendre-ho.
- Després de uns quants càlculs, l'estadístic:
  - posa μ_0 a H_0 i μ_1 a H_1
  - fixa $\alpha = 0.05$ i $\beta = 0.10$
  - proposa fer **'n' proves amb cada processador**
  - acceptar-ho si queda per damunt de un cert llindar L i rebutjar en cas contrari
- Quan posem en marxa l'estudi,
  1. Quina proporció de processadors correctes seran rebutjats?
  2. Quina proporció d'incorrectes arribaran al mercat?

Tipus d'errors en una prova d'hipòtesi. Exemple

- Ch és un navegador amb fama de ràpid, i la marca dominant MD no vol perdre la seva hegemonia. **Suposem que la velocitat mitjana de Ch per carregar una pàgina patró és 700 u., i la de MD és 600 u**. La desviació típica és 150 u. Fixem α = 0.025 (unilateral)
- Si fem 10 proves independents de càrrega per cada navegador:

Ch	$\bar{y}_A = 680$	$S_A = 89$	$n = 10$
MD	$\bar{y}_B = 597$	$S_B = 147$	$n = 10$

- Com el *P*-valor resultant és 0.07, no rebutgem la hipòtesi d'igualtat i MD proclama ('testat científicament') **que el seu navegador és tan ràpid com Ch**.
- Però aquesta conclusió no és correcta: no poder rebutjar la hipòtesi nul·la (Ch és igual a MD) no implica demostrar la seva veritat. Com ja hem dit, hi ha un cert risc de que, sent diferents els dos navegadors, no puguem trobar-ne l'evidència.
- Com en aquest cas coneixem la diferència real, anem a calcular el risc 'beta' β.

Tipus d'errors en una prova d'hipòtesi. Exemple (cont)

- La clau és estudiar la distribució de l'estadístic de referència
- Sota H_0, la diferència de mitjanes mostrals és:

$$\hat{z} = \frac{\bar{y}_A - \bar{y}_B}{\sigma\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}} \sim N(0,1) \rightarrow \bar{y}_A - \bar{y}_B \sim \mathrm{N}\left(0, \sigma\sqrt{\frac{1}{n_A}+\frac{1}{n_B}}\right) = N\left(0{,}150\sqrt{\frac{1}{10}+\frac{1}{10}}\right) = N(0{,}67.08)$$

- Definint α = 2.5% unilateral, la regió crítica es troba per diferències de les mitjanes mostrals més grans que 1.96·67.08 = 131.48 u.
- En realitat, la diferència entre mitjanes és de 100. Com les mostres provenen d'aquesta situació, comprovem com de probable és que NO puguem rebutjar H_0:

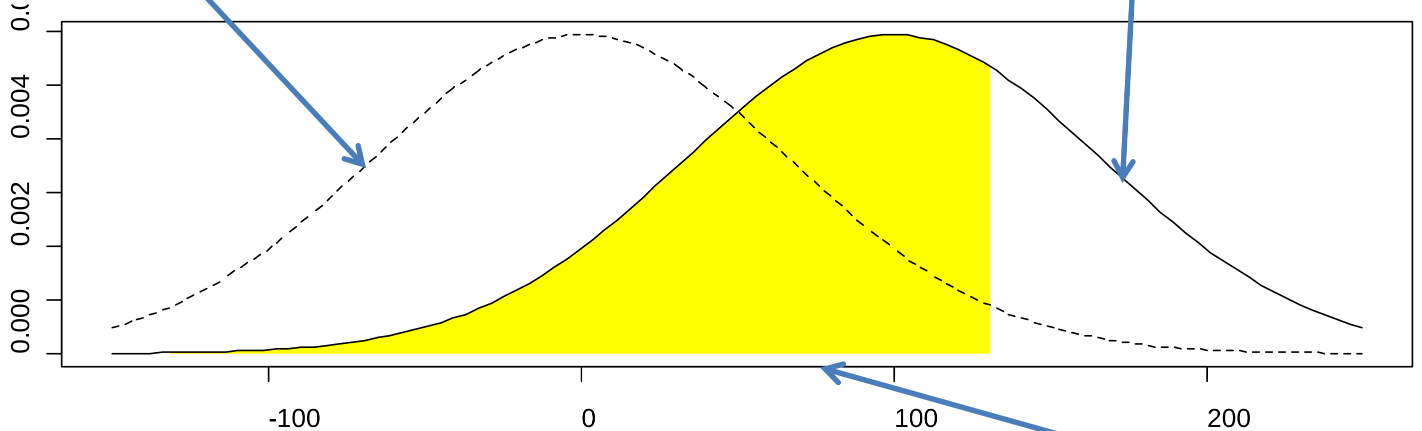
P("Error tipus II") = P($\bar{y}_A - \bar{y}_B$ < 131.48 | H_1) = P(Z < (131.48-100)/67.08) = P(Z<0.47) = 0.68

- Veiem que MD tenia molt fàcil resoldre la prova a la seva conveniència: era molt probable no trobar cap diferència significativa. La prova és **poc potent** (**potència = 1–β**).

Tipus d'errors en una prova d'hipòtesi. Exemple (cont)

Distribució "ingènua" per a la diferència de mitjanes mostrals (noteu que la campana té esperança zero)

Distribució real per a la diferència de mitjanes mostrals $(\bar{y}_A - \bar{y}_B)$. Els valors típics estaran al voltant de 100, que és la diferència autèntica de les mitjanes poblacionals $(\mu_A - \mu_B)$



Regió d'"*acceptació*"
(no es pot rebutjar la hipòtesi nul·la)

Diferència observada:
680 – 597 = 83 u.

Tipus d'errors en una prova d'hipòtesi. Exemple (Cont)

- En el nostre exemple, podem estirar de les orelles als de MD per:
  - La conclusió ha estat incorrectament exposada
  - El disseny és defectuós: li manca potència (la *n* és petita); o, més aviat, si volien demostrar equivalència, havien d'haver plantejat un altre tipus d'estudi (que no veurem en aquest curs)
- En conseqüència, l'experimentador ha d'assumir que el seu estudi està exposat a diferents perills:
  - Mostres no aleatòries (els individus no són independents entre sí)
  - Assignació de X no aleatòria (individus no similars entre els grups)
  - Variables amagades que pertorben la resposta observada
- I estar disposat a posar mesures per evitar errors com aquests. A més a més, ha de saber que l'anàlisi d'un estudi estadístic no és una demostració matemàtica i, com a mínim, la conclusió ha de ser prudent.

To do or not to do. Arbre de decisions i conseqüències

- **H_0: Un mètode nou no és efectiu?**
  - **Sí és efectiu**
    - **Rebutjar H_0** (s'implanta nou)
      - + Beneficis
    - **No rebutjar H_0** (NO s'implanta nou)
      - - Absència de beneficis potencials
      - - Costos de I+D sense retorn
      - - Pèrdua de competitivitat
      - - Frustració per no asolir objectius
  - **NO és efectiu**
    - **Rebutjar H_0** (s'implanta nou)
      - - Malestar dels usuaris
      - - Pèrdues a llarg termini
      - - Dilema: invertir més per resoldre el problema, o aturar la inversió
    - **No rebutjar H_0** (NO s'implanta nou)
      - - Decepció inicial
      - + Protecció drets usuaris
      - + Oportunitats per buscar competitivitat per altres vies

Realitat desconeguda | ***Decisió*** | ***Conseqüències***

La construcció d'un arbre com aquest ens ha de portar a determinar les probabilitats dels errors tipus I i II

Annexes

Annexe I: Comparació $\pi_1 = \pi_2$. Mostres aparellades

Al igual que en el cas independent, la variable que observem "per parelles" és una dicotomia. Per tant, la informació la posarem en forma de taula.

Per exemple, si s'ha preguntat a 100 "fibers": 1. T'agrada el PC?: molt/poc

2. T'agrada el MAC?: molt/poc

freqüència observada: f_{ij}	PC - Molt	PC - Poc	Total
MAC - Molt	61	4	65
MAC - Poc	16	19	35
Total	77	23	N = 100

Davant d'aquesta taula podem realitzar dues preguntes:

1. $P(Molt_{PC}|Molt_{MAC}) = P(Molt_{PC}|Poc_{MAC})$ [Que m'agradi el PC és independent que m'agradi el MAC? o Hi ha un caràcter comú (efecte cas) que guia ambdues respostes?]
2. $P(Molt_{PC}) = P(Molt_{MAC})$ [Existeix preferència per una màquina?]

Annexe I: Comparació $\pi_1 = \pi_2$. Mostres aparellades

- Per a contestar la **primera pregunta** podríem utilitzar l'estadístic de χ^2(Chi Quadrat) previ ja que es tracta de veure si la proporció de Molt_{PC} és la mateixa en les categories Molt_{MAC} i Poc_{MAC}. El resultat és:

$$X^2 = 50.242 > 3.841 = \chi^2_{1,0.95} \qquad o \qquad P - valor = P(\chi^2_1 > 50.242) < 0.0001$$

- Conclusió: Hi ha un "efecte cas":n'hi ha casos favorables a totes les màquines, i n'hi ha de negatius a totes

- Però la pregunta d'interès és la segona, per a la que volem comparar

$$P(\text{Molt}_{PC}) = 0.77 \quad \text{amb} \quad P(\text{Molt}_{MAC}) = 0.65$$

- Resulta que aquestes estimacions no són independents, ja que venen dels mateixos casos i la variància de la seva resta no podria ser la suma de les seves variances. [Penseu que la casella "1,1" contribueix a ambdues estimacions.]
- En realitat, comparar els efectius marginals 77 amb 65 equival a comparar els efectius 16 amb 4 que es troben fora de la diagonal havent suprimit els 61 casos comuns.
- Així, podem reestructurar la taula reunint la informació rellevant (següent diapositiva)

Annexe I: Comparació $\pi_1 = \pi_2$. Mostres aparellades

	f_k
Prefereixen PC	16
Prefereixen MAC	4
Total	20

- De ser certa H_0: $P(Molt_{PC}) = P(Molt_{MAC})$ → H_0: P(prefereixen PC) = P(prefereixen MAC) = 0.5
- I podríem realitzar la prova ja coneguda sent a = casos favorables a PC i b = casos favorables a MAC. Llavors els efectius esperats serien e = (a+b)/2
- En el nostre cas, l'estadístic valdrà:

$$\hat{X}^2 = \sum_{k=1}^{2} \frac{(f_k - e_k)^2}{e_k} = \sum_{k=1}^{2} \frac{\left(f_k - \frac{a+b}{2}\right)^2}{(a+b)/2} = \frac{(a-b)^2}{(a+b)} = \frac{(16-4)^2}{(16+4)} = 7.2$$

- Ja que el p-valor és $P(\chi_1^2 > 7.2) = 0.0073 < \alpha$ o bé que $X^2 = 7.2 > 3.84 = \chi_{1,0.95}$ podem rebutjar H_0
- **Conclusió pràctica**: Hi ha preferència pel PC.

Annexe II: Grandària Mostral. Cas I: Estimar μ

Objectiu: Estimar μ amb un IC95% d'una amplada determinada (A)

$$A = LS_{IC95\%} - LI_{IC95\%} = \left[\bar{y} + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] - \left[\bar{y} - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right] = 2 \cdot Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \rightarrow$$

$$\rightarrow n = \left(\frac{2 \cdot Z_{1-\frac{\alpha}{2}} \cdot \sigma}{A}\right)^2 = \left(\frac{2 \cdot Z_{1-\frac{\alpha}{2}}}{\frac{A}{\sigma}}\right)^2 \approx \frac{16}{\left(\frac{A}{\sigma}\right)^2}$$

Exemple: si volem estimar la mitjana del temps de posta en marxa d'un equip amb una incertesa total *A* (amplada de l'IC) que sigui la meitat de la desviació tipus, *A*/σ = 0.5, implica ***n=64 observacions***.

Annexe II: Grandària Mostral. Cas II: Estimar μ_1-μ_2

Objectiu: Estimar la diferència μ_1-μ_2 amb un IC95% d'una amplada determinada (A)

$$A = LS_{IC95\%} - LI_{IC95\%} = \left[\bar{y}_1 - \bar{y}_2 + Z_{1-\frac{\alpha}{2}} \cdot \frac{\sqrt{2}\sigma}{\sqrt{n}}\right] - \left[\bar{y}_1 - \bar{y}_2 - Z_{1-\frac{\alpha}{2}} \cdot \frac{\sqrt{2}\sigma}{\sqrt{n}}\right] =$$

$$= 2 \cdot Z_{1-\frac{\alpha}{2}} \cdot \frac{\sqrt{2}\sigma}{\sqrt{n}} \rightarrow n = \left(\frac{2\sqrt{2} \cdot Z_{1-\frac{\alpha}{2}} \cdot \sigma}{A}\right)^2 = \left(\frac{2\sqrt{2} \cdot Z_{1-\frac{\alpha}{2}}}{\frac{A}{\sigma}}\right)^2 \approx \frac{32}{\left(\frac{A}{\sigma}\right)^2}$$

Exemple: si volem estimar la diferència de mitjanes del temps de posta en marxa de dos models d'equip amb una incertesa total *A* (amplada de l'IC) que sigui la meitat de la desviació tipus comuna , $A/\sigma = 0.5$, implica *n=128 observacions*.

Annexe II: Grandària Mostral. Cas III: Contrastar $\mu_1 = \mu_2$(independ.)

Objectiu: Contrastar $\mu_1 = \mu_2$ en 2 mostres independents

$$\begin{cases} H_0: \mu_1 - \mu_2 = 0 \\ H_1: \mu_1 - \mu_2 = \Delta \end{cases}$$

$$V(\bar{y}_1 - \bar{y}_2) = V(\bar{y}_1) + V(\bar{y}_2) = \frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = (si\ \ n_1 = n_2 = n) = \frac{2\sigma^2}{n}$$

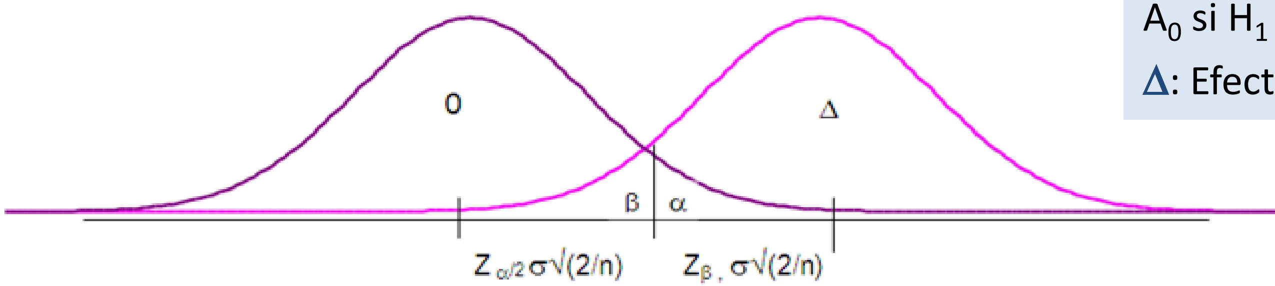
Elements:

σ: Variabilitat del fenomen en estudi

α: Freqüència admesa de A_1 si H_0 és certa

β: Freqüència admesa de A_0 si H_1 és certa

Δ: Efecte de l' intervenció



$$Z_{\alpha/2}\ \sigma\sqrt{(2/n)} + Z_{\beta}\sigma\sqrt{(2/n)} = \Delta$$

Per tan la grandària de cada grup és: $\boldsymbol{n} = \dfrac{\mathbf{2\sigma^2\left(Z_{1-\frac{\alpha}{2}} + |Z_\beta|\right)^2}}{\mathbf{\Delta^2}}$

Exemple: Quina grandària mostral seria necessària per detectar una diferència de 10 cm en la alçada mitjana d'homes i dones? ($\alpha = 0.05$; $\beta = 1–0.8$ i $\sigma = 8$ cm)

$$n = (2 \cdot 8^2(1.96 + 0.84)^2)/10^2 \approx 10$$

Annexe II: Grandària Mostral. Cas IV: Contrastar $\pi_1 = \pi_2$(independ.)

Objectiu: Contrastar $\pi_1 = \pi_2$ en 2 mostres independents

$$\begin{cases} H_0: \pi_1 - \pi_2 = 0 \\ H_1: \pi_1 - \pi_2 = \Delta \end{cases}$$

Apliquem la següent fòrmula per saber la grandària per grup:

$$n = \frac{4}{\Delta^2}$$

Exemple: Si volem demostrar que una nova versió deixa molt satisfets a un 20% més de clients (posem: π_1=0.5 i π_2=0.7) aleshores $(\pi_1 - \pi_2)^2$ = 0.04 que implica n=100 per grup.

$$n = (2.8^2(1.96 + 0.84)^2)/10^2 \approx 10$$

Annexe II: Grandària Mostral. Exercici

Pràctica de B7: Uns companys han comparat els temps de descàrrega de 2 navegadors i la seva pràctica ha anat tan bé que un altre professor els hi suggereix fer un disseny nou per publicar un article a una revista de informàtica.

Amb les dades de la seva pràctica han observat una desviació tipus S igual a 54 i una diferència de temps igual a 22. Però, per 'garantir el tir' el professor els hi suggereix treballar amb un $\sigma = 60$.

Quines grandàries mostrals necessitem en les següents situacions?

1) Volen estimar $\mu_1 - \mu_2$ amb una amplitud d'interval, A = 40?

$n = 32/(A/\sigma)^2 = 72$

2) Volen contrastar una Δ=20 amb α=0.025 i β=0.20 unilaterals.

$$n = \frac{2\sigma^2\left(Z_{1-\frac{\alpha}{2}} + |Z_\beta|\right)^2}{\Delta^2} = \frac{2 \cdot 60^2 \cdot (1.96+0.84)^2}{20\text{^}2} \approx 141.47 \rightarrow n = 142$$

Annexe II: Grandària Mostral. Potència sense augmentar n

Podem augmentar la potencia sense augmentar la n?

Ho aconsseguirem si trobem sistemes per baixar la variabilitat **s²** de la resposta. Aquesta variabilitat té 2 components:

$S^2 = S_B^2 + S_W^2$ - S_B^2: variabilitat **entre** (***B****etween*) casos

- S_W^2: variabilitat **intra** (***W****ithin*) casos

[Ex: la variabilitat de la variable "pes" ve donada pel fet de que les persones tenim constitucions diferents (S_B^2) i cada persona pot pesar lleugerament diferent depenent del moment del dia (S_W^2)]

La variància de la resposta la podem disminuir amb diferents estratègies:

1. Utilitzant un disseny aparellat, que només tingui S_W^2 $\rightarrow$ $S^2 = 2 \cdot S_W^2$
2. Definint el canvi **C** respecte al valor inicial (*baseline*) $\rightarrow$ $S^2 = 2 \cdot S_W^2$
3. Fent promitjos de K repeticions independents: $\rightarrow$ $S^2 = S_B^2 + S_W^2/K$

Aquestes estratègies es complementen, ja que controlen dos fonts diferents de variabilitat.