**NOM:**____________________________________________________________

**(Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliciteu i justifiqueu els passos en les respostes)**

## Problema 1.

Baixar una cançó del nostre grup de rock preferit d'un determinat servidor en Internet al disc dur del vostre ordinador tarda cert temps que podem considerar com una variable aleatòria X amb distribució normal amb una esperança de 10 segons i una desviació típica de 2 segons.

**A. (2 punts per cadascun dels 5 apartats)**

**1.-**

1.1 Quina és la probabilitat de que baixar una cançó trigui més de 12 segons?

1.2 I la d'obtenir la cançó en un temps inferior a 7 segons?

**2.-** Si volem calcular les probabilitats de temps de baixada d'un CD recopilatori, amb 35 cançons (suposant en la baixada de les 35 cançons és independent) hem de considerar la variable: $X = X_1+X_2+ \dots +X_{35}$.

2.1 ¿Quina és la llei de distribució d'$X$?. Com es justifica? Indica els paràmetres de la llei de distribució.

2.2 Quina és la probabilitat de que el CD complet trigui entre 320 segons i 360 segons?

**3.-** Suposant que el temps de baixada segueix una llei uniforme a l'interval de 8 a 16 segons, considerant 35 suficient per aplicar el TCL:

3.1 Quina és la distribució de la variable suma $X = (X_1+X_2+ \dots +X_{35})$?

3.2 Quina és la distribució de la variable promig $Y= (Y_1+Y_2+ \ldots +Y_{35})/35$?

**4.-** D'altra banda, s'ha observat que 4 de cada 10 CDs surten defectuosos i no podem fer l'enregistrament de la nostra música d'una manera correcta.
4.1 Quina és la llei de distribució de la variable R~"nombre de CDs defectuosos per paquet de 10 CDs"? Indica també el/s paràmetre/s d'aquesta llei.

4.2 Quina és la probabilitat de trobar exactament 5 CDs defectuosos en 1 paquet de 10?

**5.-** Donada l'alt nombre de CDs defectuosos, els usuaris fan arribar les seves queixes al distribuïdor. El nombre de reclamacions és en promig de 4,5 reclamacions/dia.
5.1 Quina llei de distribució segueix la variable Y~"Nombre de reclamacions al dia"? Indica també el/s paràmetre/s d'aquesta llei. Quina és la probabilitat de recollir menys de 7 reclamacions en 2 dies?

5.2 Quina llei de distribució segueix la variable W~"Dies entre reclamacions"? Indica també el/s paràmetre/s d'aquesta llei. Quina és la probabilitat d'estar 2 dies o més dies sense rebre reclamacions?

**NOM:**____________________________________________________________________

**(Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliciteu i justifiqueu els passos en les respostes).**

Seguint amb l'exemple anterior d'una variable aleatòria **X amb distribució normal** que representa el temps de baixar una cançó d'un determinat servidor en Internet, ara considerem que conèixer l'esperança del temps de baixada i la desviació poblacionals és poc creïble.

## B.

Durant 7 dies consecutius, i a la mateixa hora, s'ha baixat una cançó del servidor al disc dur, i s'ha observat el temps en segons que triga. Els resultats es mostren a continuació:

9,05 9,77 8,58 11,58 7,47 9,32 10,80

(amb alguns càlculs intermedis: $\Sigma x = 66{,}57$ $\Sigma x^2 = 644{,}3715$)

1) (0,5 punts) Doneu una estimació puntual de la mitjana i de la desviació del temps que triga en baixar una cançó.

2) (0,5 punts) Amb aquesta mostra, ¿en quant estimeu l'error de la mitjana o error típic?

3) (2 punts) Estimeu per IC al 95% de confiança la mitjana poblacional.

4) (2 punts) Estimeu per IC al 95% de confiança la desviació poblacional

5) (2 punts) Quin valor de n hauríem de recollir per obtenir un IC de la mitjana poblacional al 95% de confiança amb una amplada de 0,5 seg (en aquest cas sí que assumim una desviació poblacional de 2 sg)

6) (1,5 punts) Experiments realitzats amb anterioritat informen que el temps de baixada, utilizant aquest servidor i aquest tipus d'ordinador, és en mitjana de 12 segons. Poseu a prova si aquesta afirmació es pot considerar creïble o no amb un CH. Considereu un risc del 5%.

- (0,25) Hipòtesis (indicant si la prova és bilateral o unilateral):

- (0,25) Càlcul del valor de l'estadístic:

- (0,25) ¿Quin és el punt crític que utilizareu per decidir si l'estadístic anterior permet creure'ns la hipòtesis?

- (0,25) Representeu gràficament en la distribució de l'estadístic el valor obtingut i el punt crític:

- (0,25) Decisió de CH:

- (0,25) Relacionar la decisió anterior amb el resultat del IC de la pregunta 3

7) (1,5 punts) Se sap que la proporció de persones que no baixen cançons d'internet és d'un 20%. Es pren una mostra de 100 persones i s'obté que la proporció és del 22%. Hi ha evidència de que la proporció hagi augmentat? (perquè si no es pot acceptar el 20% sinó que es confirma l'augment es volen fer canvis legislatius) Per respondre plantegeu un CH i considereu un risc del 5%.

- (0,25) Hipótesis (indicant si la prova és bilateral o unilateral):

- (0,25) Càcul del valor de l'estadístic:

- (0,25) Càlcul del P-valor:

- (0,25) Representeu gràficament en la distribució de l'estadístic el valor obtingut i el P-valor:

- (0,50) Decisió de CH i interpretació:

**NOM:**__________________________________________________________________________

**(Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliciteu i justifiqueu els passos en les respostes).**

## Problema 2. (tots els apartats valen igual)

La diferència entre la velocitat de baixada contractada (VC) i la velocitat real (VR) en les connexions a Internet és una de les queixes més freqüents entre els usuaris de les línies ADSL. Els objectius d'aquest estudi són: (A) estudiar si són iguals (en mitjana) i estimar la magnitud del possible desfasament; i (B) establir una relació entre les dues. Coneixements previs aconsellen treballar amb el logaritme natural dels temps, anotats per C i R. Sigui D=R-C. Es disposa de les següents dades:

$n= 30$ $\Sigma C= 253$ $\Sigma C^2= 2150$ $\Sigma R= 245$ $\Sigma R^2= 2021$ $\Sigma D= -7.6$

$\Sigma D^2= 5.7$ $\Sigma CR= 2080$ $\Sigma CD = -63.5$ $\Sigma RD= -65.3$

**A.** Es pot considerar que els temps C i R són iguals? Per respondre-ho, fem inferència sobre la seva diferència:

**1.** D'acord amb l'enunciat, es tracta d'un disseny 'aparellat' o independent? Raoneu la resposta.

**2.** Càlculeu les mitjanes i desviacions típiques de C, R i D

**3.** Calculeu la covariància i la correlació entre R i C

**4.** Sota la PS de si són iguals els temps C i R, escriviu la hipòtesi (indicant si és bilateral o unilateral), l'estadístic i la seva distribució

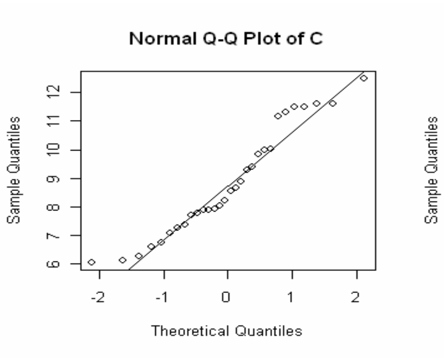
**5.** Sota PS, calculeu l'estadístic i (aproximat) el p-valor

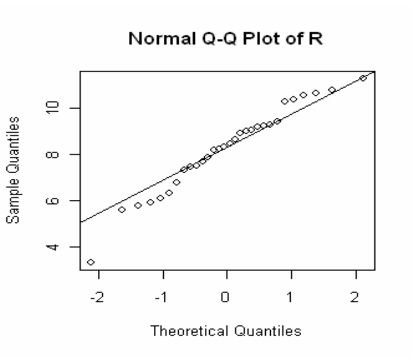
**6.** Sota PS, indiqueu la conclusió i interpretació pràctica:

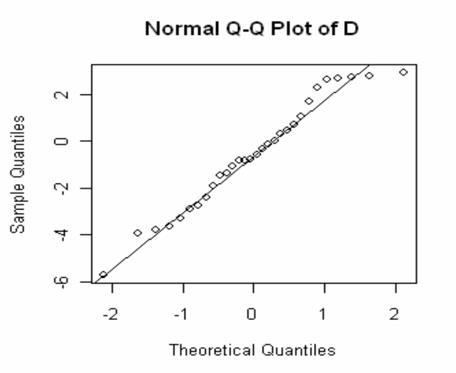
**7.** Calculeu un IC al 95% per a la mitjana de la diferència

**8.** Interpreteu l'anterior IC

**9.** Indiqueu i valoreu la premissa de Normalitat. A quina variable s'aplica?




**10.** Interpreteu i comenteu globalment l'estudi d'aquest apartat A

**NOM:**____________________________________________________________________

**(Poseu el nom i contesteu cada pregunta en el seu lloc reservat. Expliciteu i justifiqueu els passos en les respostes).**

**B.**. Si contractem C, què es pot esperar sobre R? Per respondre-ho, aplicarem el model de regressió lineal

**1.** Estimeu puntualment els coeficients de la recta de regressió:

**2.** Ompleneu la taula de descomposició de variabilitat (indicant al costat el càlcul dels valors de la columna SQ):

| | *SQ* | *Graus llibertat GdL* | *QM =SQ/GdL* | *Rati* | P-valor |
|---|---|---|---|---|---|
| *Explicada pel model* | | | | | *<0'001* |
| *Residual* | | | | | |
| *Total* | | | | | |

**3.** Estimeu i interpreteu el coeficient de determinació $R^2$

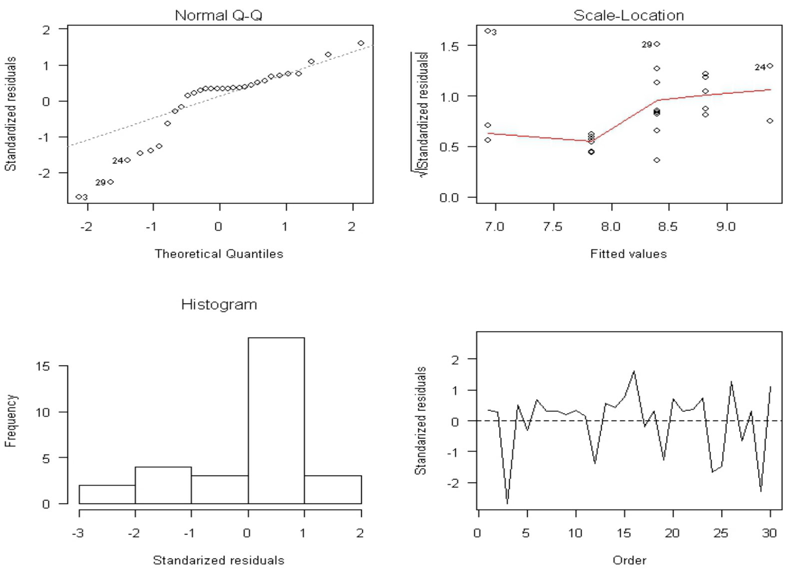
**4.** Poseu a prova H: V(R|C) = V(R): estadístic, distribució, càlcul, conclusió i interpretació:

**5.** Poseu a prova H: $\beta_1 = 0$: Estadístic, distribució, càlcul, conclusió i interpretació:

**6.** Estimeu puntualment i per interval el valor de R predit per a un contracte de C=8.01 (ln(VC=3010.917 Kb/s))

**7.** De fet si el contracte es complís la constant seria 0 i la pendent 1. Creieu que es compleix H: $\beta_1=1$ ?

**8.** Valoreu les premisses.




**9.** Interpreteu i comenteu globalment l'estudi d'aquest apartat B

**10.** Torneu a contestar la pregunta 1 de l'apartat A emprant ara la nova informació de que disposeu

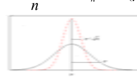
# Distribucions de variables discretes i contínues

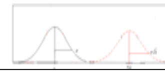
| Distribució | Declaració | Funció de probabilitat o de densitat | Funció distribució $F_X(k)=\sum_{i<=k} P_X(i)$ o $\int_{-\infty}^{k} f_X(x)dx$ | Esperança E(X) | Variància V(X) |
|---|---|---|---|---|---|
| Bernoulli | X~Bern(p) | $P_X(k)=\begin{cases} q & k=0 \\ p & k=1 \end{cases}$ | $F_X(k)=\sum_{i<=k} P_X(i)$ | $p$ | $p \cdot q$ |
| Binomial | X~B(n,p) | $P_X(k)=\binom{n}{k} p^k \cdot q^{n-k}$ <br> $k=0,1,\ldots,n$ <br> `(R:dbinom(k,n,p))` | $F_X(k)=\sum_{i<=k} P_X(i)$ <br> ( taules estadístiques ) <br> `( R:pbinom(k,n,p) )` | $p \cdot n$ | $p \cdot q \cdot n$ |
| Poisson | X~P(λ) * | $P_X(k)=\frac{\lambda^k \cdot e^{-\lambda}}{k!}$ $k=0,1,2,\ldots$ <br> `( R:dpois(k,λ) )` | $F_X(k)=\sum_{i<=k} P_X(i)$ <br> ( taules estadístiques ) <br> `( R:ppois(k,λ) )` | $\lambda$ | $\lambda$ |
| Exponencial | X~Exp(λ) * | $f_X(x)=\lambda \cdot e^{-\lambda \cdot x}$ $x>0$ <br> `( R:dexp(x,λ) )` | $F_X(x)=1-e^{-\lambda \cdot x}$ <br> `( R:pexp(x,λ) )` | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| Uniforme | X~U[a,b] | $f_X(x)=\frac{1}{b-a}$ $a<x<b$ <br> `( R:dunif(x,a,b) )` | $F_X(x)=\frac{x-a}{b-a}$ <br> `( R:punif(x,a,b) )` | $\frac{(a+b)}{2}$ | $(b-a)^2/12$ |
| Normal | X~N(μ,σ) | $f_X(x)=\frac{1}{\sigma\sqrt{2\cdot\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ <br> `( R:dnorm(k,μ,σ) )` | $F_X(x)=?$ <br> (taules estadístiques N(0,1)) <br> `( R:pnorm(k,μ,σ) )` | μ | $\sigma^2$ |

0 < p < 1; q = 1 - p; n, r enter > 0; λ, a, b, μ, σ real > 0.0
*λ paràmetre del procés Poisson: variables Poisson i Exponencial

# Propietats de les mostres i Intervals de Confiança

*TCL:* $X_1,\ldots,X_n$ i.i.d. (n→∞), amb $E(X_i)=\mu$ i $V(X_i)=\sigma^2$, llavors $\frac{\sum_{i=1}^n X_i}{n}=\bar{X}_n \approx N(\mu,\sigma^2/n)$ **( i també** $\sum_{i=1}^n X_i \approx N(n\mu,\sigma^2 n)$ **)**





*Estadístic mitjana mostral* ($\bar{x}$): $\frac{(\bar{x}-\mu)}{\sqrt{\sigma^2/n}} \approx N(0,1)$ $\frac{(\bar{x}-\mu)}{\sqrt{s^2/n}} \approx t_{n-1}$ on $\bar{x}=\sum_{i=1}^n x_i/n$

*Estadístic variància mostral* ($s^2$): $s^2\frac{n-1}{\sigma^2} \approx \chi^2_{n-1}$ on $s^2=\frac{\sum_{i=1}^n (x_i-\bar{x})^2}{n-1}=\frac{\sum_{i=1}^n x_i^2-n(\bar{x})^2}{n-1}=\frac{\sum_{i=1}^n x_i^2-\frac{\left(\sum_{i=1}^n x_i\right)^2}{n}}{n-1}$

| Paràmetre | Estadístic | Premisses | Distribució | Interval de Confiança 1-α (Risc α) |
|---|---|---|---|---|
| μ | $\hat{z}=\frac{(\bar{x}-\mu)}{\sqrt{\sigma^2/n}}$ | [ X → N ò n ≥ 30 ] i σ coneguda | $\hat{z} \to N(0,1)$ | $\mu \in (\bar{x} \pm z_{1-\alpha/2}\sqrt{\frac{\sigma^2}{n}})$ |
| μ | $\hat{t}=\frac{(\bar{x}-\mu)}{\sqrt{s^2/n}}$ | X → N | $\hat{t} \to t_{n-1}$ | $\mu \in (\bar{x} \pm t_{n-1,1-\alpha/2}\sqrt{\frac{s^2}{n}})$ |
| μ | $\hat{z}=\frac{(\bar{x}-\mu)}{\sqrt{s^2/n}}$ | n ≥ 100 | $\hat{z} \to N(0,1)$ | $\mu \in (\bar{x} \pm z_{1-\alpha/2}\sqrt{\frac{s^2}{n}})$ |
| σ (normal) | $\hat{X}^2=\frac{s^2(n-1)}{\sigma^2}$ | X → N | $\hat{X}^2 \to \chi^2_{n-1}$ | $\sigma^2 \in \left(\frac{S^2(n-1)}{\chi^2_{n-1,1-\alpha/2}},\frac{S^2(n-1)}{\chi^2_{n-1,\alpha/2}}\right)$ |
| π (Binomial) | $\hat{z}=\frac{(p-\pi)}{\sqrt{\pi(1-\pi)/n}}$ | (1-π)n ≥ 5 <br> πn ≥ 5 | $\hat{z} \to N(0,1)$ | $\pi \in (P \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}})$ <br> $\hat{\pi}=P$ o $\hat{\pi}=0'5$ |
| λ (Poisson) | $\hat{z}=\frac{(L-\lambda)}{\sqrt{\lambda}}$ | λ ≥ 5 | $\hat{z} \to N(0,1)$ | $\lambda \in (L \pm z_{1-\alpha/2}\sqrt{L})$ |

# Proves de Significació

| Paràmetre | Hipòtesis | Estadístic | Premisses | Distribució sota H | Criteri Decisió (Risc α) |
|---|---|---|---|---|---|
| | H : μ = $\mu_0$ | $\hat{z}=\frac{(\bar{y}-\mu_0)}{\sqrt{\sigma^2/n}}$ | Y→N ò n≥30 <br> y σ coneguda | $\hat{z} \to N(0,1)$ | Rebutjar H si <br> $\lvert\hat{z}\rvert > z_{1-\alpha/2}$ <br> ($\lvert\hat{z}\rvert$>1'96 amb α=5%) |
| μ | H : μ = $\mu_0$ | $\hat{t}=\frac{(\bar{y}-\mu_0)}{\sqrt{S^2/n}}$ | Y → N | $\hat{t} \to t_{n-1}$ | Rebutjar H si <br> $\lvert\hat{t}\rvert > t_{n-1,1-\alpha/2}$ <br> ($\lvert\hat{t}\rvert > t_{n-1,0'975}$ amb α=5%) |
| μ | H : μ = $\mu_0$ | $\hat{z}=\frac{(\bar{y}-\mu_0)}{\sqrt{S^2/n}}$ | n ≥ 100 | $\hat{z} \to N(0,1)$ | Rebutjar H si <br> $\lvert\hat{z}\rvert > z_{1-\alpha/2}$ |
| π (Binomial) | H : π = $\pi_0$ | $\hat{z}=\frac{(p-\pi_0)}{\sqrt{\pi_0(1-\pi_0)/n}}$ | $(1-\pi_0)n \geq 5$ <br> $\pi_0 n \geq 5$ | $\hat{z} \to N(0,1)$ | Rebutjar H si <br> $\lvert\hat{z}\rvert > z_{1-\alpha/2}$ <br> ($\lvert\hat{z}\rvert$>1'96 amb α=5%) |
| Anexe: λ (Poisson) | H : λ = $\lambda_0$ | $\hat{z}=\frac{(f-\lambda_0)}{\sqrt{\lambda_0}}$ | $\lambda_0 \geq 5$ | $\hat{z} \to N(0,1)$ | Rebutjar H si <br> $\lvert\hat{z}\rvert > z_{1-\alpha/2}$ <br> ($\lvert\hat{z}\rvert$>1'96 amb α=5%) |
| σ (normal) | H : σ = $\sigma_0$ | $\hat{X}^2=\frac{S^2(n-1)}{\sigma^2}$ | Y → N | $\hat{X}^2 \to \chi^2_{n-1}$ | Rebutjar H si <br> $\hat{X}^2 < \chi^2_{n-1,\alpha/2}$ <br> o $\hat{X}^2 > \chi^2_{n-1,1-\alpha/2}$ |
| **En les proves unilaterals s'acumula el valor de *P* a un sol costat** | | | | | **H:μ ≤$\mu_0$ → Rebutjar H si $\hat{z} > z_{1-\alpha}$** <br> **H:μ ≥$\mu_0$ → Rebutjar H si $\hat{z} < -z_{1-\alpha}$** |

# Proves de μ i σ en 2 mostres

| Paràmetre | Hipòtesis | Estadístic | Premisses | Distrib. sota $H_0$ | Decisió (Risc α) |
|---|---|---|---|---|---|
| μ | $H_0 : \mu_1 = \mu_2$ <br> $H_1 : \mu_1 \neq \mu_2$ | $\hat{z}=\frac{(\bar{y}_1-\bar{y}_2)}{\sqrt{\sigma_1^2/n_1+\sigma_2^2/n_2}}$ | [$Y_1$ $Y_2$ →N ò $n_1$ $n_2$ ≥30] mas ind. i $\sigma_1$ $\sigma_2$ coneg | $\hat{z} \to N(0,1)$ | Rebutjar si <br> $\lvert\hat{z}\rvert >$ $\mathbf{z}_{1-\alpha/2}$ <br> ($\lvert\hat{z}\rvert$>1'96 amb α=5%) |
| μ | $H_0 : \mu_1 = \mu_2$ <br> $H_1 : \mu_1 \neq \mu_2$ | $\hat{t}=\frac{(\bar{y}_1-\bar{y}_2)}{S\sqrt{1/n_1+1/n_2}}$ <br> $S^2=\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ | $Y_1$ , $Y_2$ → N <br> $\sigma_1 = \sigma_2$ <br> m.a.s indep. | $\hat{t} \to t_{n1+n2-2}$ | Rebutjar si <br> $\lvert\hat{t}\rvert > t_{n1+n2-2,\ 1-\alpha/2}$ <br> $\lvert\hat{t}\rvert > t_{n1+n2-2,0'975}$ amb α=5% |
| μ | $H_0 : \mu_1 = \mu_2$ <br> $H_1 : \mu_1 \neq \mu_2$ | $\hat{z}=\frac{(\bar{y}_1-\bar{y}_2)}{\sqrt{S_1^2/n_1+S_2^2/n_2}}$ | $n_1$ , $n_2$ ≥ 100 <br> m.a.s indep | $\hat{z} \to N(0,1)$ | Rebutjar si <br> $\lvert\hat{z}\rvert >$ $\mathbf{z}_{1-\alpha/2}$ |
| μ | $H_0 : \mu_1 = \mu_2$ <br> $H_1 : \mu_1 \neq \mu_2$ | $\hat{t} = (\bar{D}-\mu_0) / (s_D/\sqrt{n})$ | D → N <br> m.a aparellada | $\hat{t} \to t_{n-1}$ | Rebutjar si <br> $\lvert\hat{t}\rvert > t_{n-1,\ 1-\alpha/2}$ |
| σ (normal) | $H_0 : \sigma^2_1 = \sigma^2_2$ <br> $H_1 : \sigma^2_1 \neq \sigma^2_2$ | $\hat{F} = S_A^2 / S_B^2$ <br> Sent $S_A^2 > S_B^2$ | $Y_1$ , $Y_2$ → N <br> m.a.s indep | $\hat{F} \to F_{nA-1,nB-1}$ | Rebutjar si <br> $\hat{F} > F_{n1-1,n2-1,\ 1-\alpha/2}$ |
| Les corresponents proves unilaterals es fan acumulant el risc α a un costat | | | | | |

## Proves de $\pi$ en 2 mostres

| Proves de Comparació de 2 Paràmetres més usuals | | | | |
|---|---|---|---|---|
| Hipòtesi | Estadístic | Premisses | Distrib.($H_0$) | Decisió $\alpha$=0'05 |
| $H_0$: $\pi_1 = \pi_2 = \pi$<br>$H_1$: $\pi_1 \neq \pi_2$ | $z = \frac{(P_1 - P_2)}{\sqrt{P(1-P)/n_1 + P(1-P)/n_2}}$<br>$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2}$ | $e_{ij} \geq 5$ $\forall$ ij<br>m.a.s indep. | $\hat{z} \rightarrow N(0,1)$ | Rebutjar si<br>$\lvert \hat{z} \rvert >$ 1'96 |
| | $\hat{\chi}^2 = \sum_{\forall ij} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$ | | $\hat{\chi}^2 \rightarrow \chi^2_{(I-1)(J-1)}$ | Rebutjar si<br>$\hat{\chi}^2 > \chi^2_{(I-1)(J-1),0'95}$ |
| Homogeneïtat<br>$H_0$ : $\pi(J_j \mid I_i) = \pi(J_j \mid I_{i'})$ $\forall$ i,i',j<br>$H_1$ : $\exists$ j $\mid$ $\pi(J_j \mid I_i) \neq \pi(J_j \mid I_{i'})$<br>Independència<br>$H_0$ : $\pi(J_j \cap I_i) = \pi(J_j)\,\pi(I_i)$ $\forall$ i,j<br>$H_1$ : $\exists$ ij $\mid$ $\pi(J_j \cap I_i) \neq \pi(J_j) / \pi(I_i)$ | $\hat{\chi}^2 = \sum_{\forall ij} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$ | $e_{ij} \geq 5$ $\forall$ ij<br>m.a.s indep. | $\hat{\chi}^2 \rightarrow \chi^2_{(I-1)(J-1)}$ | Rebutjar si<br>$\hat{\chi}^2 > \chi^2_{(I-1)(J-1),0'95}$ |
| $H_0$: $\pi_1 = \pi_2 = \pi$<br>$H_1$: $\pi_1 \neq \pi_2$ | $\hat{\chi}^2 = \frac{(a-b)^2}{(a+b)}$ | a, b $\geq$ 5<br>m.a.s<br>aparellades | $\hat{\chi}^2 \rightarrow \chi^2_1$ | Rebutjar si<br>$\hat{\chi}^2 > \chi^2_{1,0'95}$ |
| Les corresponents proves unilaterals es fan acumulant el risc $\alpha$ a un costat. | | | | |

## Descomposició de la variabilitat

$$\bar{y} = \frac{\sum y_i}{n} \qquad s_y^2 = \frac{\sum y_i^2 - \frac{(\sum y_i)^2}{n}}{n-1} \qquad s_{xy} = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{n-1}$$

### Model quantitativa vs quantitativa (Regressió)

$$b_1 = \frac{S_{xy}}{S_x^2} = r_{XY} \frac{S_y}{S_x}$$

$$b_0 = \bar{Y} - b_1 \bar{X}$$

$$S^2 = \frac{\sum e_i^2}{(n-2)}$$

| | SQ (R: Sum Sq) | Graus llibertat GdL (R: Df) | QM =SQ/GdL (R: Mean Sq) | Rati (R: F value) | P-valor (R: Pr(>F)) |
|---|---|---|---|---|---|
| **Explicada** pel model (R: X) | $SQ_E = \Sigma(y_i - \bar{Y})^2$ | 1 | $QM_E$= $SQ_E$ / 1 | $\hat{F} = \frac{QM_E}{QM_R}$ | 1-qf(F_value, 1,n-2) |
| **Residual** (R:Residual) | $SQ_R = \Sigma(y_i - \hat{y}_i)^2$ | n-2 | $QM_R$= $SQ_R$ / (n-2) | | |
| **Total** | $SQ_T = \Sigma(\hat{y}_i - \bar{Y})^2$ | n-1 | | | |

$$\left(SQ_T = \sum (y_i - \bar{Y})^2 = (n-1) S_Y^2\right) \qquad \left(SQ_E = \sum (\hat{y}_i - \bar{Y})^2 = b_1^2 (n-1) S_x^2 = b_1 (n-1) S_{XY}\right)$$

### Model quantitativa vs categòrica

| | SQ (R: Sum Sq) | Graus llibertat GdL (R: Df) | QM =SQ/GdL (R: Mean Sq) | Rati (R: F value) | P-valor (R: Pr(>F)) |
|---|---|---|---|---|---|
| **Explicada** pel model (R: X) **(ENTRE grups o Between)** | $SQ_E = \sum_{j=1}^{j=k} n_j (\bar{y}_j - \bar{Y})^2$ | k-1 | $QM_E$= $SQ_E$ / (k-1) | $\hat{F} = \frac{QM_E}{QM_R}$ | 1-qf(F_value, k-1,N-k) |
| **Residual** (R:Residual) **(INTRA grups o Within)** | $SQ_R = \sum_{j=1}^{j=k} \sum_{i=1}^{i=n_j} (y_{ji} - \bar{y}_j)^2$ | N-k | $QM_R$= $SQ_R$ / (N-k) | | |
| **Total** | $SQ_T = \sum_{j=1}^{j=k} \sum_{i=1}^{i=n_j} (y_{ji} - \bar{Y})^2$ | N-1 | | | |

$$\left(SQ_R = \sum_{j=1}^{j=k} (n_j - 1) s_j^2\right)$$

## Model quantitativa vs quantitativa (Regressió)
## Estimació i inferència dels paràmetres

| Paràmetre | $\beta_0$ | $\beta_1$ | $\sigma^2$ |
|---|---|---|---|
| Estimador | $b_0 = \bar{Y} - b_1 \bar{X}$ | $b_1 = S_{XY} / S^2_X$ | $S^2 = \Sigma e_i^2/(n-2)$ |
| Esperança | $E(b_0) = \beta_0$ | $E(b_1) = \beta_1$ | $E(S^2) = \sigma^2$ |
| Variància | $V(b_0) = \sigma^2 (1/n + \frac{\bar{X}^2}{(n-1) S_x^2})$<br>$(S_{b_0} = \sqrt{S^2 (1/n + \frac{\bar{X}^2}{(n-1) S_x^2})})$ | $V(b_1) = \sigma^2 / (n-1) S_x^2$<br>$(S_{b_1} = \sqrt{S^2 / (n-1) S_x^2})$ | $V(S^2) = 2\sigma^4/(n-2)$ |
| Distribució | $b_0 \rightarrow N$<br>$(b_0 - \beta_0) / S_{b0} \rightarrow t_{n-2}$ | $b_1 \rightarrow N$<br>$(b_1 - \beta_1) / S_{b1} \rightarrow t_{n-2}$ | $(n-2)S^2/\sigma^2 \rightarrow \chi^2_{n-2}$ |
| Interval de Confiança | IC(95%,$\beta_0$) =<br>$= b_0 \pm t_{n-2,0.975} \cdot S_{b0}$ | IC(95%,$\beta_1$) =<br>$= b_1 \pm t_{n-2,0.975} \cdot S_{b1}$ | IC(95%,$\sigma^2$) $\rightarrow$<br>$(n-2)S^2/\chi^2_{n-2,0.975} \leq \sigma^2$<br>$\leq (n-2)S^2/\chi^2_{n-2,0.025}$ |
| $H_0$ usual | $\beta_0 = 0$ | $\beta_1 = 0$ | |
| Rebutjar $H_0$ si | $b_0 / S_{b0} > t_{n-2,0.975}$ | $b_1 / S_{b1} > t_{n-2,0.975}$ | |

## Model quantitativa vs quantitativa (Regressió)
## Predicció

| | | |
|---|---|---|
| Estimació puntual | $\hat{y}_h = b_0 + b_1 X_h$ | $y_h = \hat{y}_h = b_0 + b_1 X_h$ |
| Estimació per interval | Per al valor esperat<br>$\hat{y}_h \pm t_{n-2,0.975} S \sqrt{(1/n + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2})}$ | Per a valors individuals<br>$\hat{y}_h \pm t_{n-2,0.975} S \sqrt{(1 + 1/n + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2})}$ |