

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

#### Problema B4

Abans de llençar al mercat la nova polsera *HyperPowerBalance*, un equip d'assessors comercials tracta de veure les oportunitats d'èxit del producte. La campanya publicitària ha divulgat les seves suposades virtuts, i es pretén calcular quina proporció de la població analitzada creu que pot ser eficaç.

**Preg. 1.** S'admet que aquesta proporció podria ser de fins a un 30%, en el cas més optimista. ¿Quina grandària hauria de tenir la mostra destinada a estimar la proporció real, si volem que el interval al 95% de confiança tingui una precisió del 3% (o una amplitud total de 6%)? [1pt.]

**Preg. 2.** Després, només es poden recollir 300 enquestes, de les quals 72 persones manifesten que sí creuen en l'eficàcia de la polsera. Amb aquestes dades, feu formalment la prova d'hipòtesis per comprovar si hi ha a la població una proporció del 30% com a possible *target* del producte. Considereu una prova unilateral, amb risc  $\alpha=5\%$ , i les fases habituals: [3pt.]

- Hipòtesis
- Estadístic de la prova
- Distribució de l'estadística anterior sota la hipòtesi nul·la
- Premisses
- Càlculs
- P-valor
- Conclusió i valoració del grau d'evidència present a la mostra
- Interval de confiança (no cal que sigui unilateral) i interpretació global.

**Preg. 3.** Posem que a la prova anterior s'ha obtingut un P-valor igual a  $1/100$ . Valoreu aquestes afirmacions, i reescriuiu -les, si és el cas. [1pt.]

- La probabilitat que la hipòtesi alternativa sigui certa és el P-valor =  $1/100$
- La probabilitat que una altra mostra de 300 persones obtingui 72 respostes o més favorables a la polsera és el P-valor =  $1/100$ .

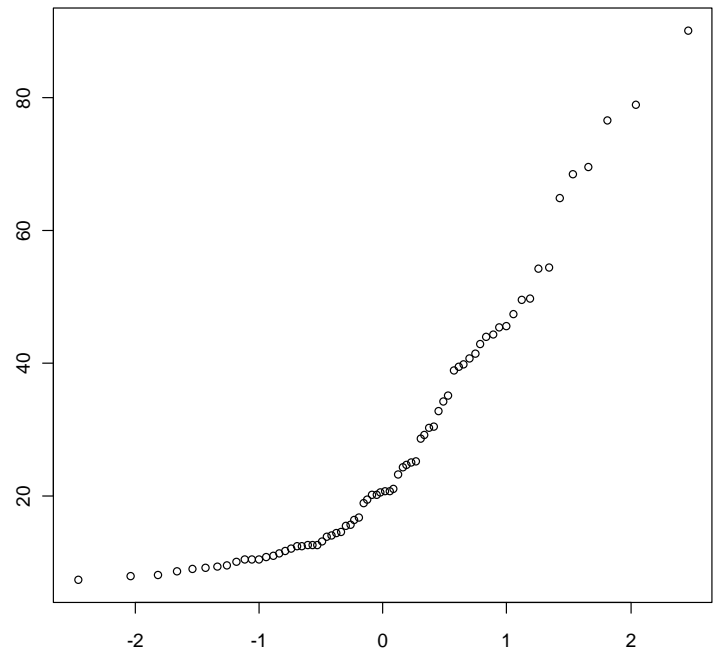
**Preg. 4.** Una altra pregunta a la enquesta demanava l'edat de la persona, i sabem que el interval per a la mitjana poblacional al 95% de confiança per al grup de dones que treballen fora de casa ( $n=27$ ) és (33.00, 38.54) anys. Es demana que trobeu el interval al 99% de confiança. [1pt.]

**Preg. 5..** Valoreu la següent afirmació, i corregiu-la si és el cas. [1pt.]

- Creiem que 99 de cada 100 dones que treballen fora de casa té una edat entre 32 i 39.5 anys, per a la població d'aquest estudi.

**Preg. 6.** A l'enquesta, una de les preguntes era "quin creus que serà el preu que costarà la nova *HyperPowerBalance*". L'empresa tenia la idea de vendre-les a 40 euros, i aquesta qüestió tenia el propòsit d'esbrinar si les expectatives de la població estaven molt allunyades de la proposta de preu.

a) A la dreta teniu el  $qqnorm()$  del preu (considerant únicament les 72 persones que van respondre que podia ser eficaç): què opineu sobre aquesta premissa? b) Ara imaginem que "es sap" que la distribució és normal. El preu mitjà a la mostra de 72 respostes és de 27.70€, i la desviació tipus mostral de 19.70€. Diríeu que en mitjana la gent creu que la polsera hauria de costar menys de 40€? Feu una prova d'hipòtesis formal (resumida), i exposeu la conclusió. [2pt.]



**Preg 7.** Hauríeu aconsellat preguntar sobre el cost de la polsera a les 300 persones enquestades, en lloc de només a aquest grup de 72? Doneu arguments a favor i en contra (penseu sobre tot en arguments estadístics). [1pt.]

**NOM:** \_\_\_\_\_

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

**Problema B5**

- a) Uns estudiants de la FIB volen verificar si dos programes d'ordenació (X, Y) triguen el mateix. Mitjançant base de dades diferents, però de dimensions similars, executen 21 vegades cada programa, cadascun amb la base de dades corresponent.

Els resultats que s'han obtingut en programa són:

Variable	Mitjana	Variància	N
X	10.38	2.52	21
Y	10.95	1.90	21

- a.1 (0.5 punts) D'acord amb l'enunciat, es tracta d'un disseny "aparellat" o "independent"? Raoneu la resposta.

- a.2 (3 punts) Es pot admetre que la variàncies en ambdues casos és la mateixa? Per respondre aquesta pregunta, plantegeu una prova d'hipòtesi, indicant:

- Hipòtesis ( i indiqueu si la prova és bilateral o unilateral)
- Calculeu el valor de l'estadístic
- Digueu quina distribució segueix l'estadístic i quants graus de llibertat té.
- Representeu gràficament el valor obtingut i el punt crític sobre la distribució de l'estadístic
- Doneu la conclusió

- a.3 (3 punts) Es pot admetre que els dos programes triguen el mateix? Per respondre aquesta pregunta, plantegeu una prova d'hipòtesis, dient:

- Hipòtesis ( i indiqueu si la prova és bilateral o unilateral)
- Especifiqueu les premisses necessàries
- Independentment dels resultats de l'apartat a.2, trobeu l'estimació de la variància comú (*pooled*) dels programes X i Y
- Determineu el valor de l'estadístic, la seva distribució i quants graus de llibertat té.
- Tenen els programes rendiments diferents? Raoneu la vostra resposta.

- b) El professor d'aquests estudiants els ha aconsellat que potser és millor executar els programes en una sola base de dades i simultàniament.

Els resultats que s'han obtingut són

Variable	Mitjana	Variància	N
X-Y	-0.57	1.25	20

b.1 (2 punts) Es pot admetre que els dos programes triguen el mateix? Per respondre aquesta pregunta, plantegeu una prova d'hipòtesi que resoldeu mitjançant el càlcul del interval de confiança per a la diferència de rendiments.

Caldrà que:

- Indiqueu la hipòtesi i si la prova és bilateral o unilateral
- Calculeu el interval de confiança al 95% per a la diferència de mitjanes poblacionals
- A quin resultat arribeu?

b.2 (1.5 punt) Compareu els resultats obtinguts a l'apartat a.3 amb els de l'apartat b.1, és a dir, si es pot admetre en ambdós casos que les mitjanes poblacionals són les mateixes o no. Tant si es pot admetre com si no, indiqueu possibles explicacions.

**NOM:** \_\_\_\_\_

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

**Problema B6**

S'ha mesurat la el temps de descarrega de 8 fitxers amb les següents dades:

Grandària arxiu (MB) : X	Temps de descarrega (s) : Y
32	10.4
64	19.3
96	33.2
128	41.8
160	50.3
192	58.7
224	74.1
256	81.7

Amb aquestes dades tenim que:

$$n=8$$

$$\sum x_i = 1152$$

$$\sum x_i^2 = 208896$$

$$\sum y_i = 369.5$$

$$\sum y_i^2 = 21471.61$$

$$\sum x_i y_i = 66937.6$$

a) Calculeu els estimadors de la constant i del pendent de la recta de regressió i representeu gràficament la recta estimada **(1 punt)**

b) Calculeu la taula de descomposició de la variància **(2 punts)**

	SQ	Graus de llibertat (Gdl)	QM = SQ/Gdl	Rati
Explicada pel model				
Residual				
Total				

c) Calculeu i interpreteu el coeficient de determinació  $R^2$  **(1 punt)**

d) Calculeu l'interval de confiança al 95% pel pendent de la recta **(1 punt)**

e) Tenim contractat 3Mbps (22.5 MB/min). Poseu a prova si aquesta és la velocitat realment disponible (heu d'adaptar les dades per a treballar amb MB/min). Quina diferència té aquesta prova amb contrastar directament si el quocient  $X/Y$  té mitjana igual a 22.5? **(2 pts)**

*Nota: tingueu en compte que la transformació de les unitats afecta tant a l'estimador com a la seva desviació estàndard.*

f) Doneu una previsió, així com el seu interval de confiança al 95%, del valor promig del temps de descarrega quan la grandària del fitxer és de 30 MB **(2 punts)**

g) Dibuixeu (inventant les dades i les figures) els gràfics NPP (Normal probability plot) i 'residuals versus fitted' (residus en front de les prediccions) de tal manera que mostrin que les premisses **sí** que es compleixen i comenteu quines premisses permeten valorar cada un. **(1 punt)**