PE (FIB) -- EXAMEN FINAL (17 de juny de 2016)

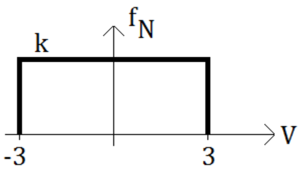
Cognoms, nom: ....

Problema B1_B2. Suposem un sistema digital de comunicacions i distingim, en ell, dues variables aleatòries: X és el símbol transmès i Y és el símbol rebut. Tant X com Y només poden prendre els valors 0 i 1. A nivell físic, el símbol a transmetre 0 es modelarà com una tensió nul·la i el símbol a transmetre 1 es modelarà com una tensió de $2\,V$. En el procés de transmissió, a la tensió transmesa se li suma un soroll aleatori N, soroll (v.a. contínua) que està uniformement distribuït entre -3 i 3. A recepció, si la tensió rebuda supera el llindar $L = 1\,V$ direm que s'ha rebut un 1 i si la tensió rebuda és inferior o igual al llindar $L = 1\,V$ direm que s'ha rebut un 0. Suposem que la probabilitat d'emetre un 0 (p_0) és doble de la probabilitat d'emetre un 1 (p_1)

(i) Calculeu els valors de p_0 i de p_1

Si $p_0 = 2p_1$ i $p_0 + p_1 = 1$, llavors ha de ser $p_0 = \frac{2}{3}$ i $p_1 = \frac{1}{3}$

(ii) Calculeu el valor que ha de prendre la k del dibuix de sota per a que f_N sigui una funció de densitat de probabilitat



Per a què f_N sigui una funció de densitat de probabilitat, l'àrea que tanca ha de ser unitària, és a dir: $6k = 1$ ó $k = 1/6$.

(iii) Calculeu les quatre probabilitats següents:

a. Probabilitat de de que s'enviï un 0 i es rebi un 0

b. Probabilitat de de que s'enviï un 0 i es rebi un 1

c. Probabilitat de de que s'enviï un 1 i es rebi un 0

d. Probabilitat de de que s'enviï un 1 i es rebi un 1

a. $P(X = 0 \wedge Y = 0) = P(X = 0\,) \cdot P(Y = 0 \mid X = 0) = P(X = 0\,) \cdot P(N \leq 1) = \frac{2}{3} \cdot \int_{-3}^{1} \frac{1}{6} dx = \frac{4}{9}$

b. $P(X = 0 \wedge Y = 1) = P(X = 0\,) \cdot P(Y = 1 \mid X = 0) = P(X = 0\,) \cdot P(N > 1) = \frac{2}{3} \cdot \int_{1}^{3} \frac{1}{6} dx = \frac{2}{9}$

c. $P(X = 1 \wedge Y = 0) = P(X = 1\,) \cdot P(Y = 0 \mid X = 1) = P(X = 1\,) \cdot P(2 + N \leq 1) = p_1 \cdot \int_{-3}^{-1} \frac{1}{6} dx = \frac{1}{9}$

d. $P(X = 1 \wedge Y = 1) = P(X = 1\,) \cdot P(Y = 1 \mid X = 1) = P(X = 1\,) \cdot P(2 + N > 1) = p_1 \cdot \int_{-1}^{3} \frac{1}{6} dx = \frac{2}{9}$

(iv) Sabent que s'ha enviat un 0, calculeu la probabilitat de que es rebi un 0 i la probabilitat de que es rebi un 1.

$$P(Y=0 \mid X=0) = \frac{P(Y=0 \wedge X=0)}{P(X=0)} = \frac{P(X=0 \wedge Y=0)}{p_0} = \frac{\frac{2}{3}p_0}{p_0} = \frac{2}{3}$$

$$P(Y=1 \mid X=0) = 1 - \frac{2}{3} = \frac{1}{3}$$

(v) Sabent que s'ha enviat un 1, calculeu la probabilitat de que es rebi un 0 i la probabilitat de que es rebi un 1.

$$P(Y=0 \mid X=1) = \frac{P(Y=0 \wedge X=1)}{P(X=1)} = \frac{P(X=1 \wedge Y=0)}{1-p_0} = \frac{\frac{1}{3}(1-p_0)}{1-p_0} = \frac{1}{3}$$

$$P(Y=1 \mid X=1) = 1 - \frac{1}{3} = \frac{2}{3}$$

(vi) Calculeu la probabilitat de que es rebi un 0. Calculeu la probabilitat de que es rebi un 1.

$$P(Y=0) = P(X=0) \cdot P(Y=0 \mid X=0) \; + \; P(X=1) \cdot P(Y=0 \mid X=1)$$

$$P(Y=0) = p_0 \cdot \frac{2}{3} + p_1 \cdot \frac{1}{3} = \frac{5}{9}$$

$$P(Y=1) = 1 - P(Y=0) = 1 - \frac{5}{9} = \frac{4}{9}$$

(vii) Sabent que s'ha rebut un 1, calculeu: probabilitat de que s'hagi enviat un 0 i probabilitat de que s'hagi enviat un 1.

$$P(X=0 \mid Y=1) = \frac{P(X=0 \wedge Y=1)}{P(Y=1)} = \frac{\frac{2}{9}}{\frac{4}{9}} = \frac{1}{2}$$

$$P(X=1 \mid Y=1) = 1 - \frac{1}{2} = \frac{1}{2}$$

Puntuació: apartats (i) i (ii): 1 punt cada un

apartats (iii): 2 punts

apartats (iv), (v), (vi) i (vii): 1.5 punts cada un

NOM:____________________________COGNOM:____

(***Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs***)

Problema 2 (B3 B4)

En una granja d'ous de gallines amb certificat de tractament ecològic asseguren que la distribució del pes (en grams) dels ous segueix el model normal amb esperança 65 gr i desviació 4 gr.

- (1 punt) Calculeu la probabilitat que el pes d'un ou estigui entre 60 gr i 70 gr

P(60 < Pes < 70) = P(Pes<70) – P(Pes<60) = P(Z<(70-65)/4) – P(Z<(60-65)/4) = P(Z<1.25) – P(Z<-1.25) =
= P(Z<1.25) – (1-P(Z<1.25)) = 2 P(Z<1.25) -1 = 2 (0.8944) -1 = 0.7888

- (1 punt) Calculeu la probabilitat que el pes d'un ou sigui superior a 64 gr

P(Pes > 64) = P(Z>(64-65)/4) = P(Z>-0.25) = P(Z<=0.25) = 0.5987 = 0.60

- (1 punt) Considerant una dotzena d'ous (12 unitats), definiu la variable nombre d'ous que superen els 64 grams, i calculeu la probabilitat que 6 o més superin aquest pes.

N és Bin(n=12, p=0.6) i NN és Bin(n=12,p=0.4)

P(N>=6) = P(NN<=5) = 0.6652

- (1 punt) Considerant un cartró d'ous (30 unitats), definiu la variable pes total dels ous, i calculeu la probabilitat que aquest pes total sigui inferior a 2 Kg.

PesT és Normal amb esperança 1950 gr (65*30) i desviació 21.91 gr (4*sqrt(30))

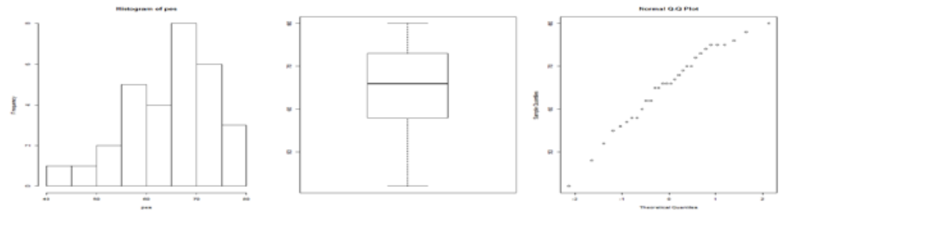
P(PesT<2000) = P(Z<(2000-1950)/21.91) = P(Z<2.28) = 0.9887

- (1 punt) Considerant un cartró d'ous (30 unitats), definiu la variable pes mitjà dels ous, i calculeu la probabilitat que aquest pes mitjà sigui inferior o igual a 64 grams.

PesM és Normal amb esperança 65 gr i desviació 0.73 gr (4/sqrt(30))

P(PesT<=64) = P(Z<=(64-65)/0.73) = P(Z<=-1.37) = 1 – P(Z<=1.37) = 1 - 0.9147 = 0.0853

Ara volem inferir resultats a partir de dades mostrals. Recollim els pes d'un cartró d'ous (30 unitats) i obtenim els següents resultats: 42, 48, 52, 55, 56, 57, 58, 58, 60, 62, 62, 65, 65, 66, 66, 66, 67, 68, 69, 70, 70, 72, 73, 74, 75, 75, 75, 76, 78, 80



$$\sum_{i=1}^{30} P_i = 1960 \qquad \sum_{i=1}^{30} P_i^2 = 130514$$

- (1 punt) Calculeu una estimació puntual de la mitjana, de la desviació, de l'error estàndard de l'estimador de la mitjana poblacional, i de la proporció d'ous amb pes superior a 64 gr

mean 65.33333
sd 9.211437
sd/sqrt(30) 1.681771
19/30 0.6333333

- (2 punts) Plantegeu i resoleu la prova d'hipòtesi sobre si el valor esperat de la mitjana és 65 gr o no, amb risc 5%

H_0: $\mu_P = 65$ Bilateral
H_1: $\mu_P <> 65$

$$t = \frac{\bar{P}-65}{s/\sqrt{n}} \sim t_{29}$$ Premisses: m.a.s i V Normal

$$t = \frac{65.33-65}{9.21/\sqrt{30}} = \frac{0.33}{1.68} = \mathbf{0.196}$$

punts crítics ($t_{29,0.975}$) -2.045 i +2.045

No hi ha evidència per rebutjar la hipòtesi nul·la (l'estadístic 0.196 està dins dels punts crítics). Podem afirmar que el pes esperat és 65 gr

```
t.test(pes,mu=65)
        One Sample t-test
data:  pes
t = 0.1982, df = 29, p-value = 0.8443
alternative hypothesis: true mean is not equal to 65
95 percent confidence interval:
 61.89373 68.77294
sample estimates:
mean of x
 65.33333
```

- (2 punts) Calculeu un IC amb confiança 95% de la proporció esperada d'ous amb pes superior a 64gr

$$Emprant\ estimació \rightarrow IC(\pi, 95\%) = p \pm z_{0.975} \cdot \sqrt{p \cdot \frac{1-p}{n}} = 0.63 \pm 1.96 \cdot \sqrt{\frac{0.63 \cdot 0.37}{30}} = 0.63 \pm 0.088 = [0.54\,, 0.72]$$

(o bé $Màxima\ indeterminació \rightarrow IC(\pi, 95\%) = p \pm z_{0.975} \cdot \sqrt{0.5 \cdot \frac{0.5}{n}} = 0.63 \pm 1.96 \cdot \sqrt{\frac{0.5 \cdot 0.5}{30}} = 0.63 \pm 0.18 = [0.42\,, 0.81]$)

```
binom.test(19,30)
        Exact binomial test
data:  19 and 30
number of successes = 19, number of trials = 30, p-value = 0.2005
alternative hypothesis: true probability of success is not equal to 0.5
95 percent confidence interval:
 0.4385598 0.8007014
sample estimates:
probability of success
              0.6333333
```

NOM:___________________________COGNOM:____

(***Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs***)

Problema 3 (B5 B6)

La pràctica d'uns alumnes consisteix en mesurar el consum de bateria que fa el GPS d'un mòbil. Dissenyen un experiment consistent en reproduir un vídeo d'una hora de duració, amb i sense el GPS activat, i amb la bateria completament carregada inicialment. Aquestes són les dades obtingudes (percentatge arrodonit de càrrega al final):

									mean	sd
Amb GPS	76	66	70	71	75	70	75	79	72.75	4.200
Sense GPS	76	79	74	74	78	78	79	75	76.625	2.134

1. Expliqueu què condicions farien que aquest experiment fos un disseny aparellat, i quines un disseny de mostres independents. Poseu exemples concrets. [1pt]

Aparellat: les observacions s'han obtingut de dos en dos, amb i sense GPS, per causa de condicions comunes; per exemple, s'han fet les proves una darrera l'altra (recarregant la bateria al mig), deixant l'estat del mòbil igual. Abans de cada nou parell de proves seria convenient deixar passar un llarg temps o reinicialitzar el mòbil. Independents: el tipus de prova (amb o sense) s'ha determinat a l'atzar a cada observació; l'estat del mòbil no pot dependre de l'ús anterior, així que potser es tindria que reinicialitzar cada vegada abans de recarregar la bateria.

2. Suposem que els dos grups a comparar són independents. Exposeu quines són les premisses de la prova de comparació de mitjanes que es vol fer. [0.5pt]

Cada mostra ha de ser una mostra aleatòria simple; els dos grups han de ser independents; la variable resposta té distribució Normal; la variància de la resposta és comuna per als dos grups.

3. Quina és la resposta que cal utilitzar: a) la càrrega de la bateria al final, b) el decrement de càrrega, c) si la bateria s'ha descarregat més d'un 25% o menys? Raoneu la resposta. [0.5pt]

Clarament l'opció c) no és la millor perquè el llindar del 25% és arbitrari i en tot cas una prova basada en una resposta dicotòmica tindria poca potència. Qualsevol de les altres dues opcions pot servir, però la b) representa una resposta lligada directament a l'objectiu del experiment i és més fàcil d'interpretar.

4. Resoleu la prova d'hipòtesis amb un enfoc unilateral amb risc α=5% i expliqueu com es trobaria (sense calcular-ho) el p-valor. [1.5pt]

Utilitzarem el decrement D = 100-Y. Les desviacions tipus són les mateixes, i les mitjanes són respectivament 27.25% i 23.375%. La hipòtesi alternativa és que el decrement esperat és major amb GPS que sense GPS.

H_0: $\mu_{amb} = \mu_{sense}$
H_1: $\mu_{amb} > \mu_{sense}$

Sota H_0, $t = \frac{\overline{d_{amb}} - \overline{d_{sense}}}{s\sqrt{\frac{1}{n}+\frac{1}{n}}} \sim t_{2n-2}$

L'estimació comuna de la desviació tipus: $s^2 = \frac{7\ 4.2^2 + 7\ 2.134^2}{14} = 11.097; s = 3.331$

L'estadístic de la prova amb les dades mostrals: $\hat{t} = \frac{27.25-23.375}{3.331\sqrt{\frac{1}{8}+\frac{1}{8}}} = 2.326$

El punt crític unilateral d'una distribució t-student amb 14 graus de llibertat: 1.761. Per tant, es veu que la hipòtesi nul·la es pot rebutjar. El p-valor seria P(t > 2.326): amb una N(0,1) valdria 0.01 (el valor correcte amb la t_{14} és 0.0178)

5. Estimeu amb un IC 95% bilateral el consum mitjà que suposa l'ús del GPS. [1pt]

IC(95%, $\mu_{amb} - \mu_{sense}$) = 27.25−23.375 ± $t_{14,\ 0.975}$ 3.331 $\sqrt{\frac{1}{8}+\frac{1}{8}}$ = 3.875 ± 2.145 1.57025 = [0.3024, 7.448]

6. Quina és la conclusió pràctica dels resultats anteriors? [1pt]

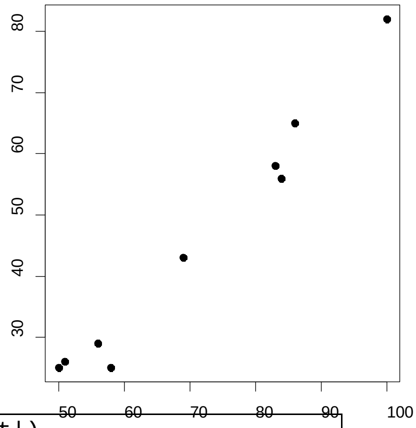
hem trobat evidències de que utilitzar el GPS durant una hora mentre el mòbil fa altres tasques incrementa el consum de bateria, entre 0.30% i 7.45% de mitjana, amb 95% de confiança.

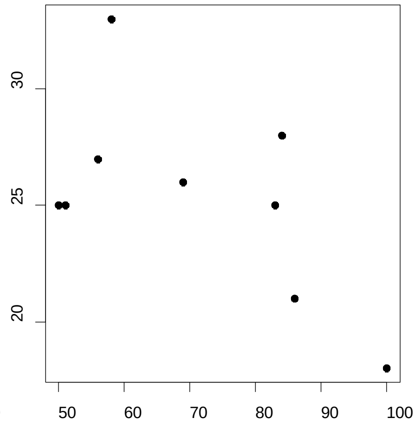
7. Per tal d'obtenir una estimació molt més precisa de l'efecte del GPS (l'objectiu és un IC 95% amb error ±1%), es planteja un projecte col·laboratiu on voluntaris de tot el mon aportarien dades. Com que es creu que la heterogeneïtat seria important, es preveu una desviació tipus poblacional del 6%. Quants voluntaris farien falta per tenir les dades, suposant que volem tantes dades amb GPS com sense? [1pt]

$1.96\ 6\sqrt{\frac{1}{n}+\frac{1}{n}} = 1; n = 277$ per grup, llavors 554 voluntaris

Continuant amb la mesura de la bateria del mòbil, es vol comprovar si la bateria es descarrega més ràpidament si la bateria està a mitja càrrega que quan està al 100% (estem suposant que el mesurador de càrrega és completament fiable). Per aquest objectiu, es repeteix l'experiència de reproduir el vídeo d'una hora (aquesta vegada sense GPS) partint de diferents nivells de càrrega de la bateria, i el resultat és el següent:

```
a l'inici:  50  51  56  58  69  83  84  86 100
al final:   25  26  29  25  43  56  54  65  82
Descens     25  25  27  33  26  27  30  21  18
```





```
              Estimate Std. Error  t value  Pr(>|t|)
(Intercept) -35.04958    5.17759   _______________
X             1.13728    0.07112   _______________

Residual standard error: 3.639 on 7 degrees of freedom
Multiple R-squared:  0.9734,    Adjusted R-squared:  0.9696
```

```
            Estimate Std. Error  t value  Pr(>|t|)
(Intercept) 35.04958    5.17759   _______________
X           -0.13728    0.07112   _______________

Residual standard error: 3.639 on 7 degrees of freedom
Multiple R-squared:  0.3474,    Adjusted R-squared:  0.2542
```

8. Expliqueu i interpreteu sintèticament els dos models ajustats a sobre. Teniu en compte: a) quines són les variables implicades, b) quin model hem trobat, c) com s'interpreten els estimadors, d) quina és la significació estadística dels resultats, e) com és la capacitat predictiva dels models. Finalment, digueu com podeu respondre a les preguntes objectiu d'aquesta part: la bateria es descarrega més lenta o més ràpidament, independentment del nivell de càrrega inicial? És convenient assegurar-se de que partim d'una bateria completament carregada? [3pt]

<table>
<tr><th>Model 1</th><th>Model 2</th></tr>
<tr><td>Variable explicativa: nivell inicial de càrrega (X)
Variable resposta: nivell final de càrrega (Y) després de la reproducció del vídeo
Y = –35.050 + 1.137 X
El terme lineal és quant augmentaria la càrrega final si a l'inici tinguéssim un 1% més
El terme independent és un corrector, que es podria interpretar com el consum mitjà (en valor absolut) de la bateria si el pendent fos exactament 1.</td><td>Variable explicativa: nivell inicial de càrrega (X)
Variable resposta: descens de càrrega (D) després de la reproducció del vídeo
D = 35.050 – 0.137 X
El terme lineal és com es modifica el decrement de càrrega si a l'inici tinguéssim un 1% més
El terme independent és un corrector, que es podria interpretar com el consum mitjà de la bateria si el pendent fos exactament 0.</td></tr>
<tr><td colspan="2">Els dos models són equivalents, donat que Y = X – D. Llavors, si el primer model és $Y = \beta_0 + \beta_1 X + \varepsilon$ el segon seria $D = -\beta_0 + (1-\beta_1) X + \varepsilon$, i les estimacions també estarien relacionades de la mateixa forma (com es pot veure als resultats). Possiblement el segon model és més clar per determinar si la descàrrega depèn de com està la bateria a l'inici.</td></tr>
<tr><td>L'estimació de la desviació residual és 3.679
El terme lineal és significativament diferent de 0: el t value seria 1.13728 / 0.07112 = 15.992, però no seria significativament diferent de 1: (1.13728 – 1)/ 0.07112 = 1.930
La capacitat predictiva és molt alta, R^2 = 97.34%, encara</td><td>L'estimació de la desviació residual és 3.679
El terme lineal és no significativament diferent de 0: el t value seria –0.13728 / 0.07112 = –1.930. El punt crític is ±2.3646.
La capacitat predictiva és baixa, R^2 = 34.74%; la càrrega a l'inici no és un bon predictor de la descàrrega de la</td></tr>
</table>

que no és gens estrany: les càrregues a l'inici i al final estan molt relacionades, com és normal.	bateria. De totes formes, el resultat és prop de la significació estadística i és negatiu, que apunta que la bateria es descarrega més si al principi no està plena.
En conclusió: el millor model per a respondre és el segon, perquè la resposta és directament el decrement de càrrega, però no hem trobat evidències forts de que aquest decrement està associat a la càrrega a l'inici. L'estimació de l'efecte (95% confiança) per a una variació d' 1% a la X és (-0.305 , 0.0309).	

9. Assumint que el model és útil per fer previsions, trobeu amb IC 95% una estimació de quin seria el nivell de càrrega final, després d'una prova de reproducció del vídeo si inicialment la bateria estigués al 100%. Ajut: $\Sigma X_i = 637$; $s^2_X = 327.1944$ [0.5pt]

error tipus de la previsió: $s\sqrt{1+\frac{1}{n}+\frac{(x_h-\bar{x})^2}{(n-1)s_X^2}} = 3.639\sqrt{1+\frac{1}{9}+\frac{(100-637/9)^2}{8\ 327.1944}} = 4.36277$

$IC(y_{100}, 95\%) = (-35.04958 + 1.13728\cdot 100) \pm 2.3646 \cdot 4.36277 = (68.362, 88.995)\%$