NOM:________________________COGNOM:_____________________________________

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

Problema A

1. Una multinacional de components informàtics té dues fàbriques diferents per produir un determinat model de pantalla d'ordinador: la fàbrica A on es produeixen el 30% de les pantalles i la fàbrica B on se n'elaboren el 70% restant. En la fàbrica A s'utilitza un procediment més laboriós que produeix un 99% d'unitats satisfactòries. La fàbrica B empra un altre procediment menys costós, però que comporta un 10% d'unitats defectuoses. Després les pantalles de les dues fàbriques s'envien al centre logístic de distribució.
   S'escull a l'atzar una pantalla en el centre logístic de distribució.
   a) Quina és la probabilitat que sigui defectuosa? (1 punt)

   P(Defectuós) = P(Defectuós| A) · P (A) + P(Defectuós| B) · P(B) = 0'01 · 0'3 + 0'1· 0'7 = 0.073

   b) Si la pantalla és satisfactòria, quina és la probabilitat que sigui de la fàbrica B? (1 punt)

   P(B | ¬ Defectuós) = P(B ∩ ¬ Defectuós) / P(¬ Defectuós) = P(¬ Defectuós| B) · P(B) / P(¬ Defectuós) = 0.9 · 0.7 / (1-0.073) = 0.6796

2. El sistema de postvenda rep les pantalles defectuoses i les classifica diàriament segons la fàbrica que les ha produït. Així, el nombre de pantalles procedents de la fàbrica A es recull a la variable aleatòria X1, i per a la fàbrica B es té la v.a. X2

x	$P_{X1}(x)$	$P_{X2}(x)$
0	0.2	0.1
1	0.25	0.2
2	0.3	0.4
3	0.25	0.3

   a) Quina és la probabilitat que en un dia s'hagin rebut menys de 2 pantalles defectuoses de la fàbrica A? (1 punt)

   $P(X1<2) = P(X1=0)+P(X1=1)= 0.2 + 0.25 = 0.45$

   b) Calculeu la taula de probabilitats conjunta de X1 i X2. (1 punt)

   Com que els esdeveniments són independents, es té que per calcular les probabilitats conjuntes P(X1=i, X2=j) = P(X1=i)·P(X2=j). Per tant, la taula quedarà:

X1 \ X2	0	1	2	3
0	0.02	0.04	0.08	0.06
1	0.025	0.05	0.1	0.075
2	0.03	0.06	0.12	0.09
3	0.025	0.05	0.1	0.075

   c) Calculeu la probabilitat que el nombre de pantalles rebudes de la fàbrica A sigui superior a les rebudes de la fàbrica B. (1 punt)

   P(X1> X2) = P(X1 =1, X2 = 0) + P(X1 =2, X2 = 0) + P(X1=3, X2 = 0) + P(X1 =2, X2 = 1) + P(X1=3, X2 = 1) + P(X1=3, X2 = 2) = 0.025 + 0.03 + 0.025 +0.06 + 0.05 + 0.1 = 0.29

d) Calculeu l'esperança d'X1 (0'5 punts)

E(X1)= 0 ·0'2 + 1·0'25 + 2·0'3 + 3·0'25 = 1'6

e) Calculeu la variància i la desviació típica de X2. (1 p)

E(X2)= 0 ·0'1 + 1·0'2 + 2·0'4 + 3· 0'3 = 1'9

E(X2^2)= 0^2 ·0'1 + 1^2·0'2 + 2^2·0'4 + 3^2· 0'3 = 4'5

Var(X2) = E(X2^2) - E(X2)^2 = 4'5 - 1.9^2 = 0'89

σ_{X2} = 0'9434

3. Es considera la variable aleatòria contínua, T, amb la següent funció de densitat:

$$f(t) = \begin{cases} \frac{1+t^2}{a}, & 0 < t < 6 \\ 0 & , \ altrament \end{cases}$$

3a) Calculeu *a* perquè f(t) sigui una funció densitat (2 p)

S'ha de complir que $\int_{-\infty}^{+\infty} f(t)dt = 1$.

En aquest cas, tenim que $\int_0^6 \frac{1+t^2}{a} dt = [\frac{t}{a} + \frac{t^3}{3a}]_0^6 = \frac{234}{3a} = 1$

Per tant, a = 78

Per aquest valor d'a, la funció f(t) assoleix sempre valors positius

3b) Calculeu la funció de distribució de T (1'5 p)

Calculem $\int_0^t \frac{1+z^2}{78} dz = [\frac{z}{78} + \frac{z^3}{234}]_0^t = \frac{t^3+3t}{234}$

$$F_T(t) = \begin{cases} 0, & t < 0 \\ \frac{t^3+3t}{234}, & 0 < t < 6 \\ 1 & , \ t > 6 \end{cases}$$

BB NOM:____________________________COGNOMS:____

Sigueu concisos i feu lletra llegible. ***Contesteu cada pregunta en el seu lloc.*** ***Expliciteu i justifiqueu els càlculs.***

Amb l'objectiu de reduir la congestió d'una ciutat, un ajuntament està considerant posar a prova 40 nous semàfors amb sensors de capacitat de manera que segons el volum de vehicles que hi ha a la cua el semàfor roman més o menys temps de color verd. Es sap que, amb aquesta nova tecnologia, el 5% dels nous semàfors (sensors) fallen.

1.a. Quina és la distribució de probabilitat que s'adequa al fenomen "Nombre de semàfors/sensors que fallen" **(0.5 punts)**
N v.a. "nombre de semàfors/sensors que fallen" ~ Bin(40. 0.05)

1.b. Quin és el nombre esperat de semàfors que fallen? **(0.25 punts)**
E(N) = n * p = 40 * 0.05 = 2

1.c. Quina és la desviació tipus del nombre de semàfors que fallen? **(0.25 punts)**
Var(N) = n*p*q = 40 * 0.05 * 0.95 = 1.9

Sigma(N) = sqrt(1.9) = 1.38

2. Quina és la probabilitat que falli algun semàfor? **(1 punt)**
P(N > 0) = 1 - P(N<=0) = 1 - (P(N = 0) =

```
> pbinom(0, 40, 0.05)
[1] 0.1285122
```

3. En la prova pilot instal·len els primers 10 semàfors d'un en un. Quina és la probabilitat que funcionin els nou primers i falli el desè? **(1 punt)**
X v.a. "Nombre d'intents fins a fallar" ~ Geomètrica(0.05)

P(X = 10)= q^(k-1)*p = 0.95^9*0.05 = 0.03

```
> dgeom(9.0.05) #En R, nombre de intents abans no falli
[1] 0.03151247
```

4. Suposem que el temps d'espera al semàfor quan està vermell és una variable aleatòria amb probabilitat constant en l'interval entre 0 i 2 minuts.

4.a. Quin és el model de probabilitat que s'adequa a aquesta situació? **(0.5 punts)**
U v.a. "Temps espera al semàfor vermell" ~ Uniforme(0,2)

4.b. Quina és el temps mitjà d'espera? **(0.25 punts)**
E(U) = (b+a)/2 = 2/2 = 1 minut

4.c. Quina és la probabilitat d'haver d'esperar menys de 30 segons? **(0.25 punts)**
30 segons = 0.5 min

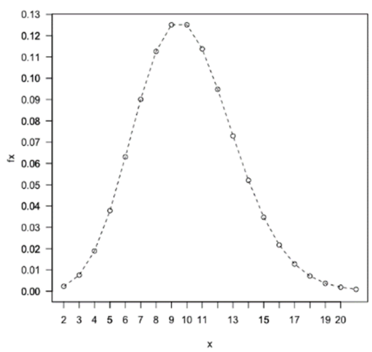
F(u) = (u-a)/(b-a) = (0.5-0)/(2-0) = 0.25

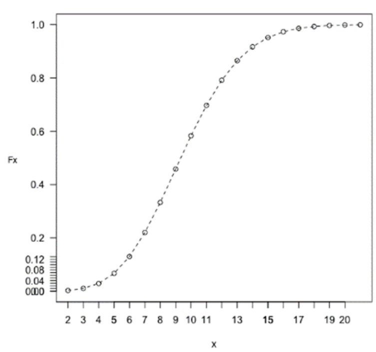
5. A un dels semàfors que podrien reduir la congestió arriben, en un dia laborable, 10 vehicles per minut.

5.a. Quin és el model de probabilitat que s'adequa a aquesta situació? **(0.5 punts)**

Y v.a. "Nombre de vehicles que arriben al semàfor per minut" ~ Poiss(10)

5.b. Aproximadament, quina és la probabilitat que arribin 6 o menys vehicles en un minut (fes servir la funció de distribució del gràfic adjunt)? **(0.5 punts)**





P(Y <= 6) ~ 0.12 (mirant el gràfic de la funció de distribució)

```
> ppois(6,10)

[1] 0.1301414
```

6. Considerem ara el temps que passa (en minuts) entre l'arribada de dos vehicles consecutius a un semàfor. Recorda que arriben 10 vehicles per minut.

6.a. Quin és el model de probabilitat que modelitza aquest fenomen? **(0.5 punts)**

T v.a. “temps d’arribada entre vehicles consecutius” ~ exp(10)

6.b.Indica quin és el temps esperat d’arribades (en segons) entre vehicles **(0.25 punts)**

E(T) = 1/lambda = 1/10 = 0.1 min => 0.1 min = 6 segons

6.c. Fa mig minut que no arriba cap vehicle al semàfor. Quina és la probabilitat que arribi un vehicle en els propers 15 segons? **(0.25 punts)**

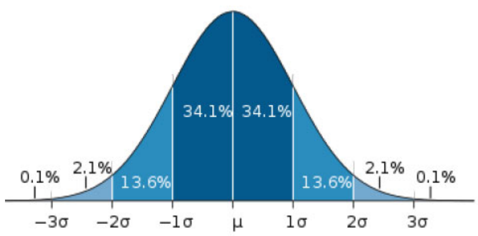
P(T < 0.75 | T > 0.5) = P(T < 0.25) = P(T <= 0.25) = 1 - exp(-10 * 0.25) = 0.92

(15 seg. = 0.25 min.)

> pexp(0.25, rate=10)

[1] 0.917915

7. En un dia festiu on la intensitat de trànsit augmenta, s’ha observat que el nombre de vehicles que arriben per minut en aquest semàfor es multiplica per 10 (recorda que en un dia laborable arriben 10 vehicles per minut). Quina és la probabilitat que arribin més de 110 vehicles en un minut (fes servir el gràfic)? **(1 punt)**



Aplicació del TCL (per lambda gran): Y1 v.a. “nombre de vehicles que arriben al semàfor per minut” ~ N(mu = 10*10, sigma = sqrt(10*10)) ~ N(100, 10)

P(Y1 > 110) = 1 - P(Y1 <= 110) = P(Z <= -10/10) = P(Z <= -1) = 13.6% + 2.1% + 0.1% = 15.8% (o 0.158)

> pnorm(-1)

[1] 0.1586553

Com a referència: 1 - ppois(110,100) = 0.1471373

8. (fent servir el gràfic) Quina és la probabilitat que arribin entre 80 i 120 vehicles per minut? **(1 punt)**

P(80 <= Y1 <= 120) = P(Y1 <= 120) - P(Y1 <= 80) = P(Z <= 2) - P(Z <= -2) = (0.136 + 0.341 + 0.341 + 0.136 + 0.021 + 0.01) – (0.021 + 0.01) = 0.954

> pnorm(2) - pnorm(-2)

[1] 0.9544997

9. Es vol calcular probabilitats sobre el nombre de vehicles que arriben als 40 semàfors en un dia festiu. Suposem que les arribades de vehicles als semàfors són independents i idènticament distribuïdes.

9.a. Indica la distribució de la variable aleatòria que ens permet estudiar aquest fenomen. **(0.5 punts)**

Aplicant el TCL,

S_{40} ~ N(40 * 100, 10*sqrt(40)) = N(4000, 63.25)

9.b. Indica quina seria la distribució de la variable aleatòria que ens permetria estudiar la mitjana de vehicles que arriben als 40 semàfors. **(0.5 punts)**

Aplicant el TCL,

M_{40} ~ N(100, 10/sqrt(40)) = N(100, 1.58)

10. Aproximadament, quina és la probabilitat que arribin més de 4100 vehicles als 40 semàfors? **(1 punt)**

P(S > 4100) = 1 – P(S <= 4100) = 1 – P(Z <= 100/63.25) =

= 1 – P(Z <= 1.58) =1 – [(P(Z<=1.56) + P(Z<=1.60))/2] =

= 1 – (0.9406 + 0.9452)/2 = 1 – 0.9429 = 0.0571

> pnorm(1.58)

[1] 0.9429

pnorm(1.52)=	0.9357
pnorm(1.54)=	0.9382
pnorm(1.56)=	0.9406
pnorm(1.60)=	0.9452
pnorm(1.62)=	0.9474
pnorm(1.64)=	0.9495

BC Final QT22 MODEL de resposta: imprescindible (>=0.8) i opcional (per aclarir).

Sigueu **concisos** *i feu lletra* **llegible**. *Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs. Cada apartat, 1 punt*

Volem comparar la eficiència en temps dels algorismes de multiplicació Karatsuba (K) i eStàndard (S). En un mateix ordinador, hem generat aleatòriament 18 parells de nombres de 900 bits i hem calculat el temps de multiplicar cada parell de nombres amb K i amb S.

1. Justifica que es tracta d'un disseny aparellat. Quina part de l'enunciat ens ho diu?

És aparellat perquè cada unitat experimental, formada per cada parell de nombres, es multiplicada per K i per S: tindrem 36 multiplicacions distribuïdes en 18 parelles.

Ens ho diu la part final ("hem calculat el temps de multiplicar cada parell de nombres amb K i amb S")

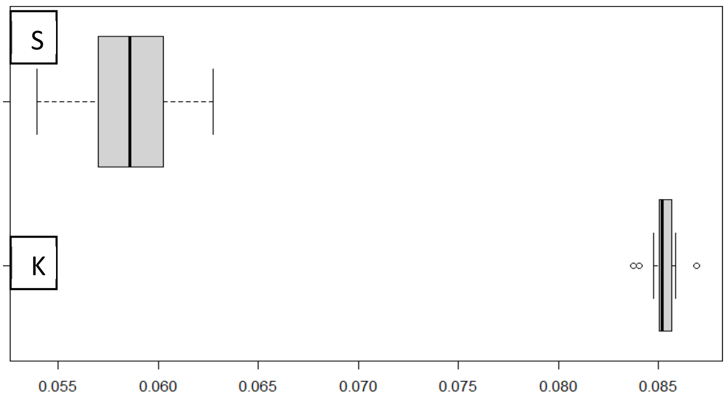
2. Com es podria valorar el grau d'aparellament assolit?

Mirant el gràfic K vs S i la correlació entre aquestes dos variables.

Si aquesta correlació fos cero i el gràfic no mostrés desviaments de la linealitat, podrien dir que les variables són independents. Cas contrari tindrien el grau d'aparellament.

Si el enunciat ens proporciona, p.e., les desviacions tipus de K, de E i de K-E, podrien deduir la covariància i la correlació.

3. Interpreteu aquest gràfic descriptiu on les abscisses són el temps en segons. Proposaríeu transformar les variables amb logaritmes? Perquè?

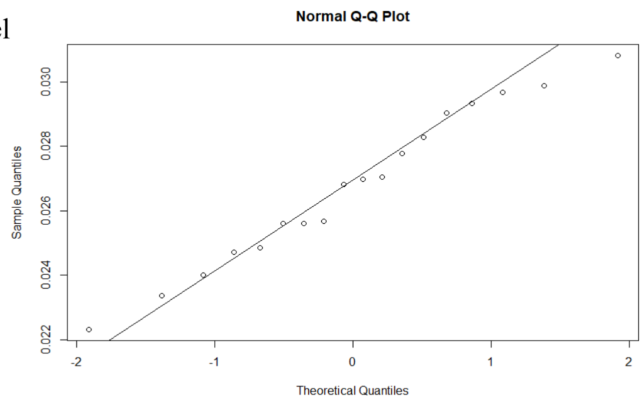


Clarament S és més ràpid: la seva mitjana està al voltant de 0.059s, per 0.086s la de K.

Els box-plots mostren tots dos distribucions simètriques, amb una dispersió (distància inter-quartil) més gran per a S.

No transformaria les variables amb logaritme: no tindria gairebé cap efecte en la forma de la distribució. En tot cas podria ser útil si la variància més gran la tingués K (que també té la mitjana més gran), però no és el cas.

4. Interpreteu el diagrama quantil-quantil de les diferències K−S del temps.



Recolza la premissa de normalitat.

Els punts mostrals (ordenades) i els teòrics si la distribució fos normal (abscisses) tenen una bona correspondència i estan raonablement alienats en la recta.

La mitjana i error tipus (*et*) de les diferències valen 0.02678 i 0.00057, respectivament.

5. Interpreteu.

S guanya a K per 0.0268s (27 mil·lèsimes), si bé el procés aleatori de generació de les dades conté una incertesa esperada de 0.006 (6 mil·lèsimes) en aquesta mitjana

6. Calculeu la desviació estàndard S (o tipus) de la diferència.

$et = S/\sqrt{n} \rightarrow S = 0.00057 \cdot \sqrt{18} \approx 0.00242$

7. Quina diferència hi ha entre la desviació i l'error tipus: què mesuren?

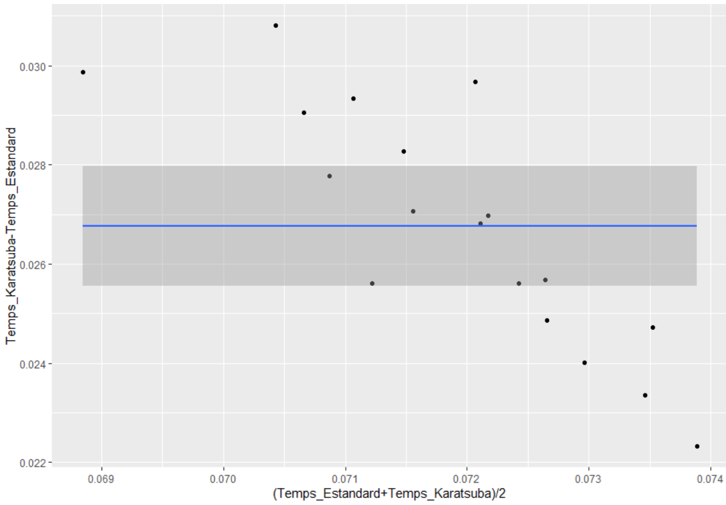
"***S***" descriu la dispersió de les dades (S^2 o variància és la mitjana de totes les distàncies quadrades a la mitjana).

"***et***" diu el error esperat al creure que la desconeguda mitjana poblacional (en totes, o infinites, execucions) iguala a la mitjana mostral calculada.

8. Calculeu el IC95% per la diferència de temps entre els dos algoritmes.

$$IC(\mu_M, 95\%) = \overline{D}_{K-S} \pm t_{0.975,17} \cdot es = \ 0.02678 \ \pm \ 2.110 \cdot 0.00057 = [\mathbf{0\ 025577, 0, 027983}] = [\mathbf{0\ 0256, 0, 0280}]$$

9. Interpreteu globalment els resultats. Utilitzeu el següent gràfic com a suport a la interpretació.



S guanya a K per 26.8ms en mitjana, encara que la incertesa per no haver provat amb tots el possibles nombres de 900 bits amplia aquesta estimació puntual a l'interval [0.0256,0.0280].

El gràfic ens diu que aquesta estimació pot ser millorada, ja que la diferència depèn inversament de la seva mitjana. Potser seria convenient estudiar si alguna variable està relacionada amb els temps d'algun dels algorismes. *O potser és una conseqüència de les variàncies desiguals: com el temps de K és gairebé constant, amb els casos més costosos per S tindríem una mitjana dels temps més gran però també una diferència més petita.*

10. Té alguna importància per interpretar aquest IC que les desviacions tipus de E i S semblin diferents? *Pista: considereu la premissa d'additivitat de l'efecte.*

Un efecte additiu, constant, no és compatible amb aquestes dades. Una vegada més, ens indica que haurem de continuar l'estudi.

qnorm(0.900) = 1.282	qnorm(0.975) = 1.960	qt(0.950,18)=1.734	qt(0.975,18)=2.101	qt(0.950,36)=1.688	qt(0.975,36)=2.028
qnorm(0.925) = 1.440	qnorm(0.990) = 2.326	qt(0.950,17)=1.740	qt(0.975,17)=2.110	qt(0.950,35)=1.690	qt(0.975,35)=2.030
qnorm(0.950) = 1.645	qnorm(0.995) = 2.576	qt(0.950,16)=1.746	qt(0.975,16)=2.120	qt(0.950,34)=1.691	qt(0.975,34)=2.032

BD NOM:_____________________________COGNOMS:____

Sigueu concisos i feu lletra llegible. ***Contesteu cada pregunta en el seu lloc.*** ***Expliciteu i justifiqueu els càlculs.***

El forat de gènere (*gender pay gap*) es refereix a la diferència de sou existent entre un treballador home i una treballadora dona. Una empresa de consultoria està estudiant la seva situació particular al respecte, i recull dades dels seus assalariats (dades provinents d'una mostra a l'atzar de 30 homes i 20 dones). La variable resposta **Y** és el salari anual actual en milers d'euros; **G** és el gènere ("m", home, "w", dona); **xp** és el nombre d'anys de vida laboral de l'empleat; finalment, **age** és la seva edat. A sota teniu un extret de diversos models aplicats a la resposta.

```
A            Estimate Std. Error t value Pr(>|t|)
(Intercept)  48.0241     0.7107  67.576  < 2e-16
Gw           -7.4136     1.1237  -6.598 3.05e-08
Residual standard error: 3.893 on 48 degr of
freedom
Multiple R-squared:  0.4756,  Adjusted R-squared:
0.4646
```

```
B            Estimate Std. Error t value Pr(>|t|)
(Intercept)  31.7455     1.7458  18.184  < 2e-16
xp            1.0350     0.1301   7.958 2.55e-10
Residual standard error: 3.529 on 48 degr of freedom
Multiple R-squared:  0.5689,    Adjusted R-squared:
0.5599
```

```
C            Estimate Std. Error t value Pr(>|t|)
(Intercept)  36.9077     1.5525  23.773  < 2e-16
xp            0.7922     0.1051   7.535 1.27e-09
Gw           -5.0984     0.8237  -6.190 1.39e-07
Residual standard error: 2.647 on 47 degr of
freedom
Multiple R-squared:  0.7625,  Adjusted R-squared:
0.7524
```

```
D            Estimate Std. Error t value Pr(>|t|)
(Intercept)  44.0387    10.9254   4.031 0.000207
xp            1.1575     0.5638   2.053 0.045789
age          -0.3088     0.4683  -0.659 0.512878
Gw           -5.2430     0.8572  -6.116 1.94e-07
Residual standard error: 2.663 on 46 degr of freedom
Multiple R-squared:  0.7647,    Adjusted R-squared:
0.7494
```

Es tracta d'un estudi experimental o observacional? Canviaria la vostra resposta si fossin 25 homes i 25 dones? Raoneu els vostres arguments. (1 pt)

És un estudi observacional, perquè òbviament no hem pogut assignar el gènere als participants i seguir els esdeveniments. El fet que sigui un nombre desigual el d'homes i dones no és rellevant, i si fos el mateix nombre no canviaria la situació.

Quin és el salari mitjà dels homes i dones d'aquesta mostra? (1 pt)

Del model A, 48,024.1€ per homes, i 48,024.1 − 7,413.6€ = 40,610.5€ per dones.

Estimeu per interval de confiança 95% la mitjana de la diferència de salari. Hi ha una única resposta? Hi ha coincidència? Com s'interpreta? (2 pts)

Podem obtenir diferents opcions. Amb el model A que només considera el gènere, 7,413.6 ± 2.011 ×1,123.7 = (5,154 – 9,673)€

Amb el model C, que també incorpora la vida laboral (és a dir, per a un mateix temps de vida laboral), 5,098.4 ± 2.012 ×823.7 = (3,441 – 6,756)€

Amb el model D, que a més a més té en compte l'edat de la persona, 5,243 ± 2.013 × 857.2 = (3,517 – 6,969)€

Com es pot veure, les diferents respostes tenen una part coincident [la diferència salarial mitjana estaria entre 5.5K€ i 9.7K€, o bé 3.4K€ i 6.8K€]; sembla que tenir en compte la vida laboral redueix una mica el gap. També es veu un descens en l'error tipus de l'estimació (de 1,123.7 a 823.7). En canvi, afegir al model l'edat no aporta cap benefici.

Com es valora en termes de salari cada any addicional de vida laboral? Hi ha una única resposta? Hi ha coincidència? Com valoreu la precisió de l'estimació? Com s'interpreta? (2 pts)

Novament tenim diferents respostes segons el model escollit:

- explicant el salari només per la vida laboral: 1,035€ / any (se=130€)
- afegint el gènere: 792.2€ / any (se=105.1€)
- amb l'edat: 1,157.5€ / any (se=563.8€)

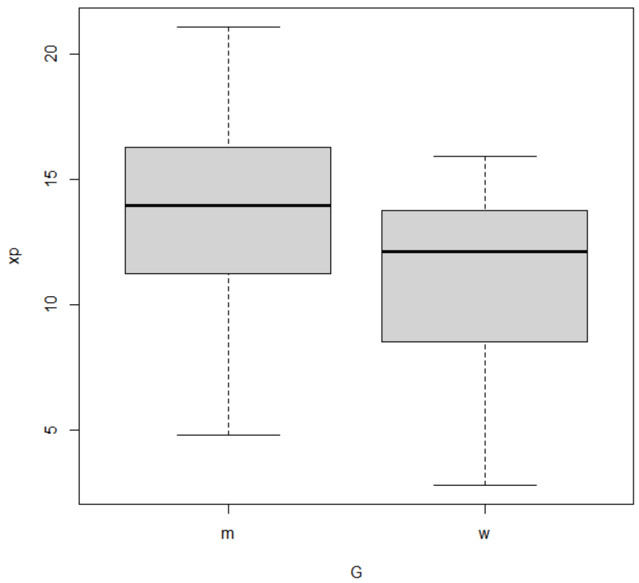
S'observa que amb gènere o sense l'estimació és bastant diferent, encara que l'error tipus és semblant. En canvi amb el factor "edat" l'error d'estimació es dispara (cinc vegades), per tant es tracta d'un valor molt poc fiable. El més correcte seria pensar que cada any més a la vida laboral suposa un increment d'uns 800 euros (amb 95% de confiança, aproximadament entre 600 i 1000).

A la mostra, la desviació tipus del salari és de 5,320€. Compareu aquest valor amb les diferents desviacions residuals dels models **A** a **C**, destacant el diferent significat de cadascú. (1 pt)

A: 3,893€ és la estimació conjunta de la desviació tipus del salari en grups on el gènere és el mateix.
B: 3,529€ és la estimació conjunta de la desviació tipus del salari per a treballadors amb la mateixa vida laboral
C: 2,647€ és la estimació conjunta de la desviació tipus del salari per a treballadors del mateix gènere i amb la mateixa vida laboral

Es veu clarament que, a la que hi afegim factors explicatius, la variabilitat residual va disminuint.

Què ens aporta la figura següent? Podeu relacionar aquesta informació amb el que heu vist anteriorment? (1 pt)



Aquest diagrama ens diu que els homes de la mostra tenen més experiència (o que la seva vida laboral és major, en general). Llavors, quan es compara el salari d'homes i dones no només estem comparant el gènere, sinó que també hi intervé un altre factor molt rellevant en el salari, ja que més vida laboral està relacional amb més salari. Per aquesta raó observem un gap més gran amb el model A, i menor si hi afegim el factor de la vida laboral, ja que llavors interpretem l'estimador del gap considerant que les altres variables del model són fixes (comparem homes i dones amb la mateixa vida laboral).

Un manager de la consultora no sap interpretar el valor -0.3088 del darrer model. Segons ell, no és versemblant que a més edat menys sou. Com li podeu ajudar? (1 pt)

L'estimació del factor *edat* present a aquest model, junt amb *gènere* i *vida laboral*, és un valor negatiu, però el seu error tipus és molt gran (0.4683). Per tant, tenim un rati senyal/soroll = -0.659, significant que aquesta estimació és molt imprecisa (compte, no es pot dir que l'edat no està relacionada amb el salari). Casualment hem obtingut un valor negatiu, i amb una altra mostra podria ser molt superior a 0.

(Addicionalment, la causa d'aquesta estimació tan dolenta és la molt previsible alta correlació entre edat i vida laboral. Posar dos predictors que pràcticament diuen el mateix no millora el model, al contrari, augmenta l'error tipus d'aquests factors, com es pot veure a les sortides. Observeu que el R^2 pràcticament no ha canviat)

Expliqueu què obtenim amb les següents instruccions: (1 pt)

```
ind = data.frame(age=50, G='w', xp=25)
predict(lm(Y~xp+G), ind, interval='confidence')
       fit      lwr      upr
1 51.61551 48.44504 54.78598
```

Tenim una predicció individual del salari (és a dir, no per a la mitjana sinó per a una persona) quan es tracta de una dona de 50 anys amb 25 anys de vida laboral, i hem obtingut que el seu salari estaria entre 48.4K€ i 54.8K€.

qt(0.95, 46)=	1, 67866	qt(0.975, 46)=	2, 01290	qt(0.95, 48)=	1, 67722	qt(0.975, 48)=	2, 01063	qnorm(0.95)=	1.64485
qt(0.95, 47)=	1, 67793	qt(0.975, 47)=	2, 01174	qt(0.95, 49)=	1, 67655	qt(0.975, 49)=	2, 00958	qnorm(0.975)=	1.95996