

NOM: _____ COGNOM: _____
(Contesteu cada pregunta en el seu lloc. Expliciteu i justifiqueu els càlculs)

Problema 1 (Bloc C)

Nota: el separador decimal en tot l'exercici és el punt (".")

El *Cyber Monday* és un dilluns on les empreses que tenen servei de venda online fan descomptes dels seus productes. S'han recollit dades de preus de productes de 2 grans tendes ("Amazones" i "Market") abans i durant el *Cyber Monday* (A i C respectivament). Algunes dades de 2 productes (mòbils i ordinadors portàtils) de gama intermèdia són:

			(A) Preus abans Cyber Monday		(C) Preus Cyber Monday	
Tenda	Producte	N	$\sum A_i$	$\sum(A_i^2)$	$\sum C_i$	$\sum(C_i^2)$
Market	mòbil	12	5395	2622825	5214	2455166
Amazones	mòbil	12	4631	1963139	4568	1929330
Market	ordinador	12	5925	3174325	5782	3085308
Amazones	ordinador	12	6273	3534247	5979	3266463

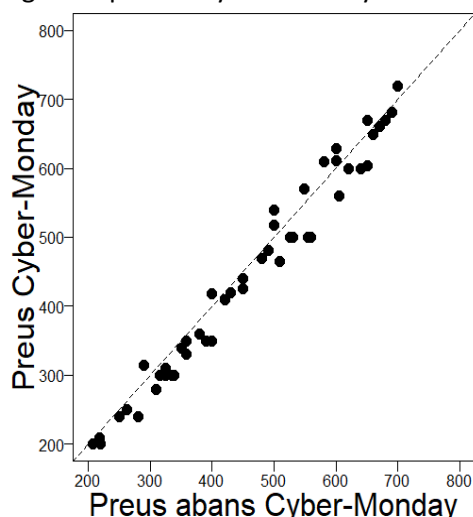
		(A) Preus abans Cyber Monday		(C) Preus Cyber Monday	
Tenda	N	$\sum A_i$	$\sum(A_i^2)$	$\sum C_i$	$\sum(C_i^2)$
Market					
Amazones					

- 1.- (0.5 punts) Indiqueu els valors a la taula de la dreta per a les dues tendes (sense distingir per tipus de productes)
- 2.- (1.5 punts) Doneu una estimació puntual de la mitjana de **preus al Cyber Monday** a la tenda **Market** (indistintament que siguin mòbils o ordinadors). Quin és l'error estàndard d'aquesta estimació?

- 3.- (1 punt) Calculeu un interval de confiança (IC) del 99% per a la mitjana de **preus al Cyber Monday** a la tenda **Market** (indistintament que siguin mòbils o portàtils) suposant una desviació poblacional de 150 euros

- 4.- (1 punt) L'IC del 95% de la mitjana dels **preus de mòbils al Cyber Monday** amb σ desconeguda a la tenda **Amazones** és [297, 464] amb els 12 productes recollits (per tant l'amplada és 167). Amb una mostra de 24 productes de la mateixa tenda i mateixos indicadors, quin dels següents 4 intervals de confiança del 95% seria més plausible obtenir? Per què?
- a) [310, 451] b) [340, 421] c) [322, 439] d) [332, 449]

- 5.- (1.5 punts) El següent gràfic representa els preus al Cyber Monday versus els preus abans del Cyber Monday pels 48 productes. La línia puntejada representa la recta bisectriu ($y = x$). Observant el gràfic, indiqueu la proporció de productes que pugen de preu el Cyber Monday i calculeu i interpreteu un IC95% d'aquesta proporció



6.- (1 punt) Argumenteu quines característiques han de tenir les dades per tal de calcular un interval de confiança per a la diferència de mitjanes de preus entre les dues tendes (*Amazones* versus *Market*) com a mostres aparellades o independents

7.- (1 punt) A partir només de la següent sortida de R, calculeu un interval de confiança del 90% per a la diferència de preus del Cyber-Monday menys els preus abans del Cyber Monday.

```

Paired t-test
data:  preus_cyber(C) and preus_abans(A)
t = -4.1278, df = 47, p-value = 0.0001487
alternative hypothesis: true mean difference is not
equal to 0
95 percent confidence interval:
 -21.102015  -7.272985
sample estimates:
mean difference  -14.1875

```

8.- (1.5 punts) Calculeu un interval de confiança del 95% per a la diferència de mitjanes de **preus abans del Cyber Monday entre mòbils (m) i ordinadors portàtils (o)** a la tenda **Amazones** assumint variàncies iguals. Interpreteu l'interval, argumenteu si podem concloure que els mòbils i els portàtils tenen diferent preu (quantificant la diferència si en tenen)

9.- (1 punt) Troba la grandària mostral necessària per a que l'anterior interval de confiança del 95% tingui una amplada de 150€ assumint que la σ comuna als 2 grups és 100€ i que volem el mateix nombre de portàtils que de mòbils.

qnorm(0.900) = 1.282	qnorm(0.975) = 1.960	qt(0.950,22)=1.717	qt(0.975,22)=2.074	qt(0.950,46)=1.679	qt(0.975,46)=2.013
qnorm(0.925) = 1.440	qnorm(0.990) = 2.326	qt(0.950,23)=1.714	qt(0.975,23)=2.069	qt(0.950,47)=1.678	qt(0.975,47)=2.012
qnorm(0.950) = 1.645	qnorm(0.995) = 2.576	qt(0.950,24)=1.711	qt(0.975,24)=2.064	qt(0.950,48)=1.677	qt(0.975,48)=2.011

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

S’analitza si la diferència en el temps d’execució [ms] entre C i C++ a l’hora de trobar els valors propis d’una matriu depèn de la mida de la matriu (aquí, les dimensions seran $n \times n$ on n és 10, 20, ..., 190, 200). Cada operació amb una matriu a l’atzar d’una mida determinada es fa amb cada llenguatge, obtenint-ne els temps d’execució (C i + respectivament). Per conveniència, treballarem amb el logaritme del temps, i la seva diferència “ $D = \ln(+)-\ln(C)$ ” equival al logaritme del rati: “ $\ln(+/C)$ ”. A sota veieu les sortides de R per a dos models que hem provat (dades\$Y és el logaritme dels temps; dades\$L és llenguatge: C o +).

Call: lm(formula = D ~ 1) Residuals: Min 1Q Median 3Q Max -0.41595 -0.34077 -0.05871 0.24805 0.77256 Coefficients: Estimate Std.Error t value Pr(> t) (Intercept) 1.23682 0.08229 15.03 5.32e-12 Residual standard error: 0.368 on 19 degrees of freedom	Call: lm(formula = Y ~ L, data = dades) Residuals: Min 1Q Median 3Q Max -0.82690 -0.38031 -0.04163 0.33268 1.30651 Coefficients: Estimate Std.Error t value Pr(> t) (Intercept) 0.5784 0.1169 4.950 1.55e-05 LC -1.2368 0.1653 -7.484 5.50e-09 Residual standard error: 0.5226 on 38 degrees of freedom. Multiple R-squared: 0.5958, Adjusted R-squared: 0.5852 F-statistic: 56.01 on 1 and 38 DF, p-value: 5.496e-09
---	--

A) Descriuiu quins són els dissenys i les premisses que hi ha darrera de cada opció (1, esquerra; 2, dreta).

B) Digueu si són dues opcions vàlides; si ho són, justifiqueu la resposta; si no ho són, digueu quina és l’opció apropiada.

C) Independentment de la resposta anterior, interpreteu el resultat obtingut amb l’opció 1.

D) El mateix amb l’opció 2 (a aquesta pregunta i a l’anterior, no cal considerar aspectes menors).

E) Què representa el “Residual Standard error” en cada cas?

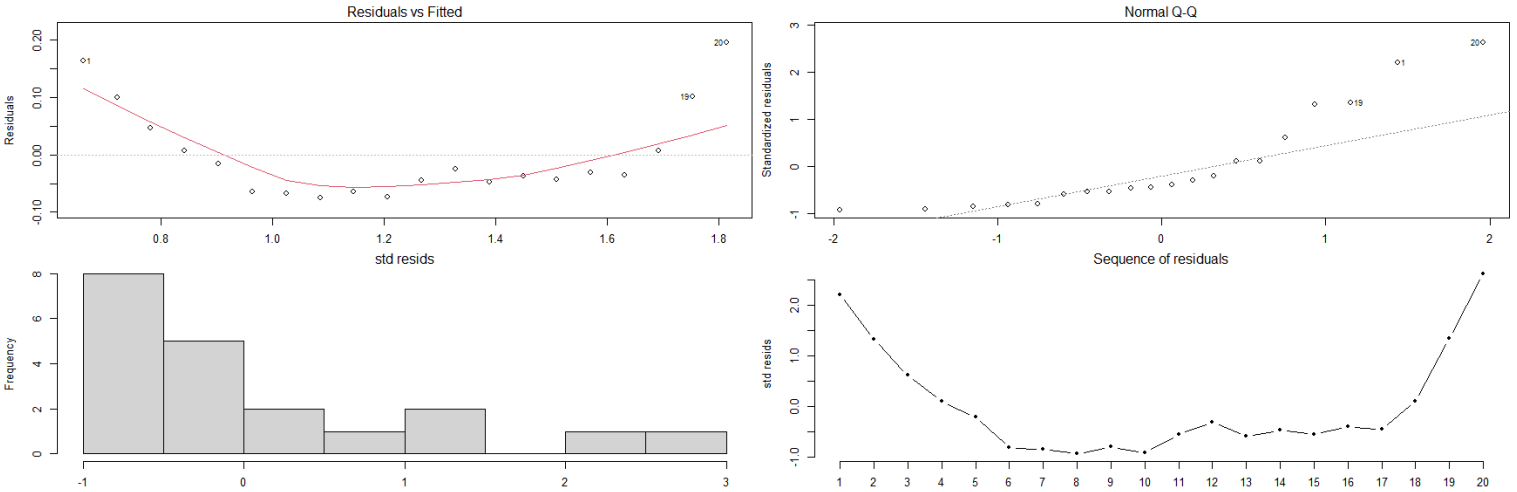
F) Expliqueu de què ens informa el valor de “Multiple R-squared”. Perquè una opció inclou el coeficient de determinació i l'altra no?

G) Calculeu un IC al 95% de confiança per al paràmetre del model 1, i explica què ens diu respecte als dos llenguatges.

H) Seguidament teniu un fragment de la sortida del model “D en funció de la mida (N)”. Comenteu què representen els dos valors de la columna “Estimate”, d’acord amb el model estadístic que hem aplicat en aquesta ocasió.

```
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.5993784  0.0383419   15.63 6.45e-12
N           0.0060709  0.0003201    18.97 2.40e-13
Residual standard error: 0.08254 on 18 degrees of freedom
Multiple R-squared:  0.9523,    Adjusted R-squared:  0.9497
```

I) Segons la sortida anterior, utilitzeu un interval de confiança 95% per a dir com afecta la mida de la matriu en la relació entre C++ i C.



J) A la vista d’aquests gràfics, com us sembla que ha estat el compliment de les premisses del model emprat? Justifiqueu en detall la resposta.

qt(0.95, 18)=	1,7341	qt(0.975, 18)=	2,1009	qt(0.95, 19)=	1,7291	qt(0.975, 19)=	2,0930
qt(0.95, 38)=	1,6860	qt(0.975, 38)=	2,0244	qt(0.95, 39)=	1,6849	qt(0.975, 39)=	2,0227