

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

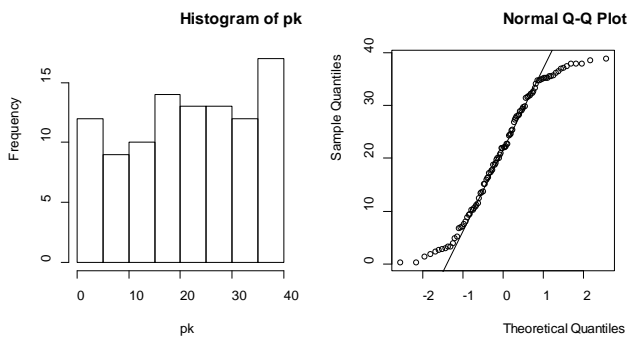
(Contesteu cada pregunta en el seu lloc. Expliqueu i justifiqueu els càlculs.)

**Problema 1 (B4).** El Servei Català de Trànsit (SCT) ens controla. Des de fa un temps, el SCT es capaç de captar les senyals dels conductors que envien “whatsapp” mentre estan conduint a les autopistes que gestiona. L’aplicació que han construït recull la informació corresponent a 4 variables: identificador de l’autopista, data, hora i punt quilomètric de l’autopista des d’on es rep el senyal. Un hacker amic nostre ha aconseguit accedir a una mostra de 100 deteccions d’una mateixa autopista. Les dades dels punts quilomètrics i la seva descriptiva es mostren a les següents taules:

0.3	0.3	1.4	1.8	2.3	2.7	2.8	2.9	3.3	3.3	3.9	4.9	5.2	6.8	6.9	7.1	7.6	8.0	8.8	9.3
9.4	10.2	10.2	10.6	10.8	11.2	11.5	12.5	13.4	13.5	13.8	15.1	15.2	16.0	16.2	16.4	17.2	17.3	17.7	17.9
18.8	18.8	19.1	19.7	20.0	20.1	20.7	21.0	21.9	22.0	22.1	22.1	22.3	22.8	22.8	24.3	24.5	24.6	25.3	25.5
26.9	27.4	27.7	27.9	28.2	28.3	29.0	29.1	29.4	29.7	29.9	31.4	31.6	31.7	31.9	32.2	32.4	32.8	33.3	34.2
34.8	34.8	34.9	35.1	35.2	35.3	35.3	35.5	35.6	35.8	36.2	36.6	37.0	37.2	37.5	37.9	37.9	37.9	38.6	38.9

<pre>&gt; length(pk) [1] 100</pre>	<pre>&gt; summary(pk)   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  0.30  11.10   22.05   21.37  31.98   38.90</pre>	<pre>&gt; sd(pk) [1] 11.66476</pre>	<pre>&gt; sum(pk&lt;10) [1] 21</pre>
------------------------------------	--	-------------------------------------	--------------------------------------

1) (1 punt). Proposa una distribució que et sembli raonable per aquestes dades i justifica-ho.



2) (1 punt). Ens interessa saber de quina autopista provenen les dades però l’identificador de l’autopista és un codi que no permet esbrinar de quina via es tracta. Si sabéssim la longitud total de l’autopista de la qual hem recollit les dades, la podríem comparar amb les longituds reals de les autopistes de Catalunya. Independentment de la resposta del primer apartat, suposa que les dades es distribueixen com una uniforme continua. Dóna almenys 2 estimacions puntuals raonables pel punt quilomètric màxim d’aquesta autopista (inclou la forma analítica, és a dir la fórmula, i el càlcul dels teus estimadors).

3) (4 punts. 0.8 per apartat) Coneixem que l’autopista C-58 té exactament 40.0 km. Posem a prova la hipòtesi de que l’autopista de la qual hem recollit les dades tingui aquesta longitud suposant que segueix una distribució uniforme continua. Tingués present que una forma equivalent és contrastar si el punt quilomètric mitjà és 20 amb un  $\alpha=0.05$ .

a) Hipòtesis

- b) Estadístic, distribució, premisses
- c) Valor de l'estadístic i punt crític amb un  $\alpha=0.05$
- d) Conclusió. Veient els resultats, ¿creus que hem identificat l'autopista com la C-58?
- e) Calcula el valor mínim de la mitjana de la mostra amb el qual acceptariem la hipòtesi alternativa de que l'autopista té més de 40 km amb un  $\alpha=0.05$  unilateral.
- 4) (4 punts. 1 per apartat). L'exercici anterior l'hem basat en la premissa de que era una distribució uniforme, però no estem del tot segurs. Per intentar discernir si es tracta d'una uniforme  $[0,40]$ , fes un contrast d'hipòtesi amb un  $\alpha=0.05$  sobre si la probabilitat de tenir *whatsapps* en els 10 primers quilòmetres és igual a la probabilitat teòrica de tenir-los abans de 10 en una uniforme  $[0,40]$ .
- a) Hipòtesi
- b) Estadístic, distribució, premisses
- c) Valor de l'estadístic i punt crític
- d) Conclusió. ¿Podem afirmar que la distribució de referència és una uniforme?

**Bloc 5.** El passat 21 d'abril de 2014 *La Vanguardia* va publicar un article titulat “Conducció política temerària” que, basant-se en un sondeig del CIS (Baròmetre de setembre de 2013), donava a entendre que hi ha una relació entre conduir temeràriament i votar a certs partits. Concretament s’assenyalava als votants d’UPyD com els més disposats a infringir les regles de circulació mentre que els de CiU estarien entre els més respectuosos amb elles. A l’hora de buscar possibles causes d’aquesta sorprenent relació entre vot i conducció, l’últim paràgraf de l’article apuntava la possibilitat de que l’edat dels votants hi pogués tenir a veure (l’article deia que els votants d’UPyD són en mitjana més joves que els de PP, PSOE, CiU o CC, per exemple).

Farem servir les dades del Baròmetre de setembre de 2013 del CIS per comprovar si les afirmacions de l’article estaven ben fonamentades. Ens centrarem únicament en les dades corresponents als enquestats que van manifestar que a les eleccions de 2011 (*Recuerdo.de.voto.2011*) van votar o bé a UPyD o bé a CiU. Estudiarem la seva edat (variable *Edad*). Pel que fa a la temeritat en la conducció, ens limitarem a estudiar les respostes a la preguntes de si s’excedeix el límit de velocitat en ciutat, codificada així:

1:Siempre, 2:Muchas veces, 3:Algunas veces, 4:Pocas veces, 5:Nunca.

- En primer lloc volem comparar les edats esperades de votants d’UPyD i de CiU, a les que anomenarem  $\mu_1$  i  $\mu_2$ , respectivament. En R definim els vectors *Edad.UPyD* i *Edad.CiU* que contenen les edats dels enquestats que han declarat haver votat el 2011 a UPyD o CiU, (respectivament). Les mides mostrals són, respectivament,  $n_1 = 73$  i  $n_2 = 65$ , respectivament. A continuació oferim unes descriptives numèriques i gràfiques d’aquestes dues variables.

```
> summary(Edad.UPyD)
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 20.00  31.00   38.00   39.66  47.00   77.00
```

```
> sd(Edad.UPyD)
```

```
[1] 13.43836
```

```
> summary(Edad.CiU)
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 21.00  39.00   50.00   52.31  65.00   90.00
```

```
> sd(Edad.CiU)
```

```
[1] 17.61917
```

1. **(1 punt)** Es tracta de mostres aparellades o de mostres independents? Per què?
2. **(2 punts)** Constrasteu la igualtat de les variàncies,  $\sigma_1^2$  i  $\sigma_2^2$ , de les edats de votants d’UPyD i de CiU, respectivament. Indiqueu les hipòtesis nul·la i alternativa. Quin estadístic fareu servir per fer la prova? Quina distribució té sota la hipòtesi nul·la? Sota quina condició? Creus que aquesta condició es verifica en aquest cas? Doneu el valor de l’estadístic de la prova, digueu quin és el p-valor de la prova i si rebutjaríeu  $H_0$  amb un risc  $\alpha = 0.01$ .

*Nota: Feu servir la següent informació (recordeu que  $\text{pf}(\mathbf{x}, \mathbf{g1}, \mathbf{g2})$  és la funció de distribució d’una  $F$  amb graus de llibertat  $\mathbf{g1}$  i  $\mathbf{g2}$  avaluada al punt  $\mathbf{x}$ ):*

```
> pf(1.71901, 72, 64)
```

```
[1] 0.9858147
```

```
> 2*(1-pf(1.71901, 64, 72))
```

```
[1] 0.02610106
```

```
> pf(1.874246, 64, 72)
```

```
[1] 0.995
```

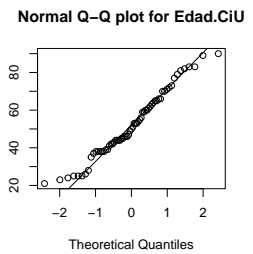
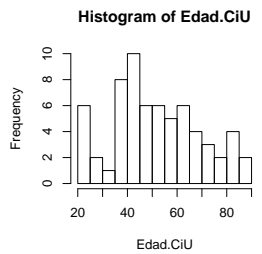
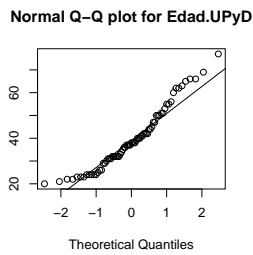
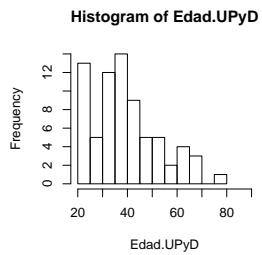
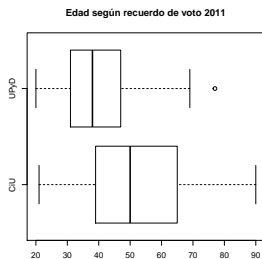
```
> pf(1.893433, 72, 64)
```

```
[1] 0.995
```

3. **(3 punts)** Constrasteu ara la igualtat de les edats esperades dels votants d’UPyD i de CiU,  $\mu_1$  i  $\mu_2$ . Indiqueu les hipòtesis nul·la i alternativa, tenint en compte el que diu la notícia del diari per decidir si fareu una prova bilateral o unilateral. Quin estadístic fareu servir per fer la prova? Quina distribució té sota la hipòtesi nul·la? Sota quines condicions? Es verifiquen aquestes condicions? Quina és la regió crítica de la prova amb un risc  $\alpha = 0.01$ ? Doneu el valor de l’estadístic de la prova, digueu quin és el p-valor de la prova i si rebutjaríeu  $H_0$  amb un risc  $\alpha = 0.01$ .

*Nota: Aproximeu la distribució  $t_n$  per una  $N(0,1)$  si els graus de llibertat  $n$  són més grans que 100.*

4. **(2 punts)** Doneu un interval de confiança 99% per a la diferència de les edats esperades dels votants d’UPyD i de CiU.



- Ara compararem els estils de conducció dels votants d'UPyD i de CiU. Les preguntes del baròmetre del CIS relatives a aquest tema només les havien de respondre aquells enquestats que habitualment condueixen un vehicle de motor. Hi va haver 63 votants d'UPyD que condueixen, i 54 de CiU.

Definim la variable `Exc.lv.ciud` que val `TRUE` si un conductor enquestat declara que excedeix el límit de velocitat a la ciutat *sempre*, *moltes vegades* o *algunes vegades*, i val `FALSE` en altres casos. El creuament de les variables `Exc.lv.ciud` i `voto.2011.cond` (que indica quin partit van votar, UPyD o CiU, els votants d'aquests dos partits que condueixen habitualment) ve donat per la següent taula:

voto.2011.cond	Exc.lv.ciud	
	TRUE	FALSE
UPyD	25	38
CiU	12	42

5. (2 punts) Sigui  $p_1$  (respectivament,  $p_2$ ) la probabilitat que un conductor votant d'UPyD (respectivament, de CiU) triat a l'atzar excedeixi el límit de velocitat a la ciutat sempre, moltes vegades o algunes vegades. Contrasteu la igualtat de  $p_1$  i  $p_2$ . Indiqueu les hipòtesis nul·la i alternativa, tenint en compte el que diu la notícia del diari per decidir si fareu una prova bilateral o unilateral. Quin estadístic fareu servir per fer la prova? Quina distribució aproximada té sota la hipòtesi nul·la? Quina és la regió crítica de la prova amb un risc  $\alpha = 0.05$ ? Doneu el valor de l'estadístic de la prova, digueu quin és el p-valor de la prova i si rebutjaríeu  $H_0$  amb un risc  $\alpha = 0.05$ .

NOM: \_\_\_\_\_ COGNOM: \_\_\_\_\_

(Contesteu cada pregunta en el seu lloc. Expliqueu i justifiqueu els càlculs. Totes les preguntes valen igual)

### Problema 3 (B6).

El tema de B7 d'un grup d'estudiants consisteix en determinar si la distància al mòdem wifi és determinant en la velocitat de connexió. Per això, dissenyen un estudi en el que prenen la velocitat de descàrrega amb una pàgina estàndard, ubicant l'ordinador a diferents llocs de casa i mesurant la distància al mòdem en línia recta. Aquestes són les dades recollides:

D [m.] 0.80 1.100 2.000 2.100 3.300 3.400 4.600 4.800 5.000 5.000 5.000 5.600 5.700 5.800 7.400  
 V [MBs] 2.98 2.102 2.485 3.016 3.037 1.965 2.704 2.942 3.123 2.683 2.474 2.544 2.412 1.357 2.443

1. Amb R han estimat els paràmetres d'un model que relaciona "distància" amb "velocitat" però, ¿quina ha estat la variable resposta i quina la variable explicativa, entenent que han fet un model *sensat*? Justifiqueu si la situació oposada tindria més o menys sentit. El fragment de la sortida que hem inclòs a continuació us pot ajudar a contestar.

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	2.75517	_____	9.048	5.67e-07	***
x	-0.04968	0.06762	_____	0.476	

2. Trobeu els valors mancants de la sortida anterior.

3. Escriviu a continuació quines són les dues proves d'hipòtesis formals que s'estan contrastant a la sortida anterior. Expliqueu també què volen dir cadascuna d'elles:

-  $H_0$ :

-  $H_0$ :

4. Corregiu/milloreu les afirmacions següents, o confirmeu si és que no contenen cap error:

- Cada metre que ens apropem al mòdem significa que la velocitat baixa 0.05MBs (aprox.)
- Si ens allunyem un metre del mòdem la velocitat augmenta en 2.76MBs (aprox.)
- Quan la distància al mòdem disminueix 5cm aconseguim una millora en la velocitat de 1MBs
- No es veu que la velocitat es modifiqui si la distància al mòdem canvia

5. Sabent que la mitjana de D és 4.106667, la de V és 2.551133, les respectives variàncies són 3.656381 i 0.2263794, i la correlació val -0.1996768, reproduïu el càlcul del valor 0.06762 de la taula anterior. Calculeu també l'estimació de la desviació residual i el coeficient de determinació corresponent a aquestes dades. Justifiqueu totes les passes.

6. Construïu un interval de confiança al 99% per al pendent de la recta, i expliqueu què significa el que heu calculat.

7. Un amic té molta por de les radiacions electromagnètiques que desprèn el mòdem, però pensa que si s'allunya massa està perdent velocitat. Creieu que aquests resultats (concretament, l'apartat 6) el poden ajudar per decidir on ha de col·locar l'ordinador, i perquè?

8. Com tothom diu la seva, un altre amic diu que aquest estudi no conclou res perquè la grandària de la mostra no és l'adequada. Que el que cal és que la precisió de l'estimació sigui alta i, per tant, que l'error estàndard de l'estimador del pendent no sigui superior a 0.025 (*ell sabrà perquè*). Assumint (a) que la desviació residual poblacional és 0.5 i (b) que la variabilitat de les distàncies preses al nou estudi serà més o menys igual: quantes observacions s'haurien d'obtenir?

9. Imagineu que l'estudi s'ha repetit, i hem obtingut el resultat adjunt. Expliqueu si seria correcte utilitzar el model lineal simple que coneixeu per relacionar les variables. Digueu quines assumpcions ("premisses") del model es podrien donar per vàlides i quines no.

