

# Las compañías telefónicas nos timan

## Abstract

**Objectiu:** Veure si hi ha diferències entre la velocitat de baixada real (VBR) i la contractada (VBC).

**Encara que tots sabem que sí.**

**Mètodes:** Es va examinar la relació entre la VBR i la VBC mitjançant regressió lineal simple.

**Resultats:** En el model que hem ajustat, es veu que efectivament les companyies de telèfons ens estan timant, ja que el coeficient de VBC és inferior a 1:

$$VBR = 1299 + 0.44 \cdot VBC$$

**Conclusió:** La VBR és pot saber multiplicant la VBC per 0.44 i sumant-li 1299.

## Justificació

La diferència entre la velocitat contractada i la velocitat real en les **conexions** a Internet és una de les queixes més freqüents entre els usuaris de les **línies** ADSL. En un inici, les diferents companyies proveïdores d'Internet només garantien un velocitat de baixada del 10% i no va ser fins al juliol de 2008 que es va posar en marxa un decret que obligava a les operadores a garantir un 80% de la velocitat contractada.

### Objectiu

L'objectiu d'aquest estudi és mirar si es poden trobar diferències entre la velocitat de baixada real (VBR) i la contractada (VBC).

## Material y mètodes

Es va proposar recollir la informació sobre la connexió d'ADSL a **uns** estudiants de la FIB. **Uns quants** van participar en l'estudi.

Els alumnes van realitzar una única prova de connexió. No es van incloure les mesures realitzades des de mòdems amb velocitats de 56 kb/s o inferiors, des de dispositius d'Internet mòbils o des de xarxes locals.

La variable resposta en aquest estudi és la VBR i la variable predictora és la VBC.

### Anàlisi estadística

La descriptiva de les variables numèriques es realitza a través de les mitjanes i **Error Estàndards (EE)** i la de les variables categòriques mitjançant els percentatges de cadascuna d'elles. **En l'anàlisi principal es va fer una regressió lineal simple i després vàrem comprovar les premisses.**

## Resultats

### Descriptiva

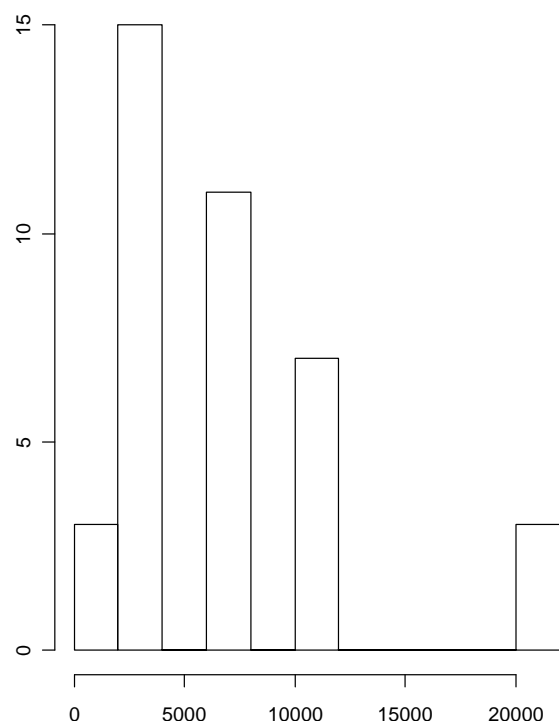
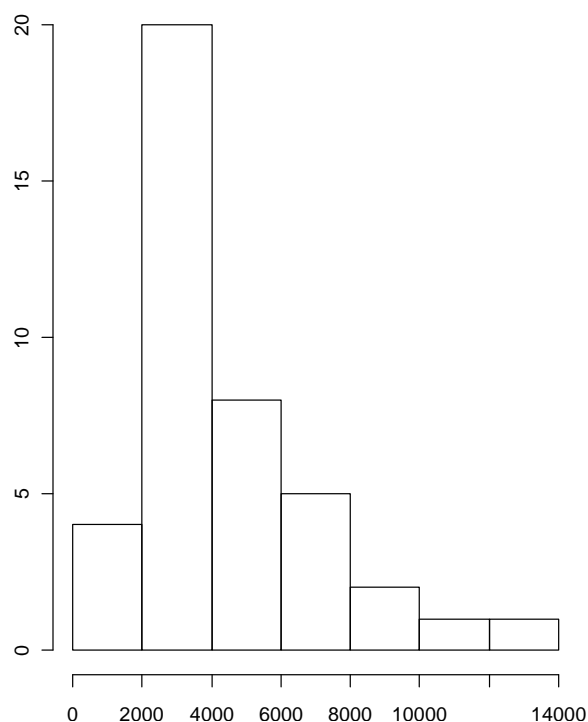
En la **taula de més avall** es mostra la descriptiva global de les variables numèriques amb la mitjana (**EE**) i de les variables categòriques amb el número de participants en cada categoria i el seu percentatge. S'observa que el promig de les velocitats reals són inferiors a les contractades tant de pujada com de baixada. La companyia amb més usuaris **a la població** és Telefònica i **hi ha més gent**

amb cable que amb wifi. Com era d'esperar hi ha més alumnes de fora de Barcelona perquè fora de Barcelona viu més gent que a Barcelona.

	Global			Global	
	n	Mitjana (EE)		n (%)	
Velocitat de baixada real	41	4198.5(2677.3)	Proveïdor	Jazztel	3(7.5%)
				Ono	5(12.5%)
Velocitat de baixada contractada	39	6406.6(5004.0)		Orange	2(5.0%)
				Telefónica	26(65.0%)
Velocitat de pujada real	41	295.3(141.2)		Vodafone	1(2.5%)
				Ya.com	3(7.5%)
Velocitat de pujada contractada	37	354.9(181.1)	Connexió	Cable	23(56.1%)
				Wifi	18(43.9%)
Latència	41	142.6(66.9)	Localització	Barcelona	10(24.4%)
				Fora	31(75.6%)
Distància a la central	41	1939.4(1545.2)			

### **Anàlisi principal**

L'objectiu principal d'aquest estudi és **comparar** la VBR amb la VBC. En la **figura de més avall** es mostra la distribució de les dues variables que intervenen en la regressió. **Hi ha un alumne que té molta VBR. No es veuen masses diferències entre VBR i VBC ja que en els dos histogrames els valors arriben a la dreta de tot.**



## Modelatge

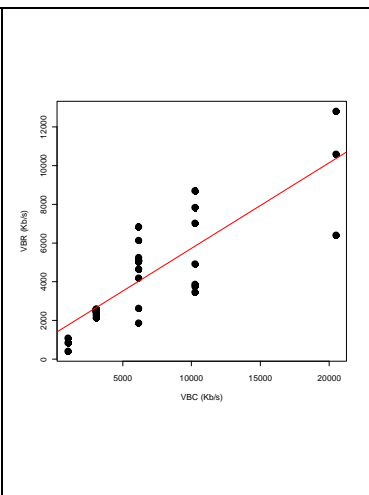
Al ajustar un model lineal de VBR en funció de VBC s'obté la següent sortida en R:

```
Call:
lm(formula = down.speed ~ veloc.cont.down, data = adsl)

Residuals:
    Min       1Q   Median       3Q      Max
-3945.9  -586.6  -108.6   1038.3   2869.2

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.299e+03  3.833e+02   3.388  0.00168 **
veloc.cont.down 4.414e-01  4.738e-02   9.317 3.04e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

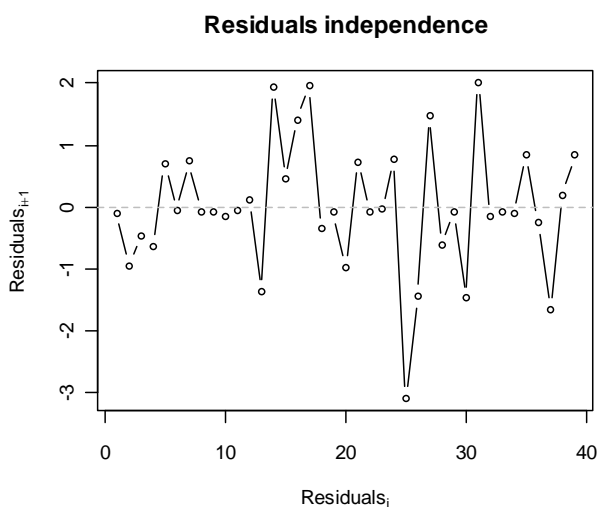
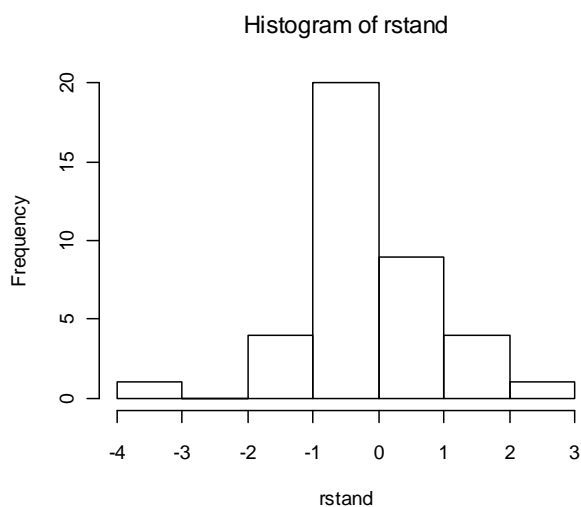
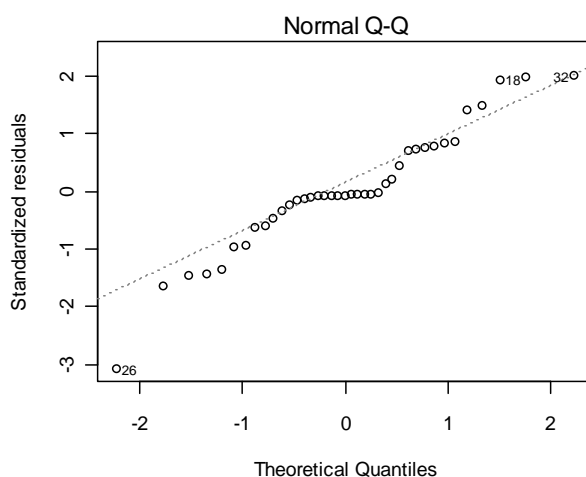
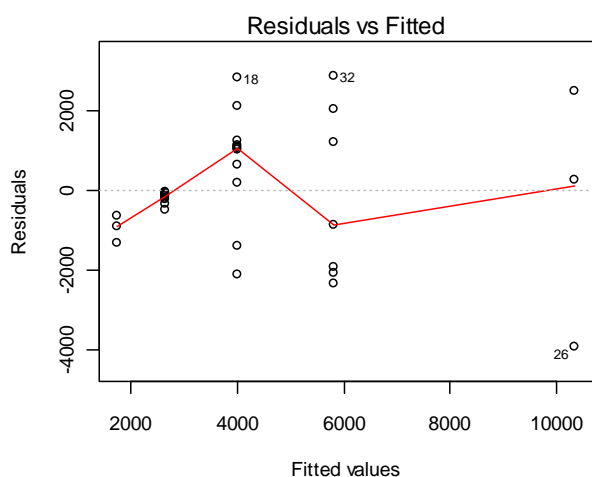
Residual standard error: 1461 on 37 degrees of freedom
(2 observations deleted due to missingness)
Multiple R-squared:  0.7012,    Adjusted R-squared:  0.6931
F-statistic: 86.81 on 1 and 37 DF,  p-value: 3.042e-11
```



El coeficient de 0.44 de la variable explicativa, **am una DE de 0.0047** en el model s'ha d'interpretar com **que els alumnes de la mostra tenen una VBR que està en el 95% dels casos en 0.44 +/- 2.0047 la VBC**. Aquest valor queda molt per sota en el millor dels casos del 80% d'eficiència exigida pel decret de l'any 2008. De totes formes, a continuació s'ha de validar a través de l'anàlisi dels residus.

## II. Comprovació de les premisses

La següent figura mostra els quatre gràfics de residus necessaris per fer l'anàlisi de les premisses. El baix nombre d'observacions i el fet que molts valors de la variable explicativa estiguin repetits complica extreure conclusions. El primer gràfic serveix per avaluar la linealitat. La linealitat sembla complir-se ja que el núvol de punts es distribueix simètricament respecte al 0 pels diferents valors predits. En quant a la homocedasticitat, en el QQ-Norm es veu que els punts es situen raonablement sobre la recta, per tant no existeix heterocedasticitat i per tant, la premissa de Normalitat també es compleix. La independència sí que seria discutible ja que veiem que sí que hi ha un patró en les ratxes d'observacions que queden per sota del zero: 4, 1, 4, 1, 3, 2, 2, 3, 3 i 2. Si ens fixem, al principi només apareixen uns i quatsres i, de sobte, només apareixen dosos i tresos. Això és molt sospitos i fa entreveure que no hi ha independència, però aquesta premissa és poc important, així que donem el model per vàlid.



El model ajustat resultant és:

$$VBR = 1299 + 0.44 \cdot VBC$$

## Discussió

El valor de la constant s'interpreta com que un individu que no tingui línia d'ADSL contractada s'espera que tingui 1299 Kb/s de VBR. Això està molt bé ja que ajuda a potenciar l'accés a Internet a les classes socials més desafavorides.

Per altra banda, el valor 0.44 de la constant s'interpreta com que la VBR és menys de la meitat de la VBC en els individus de la mostra.

La  $R^2$  de 0.70 és molt alta i ens està indicat que el model és vàlid i, per tant que es compleixen les premisses.

Aquests resultats són extrapolables a qualsevol població ja que les línies d'ADSL són iguals a tot arreu i, per tant, no és necessari fer més estudis. En tot cas, si es fan, s'haurien de fer amb més participants perquè el valor de 0.44 del coeficient baixés encara més i, per tant, fos més significatiu.

## ANNEX I. SCRIPT EN R

```
#####  
# Lectura de les dades  
#####  
  
colClasses <-  
c(rep("factor",2),rep("numeric",3),rep("factor",2),"numeric","factor",rep("numeric",2),rep("factor",3),rep("numeric",4))  
adsl <- read.table(url("http://www-eio.upc.es/teaching/pe/DADES/dades_ADSL.txt"), header=TRUE,  
na.strings='00',colClasses=colClasses,dec=",")  
  
#####  
# Depuració i recodificació  
#####  
  
adsl$dia <- as.Date(adsl$dia,"%d/%m/%y")  
# Donar format de data  
adsl$veloc.cont.down <- adsl$veloc.cont.down*1024  
# Es passa de Mb/s a Kb/s  
adsl$log.cont.down <- log(adsl$veloc.cont.down)  
# Es treu el logaritme de la variable en Kb/s  
  
#####  
### Inspecció de les dades  
#####  
  
names(adsl)  
head(adsl)  
summary(adsl)  
  
#####  
### Descriptiva  
#####  
  
##### Variables numèriques #####  
  
##### Descriptiva univariant variables numèriques  
summary(adsl[,varnum])  
apply(adsl[,varnum],2,sd,na.rm=TRUE)  
par(mfrow=c(2,3))  
apply(adsl[,varnum],2,boxplot)  
  
##### Taula descriptiva Global  
Desc1 <- matrix(NA,nrow=6,ncol=3)  
namesvarnum <- c("Velocitat de baixada real","Velocitat de baixada contractada",  
"Velocitat de pujada real","Velocitat de pujada  
contractada",  
"Latència","Distància a la central")  
namesDes <- c("n","mitjana","desviació")  
rownames(Desc1) <- namesvarnum  
colnames(Desc1) <- namesDes  
  
for (i in 1:6){  
  Desc1[i,1] <- sum(!is.na(adsl[,varnum[i]]))  
  Desc1[i,2] <- mean(adsl[,varnum[i]],na.rm=TRUE)  
  Desc1[i,3] <- sd(adsl[,varnum[i]],na.rm=TRUE)  
}  
print(Desc1)
```

```
##### Variables categòriques #####

##### Descriptiva univariant variables categòriques
summary(adsl[,varfact])
Table <- apply(adsl[,varfact],2,table)
P <- list()
par(mfrow=c(2,2))

for (i in 1:4){
  P[[i]] <- prop.table(Table[[i]])*100
  barplot(P[[i]],las=2)
}
print(P)

#####
### Anàlisi principal
#####

### Boxplot de les dues variables que intervenen en la regressió
par(mfrow=c(2,2),mar=c(1.5,4,3,1))
boxplot(adsl$down.speed,main="VBR",xlab="",ylab="Kb/s")
boxplot(log(adsl$down.speed),main="log(VBR)",xlab="",ylab="log(Kb/s)")
boxplot(adsl$veloc.cont.down,main="VBC",xlab="",ylab="Kb/s")
boxplot(log(adsl$veloc.cont.down),main="log(VBC)",xlab="",ylab="log(Kb/s)")

##### Regressió lineal #####
mod.lm1 <- lm(down.speed~veloc.cont.down,data=adsl)
summary(mod.lm1)
confint(mod.lm1)

### Gràfic
plot(adsl$veloc.cont.down,adsl$down.speed,pch=19,cex=1.8,xlab="VBC
(Kb/s)",ylab="VBR (Kb/s)")
abline(mod.lm1$coef[1],mod.lm1$coef[2],col=2,lwd=2)

### Residus
par(mfrow=c(2,2))
plot(mod.lm1, which = c(1:2,5))
resi <- mod.lm1$res
resi0 <- resi[1:(length(resi)-1)]
resil <- resi[2:length(resi)]
plot(resi0,resil,cex=0.8,xlab=expression(Residuals[i]),ylab=expression(Residuals
[i+1]))
mtext("Residuals Independence",3,0.2)
abline(h=0,col="grey",lty=2)

##### Regressió lineal amb transformació logarítmica a les dues variables #####
mod.lm2 <- lm(adsl$log.obs.down~log(veloc.cont.down),data=adsl)
summary(mod.lm2)
confint(mod.lm2)

### Gràfic
par(mfrow=c(1,1))
plot(log(adsl$veloc.cont.down),adsl$log.obs.down,pch=19,cex=1.8,xlab="log(VBC)
(log[Kb/s])",ylab="log(VBR) (log[Kb/s])")
points(log(adsl$veloc.cont.down)[3],adsl$log.obs.down[3],col=2,pch=19,cex=1.8)
abline(mod.lm2$coef[1],mod.lm2$coef[2],col=2,lwd=2)

### Residus
par(mfrow=c(2,2))
plot(mod.lm2,which = c(1:3,5))
```



```

resi <- mod.lm2$res
resi0 <- resi[1:(length(resi)-1)]
resi1 <- resi[2:length(resi)]
plot(resi0,resi1,cex=0.8,xlab=expression(Residuals[i]),ylab=expression(Residuals
[i+1]))
mtext("Residuals Independence",3,0.2)
abline(h=0,col="grey",lty=2)

### Treure el individu 3?
mod.lm3 <- lm(adsl$log.obs.down[-3]~log(veloc.cont.down[-3]),data=adsl)
summary(mod.lm3)

### Gràfic
par(mfrow=c(1,1))
plot(log(adsl$veloc.cont.down)[-3],adsl$log.obs.down[-
3],pch=19,cex=1.8,xlab="log(VBC) (log[Kb/s])",ylab="log(VBR) (log[Kb/s])")
abline(mod.lm3$coef[1],mod.lm3$coef[2],col=2,lwd=2)

```