

Estadística Descriptiva Univariant (EDu)

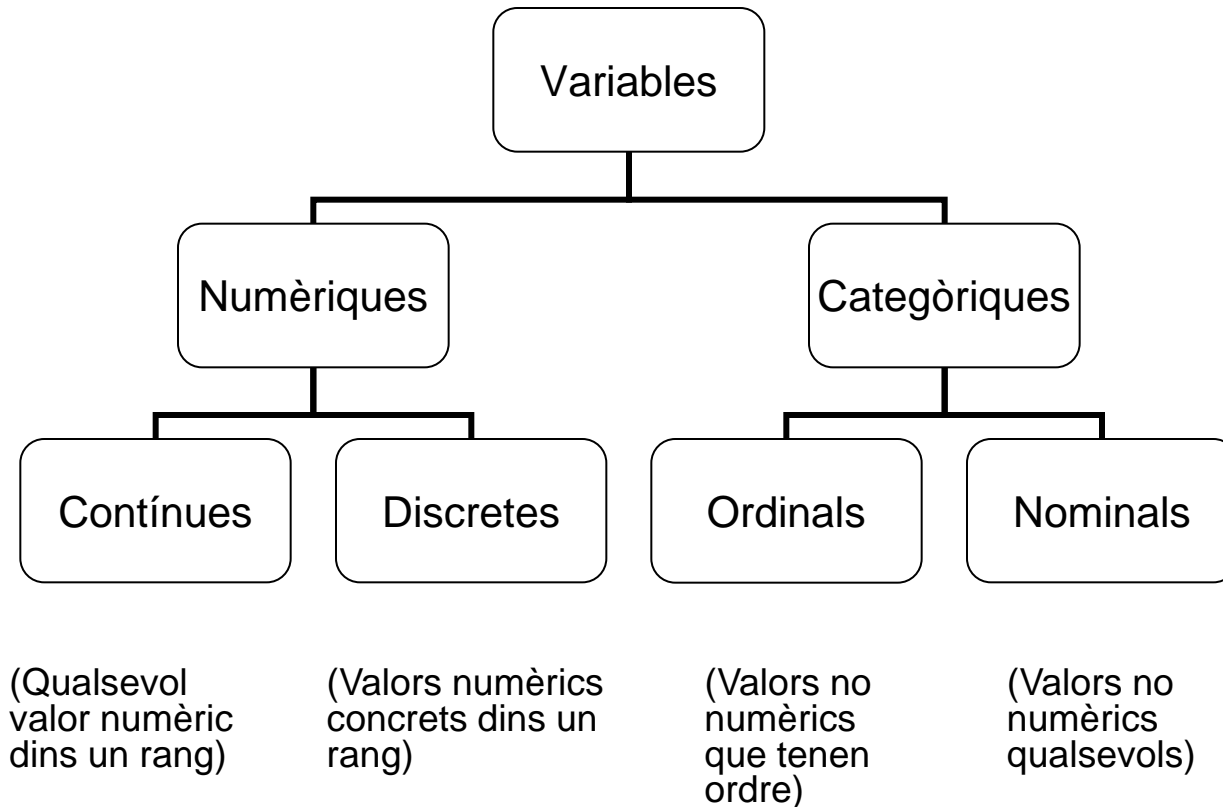
Objectiu: Descriure (estadísticament) ...
... les variables (una mostra o vector de valors) ...
... d'una en una.

Les variables es poden classificar en:

numèriques (contínues o discretes)
i categòriques (ordinals o nominals).

(hi ha diverses notacions per classificar els tipus de variables)

Tipus de variables

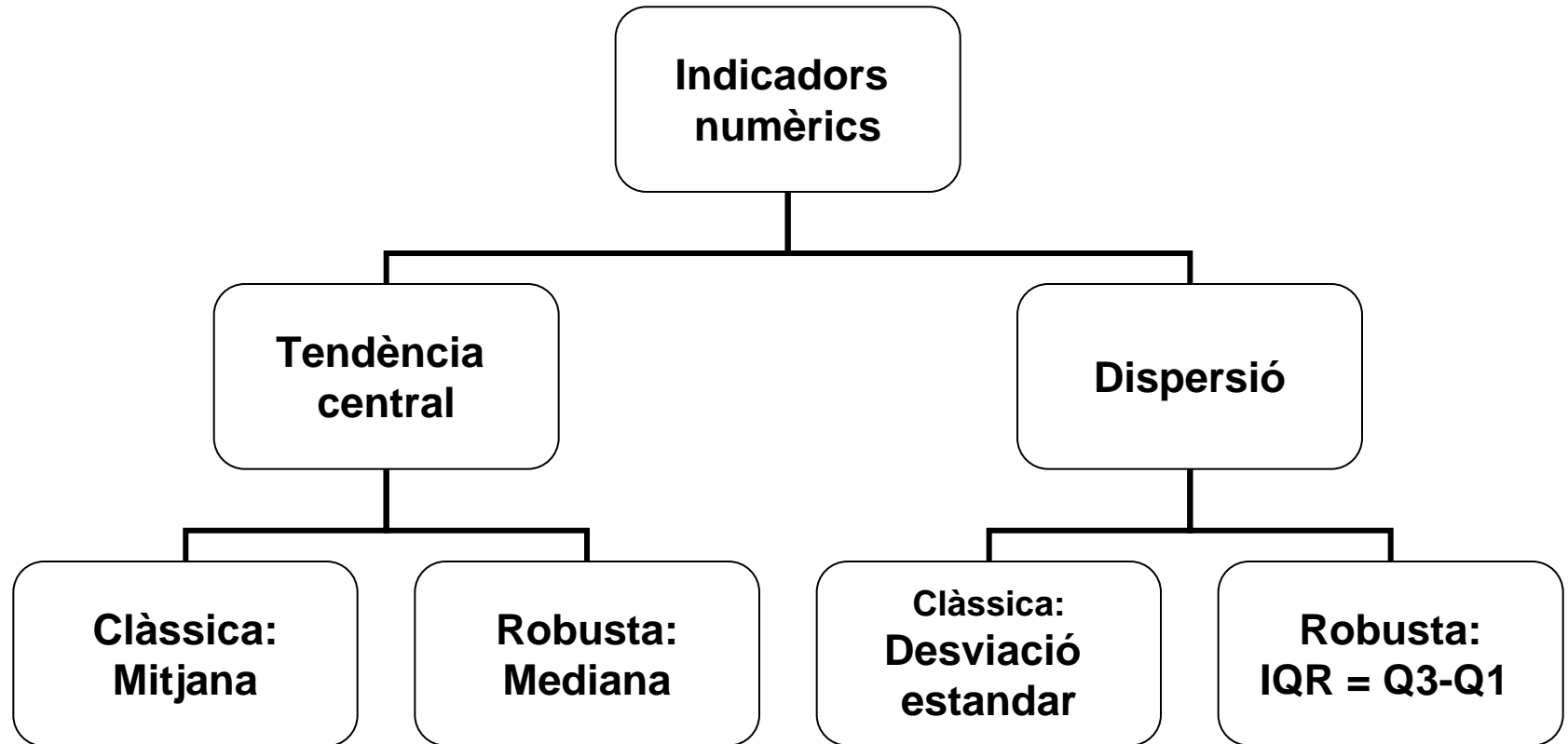


(una mostra de valors pot correspondre's amb més d'un tipus)

Descripció i anàlisi de la mostra

- Depuració de les dades
 - Dades mancants, MISSINGS “*”
 - Dades anormals (Outliers)
- Tractament estadístic
 - EINES NUMÈRIQUES (mitjana, mediana, desviació estandar,..)
 - EINES GRÀFIQUES (dotplot, boxplot, diagrama de barras..)

EINES NUMÈRIQUES



(Sobretot tenen sentit per variables numèriques contínues)

Càlcul dels indicadors clàssics

n = 20

x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
9.50	0.46	0.22
9.50	0.46	0.22
8.55	-0.49	0.24
6.65	-2.39	5.69
9.10	0.06	0.00
8.90	-0.14	0.02
9.75	0.71	0.51
9.15	0.11	0.01
9.75	0.71	0.51
9.35	0.31	0.10
9.80	0.76	0.59
10.00	0.96	0.93
9.15	0.11	0.01
8.90	-0.14	0.02
9.10	0.06	0.00
9.80	0.76	0.59
9.30	0.26	0.07
8.55	-0.49	0.24
6.55	-2.49	6.18
9.35	0.31	0.10

Mitjana

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = 9.04$$

Variancia

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = 0.85$$

$$S_x^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - n(\bar{x})^2 \right]$$

Desviació tipus o estàndar

$$s_x = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = 0.92$$

$$S_x = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - n(\bar{x})^2 \right]}$$

x_i	x_i^2
9,50	90,25
9,50	90,25
8,55	73,10
6,65	44,22
9,10	82,81
8,90	79,21
9,75	95,06
9,15	83,72
9,75	95,06
9,35	87,42
9,80	96,04
10,00	100,00
9,15	83,72
8,90	79,21
9,10	82,81
9,80	96,04
9,30	86,49
8,55	73,10
6,55	42,90
9,35	87,42

Càlcul dels indicadors Robusts

n= 20

x_i	x_i
9.50	6.55
9.50	6.65
8.55	8.55
6.65	8.55
9.10	8.90
8.90	8.90
9.75	9.10
9.15	9.10
9.75	9.15
9.35	9.15
9.80	9.30
10.00	9.35
9.15	9.35
8.90	9.50
9.10	9.50
9.80	9.75
9.30	9.75
8.55	9.80
6.55	9.80
9.35	10.00

Ordenar



Posició Q1 = $(n+1)/4$, **Q1** = (valor inferior *3 + valor superior)/4

Posició **Me** = $(n+1)/2$, **Q2** = (valor inferior + valor superior)/2

Posició Q3 = $(n+1)*3/4$, **Q3** = (valor inferior + valor superior * 3)/4

IQR = Q3-Q1

	Posició	Resultat
Q1 →	5.25	8.9
ME →	10.5	9.23
Q3 →	15.75	9.69
IQR →		0.79

EINES GRÀFIQUES

- Histograma

(sobretot té sentit per variables numèriques contínues)

- Diagrama de barres

(sobretot té sentit per variables numèriques discretes i per categòriques)

- Pastís

(sobretot té sentit per variables numèriques discretes i per categòriques)

- Box-plot: indicadors ROBUSTS + Outliers

(sobretot té sentit per variables numèriques contínues)

R i algunes dades

- R:

([http:// cran.r-project.org](http://cran.r-project.org))

(Inici_R.ppt)

- Exemple cas Adsl

Dades sobre la potència de la connexió Adsl domèstica (recollides per estudiants)
a analitzar estadísticament

Cada alumne participant va recollir, entre altra, la següent informació:

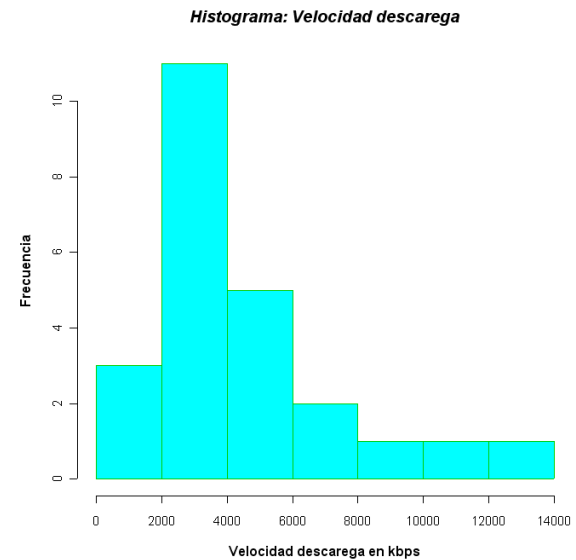
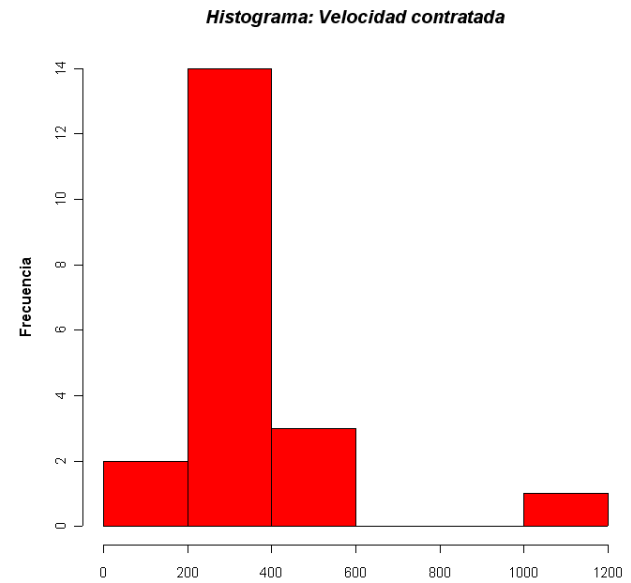
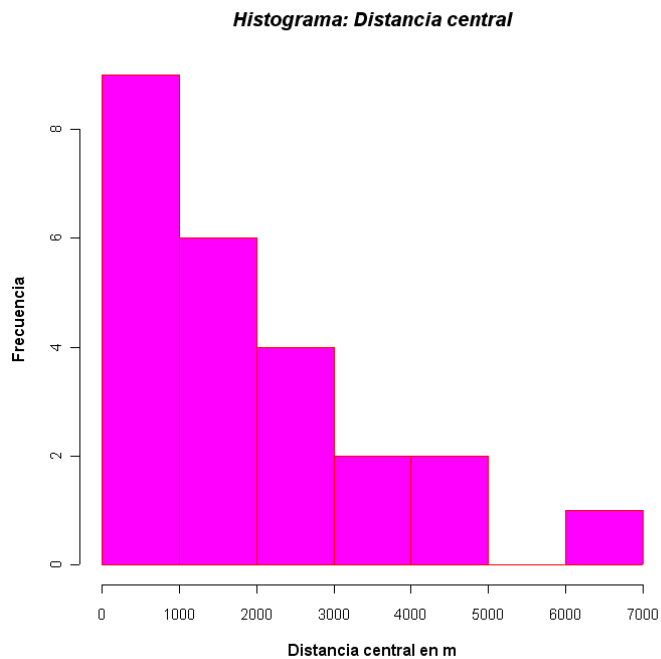
(*) proveïdor ADSL

(*) velocitats contractades (pujada/baixada) en Mbps

(*) velocitat real (<http://www.internautas.org/testvelocidad/>)

(*) distància a la central (a <http://www.adslnet.es/distancia-adsl>)

Ex. histograma (cas ADSL)



Ex. diagrama de barres (cas ADSL)

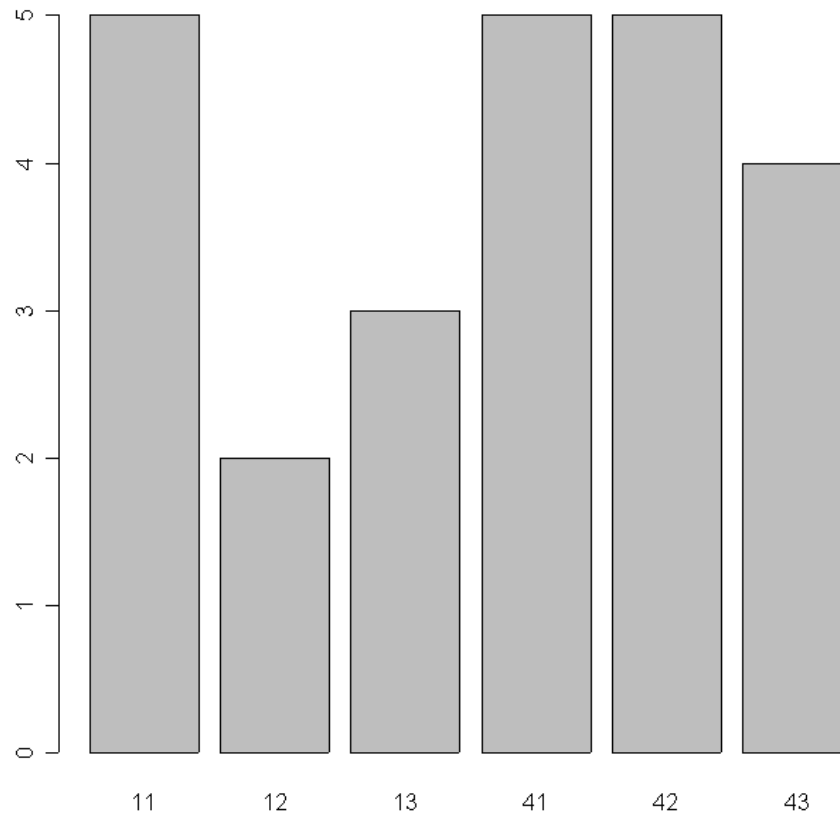
#Distribució dels participants en els 6 grups

```
> table(adsl$grp)
```

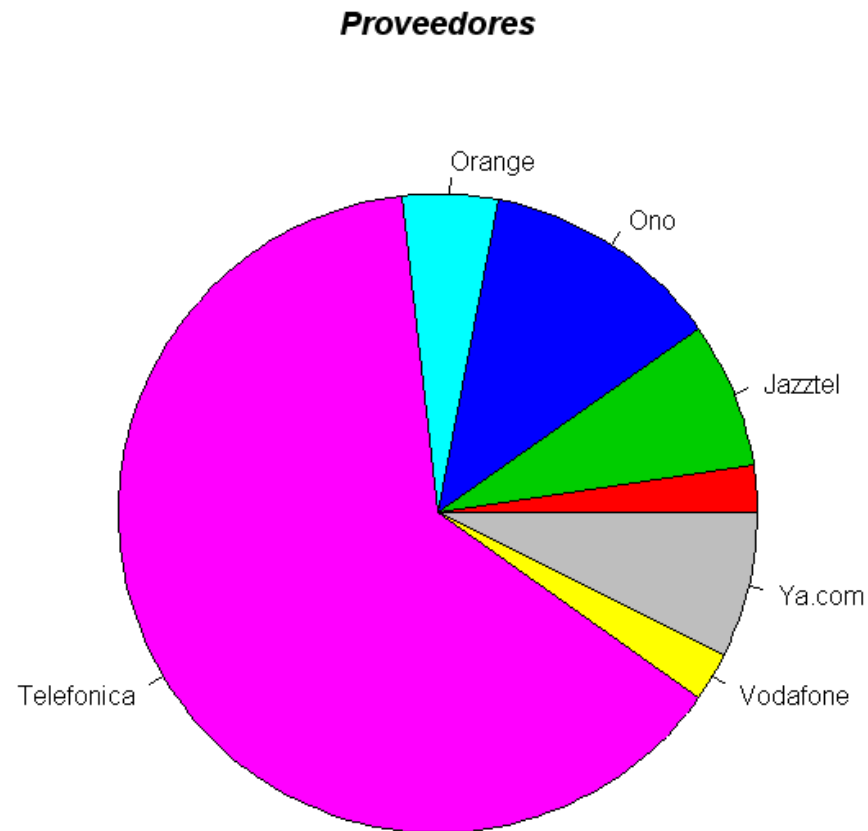
```
11 12 13 41 42 43
```

```
5  2  3  5  5  4
```

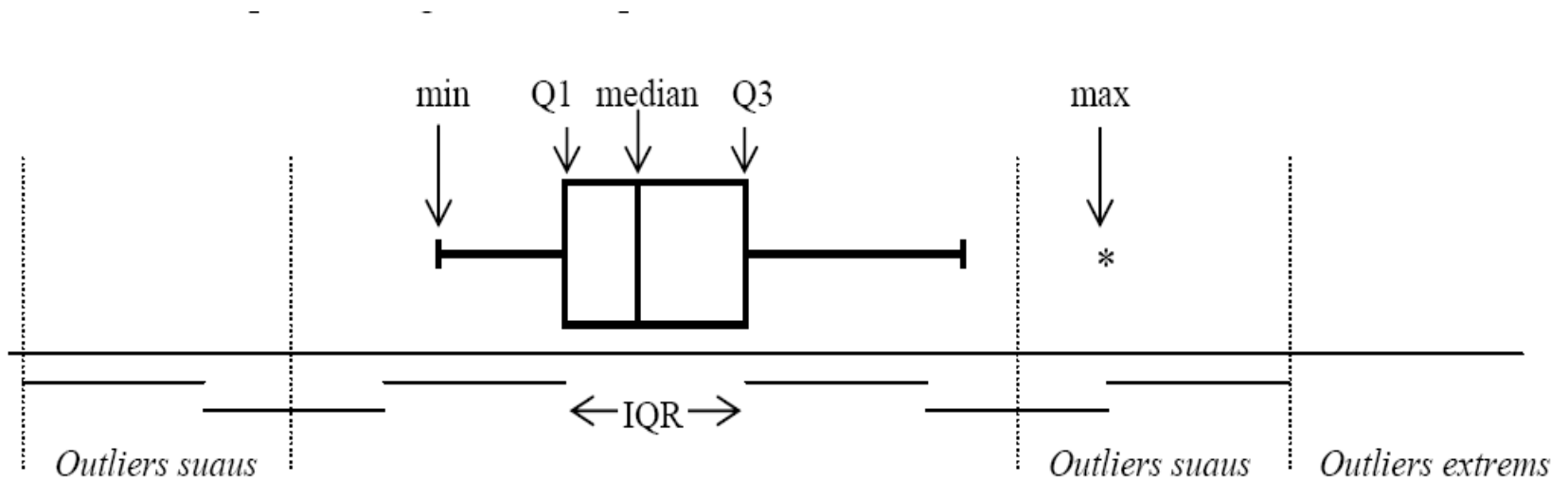
```
> barplot(table(adsl$grp))
```



Ex. pastís (cas ADSL)



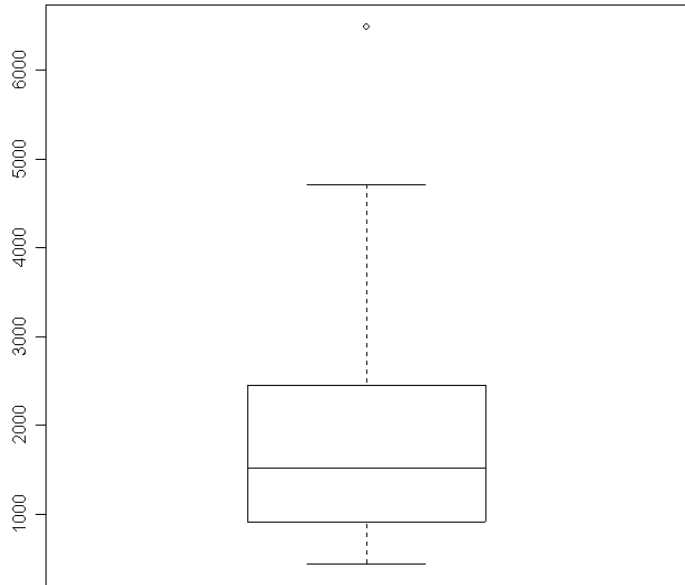
El Box-plot



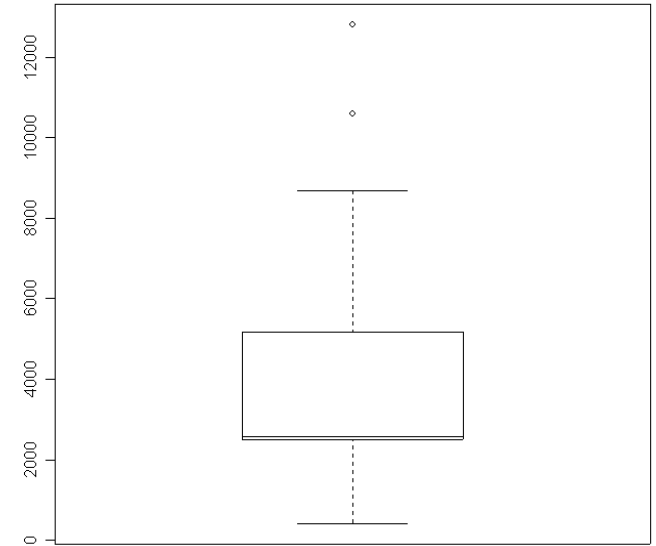
box-plot: indicadores ROBUSTS + Outliers

Ex. Box-plot (cas ADSL)

Boxplot: Distancia central



Boxplot: Velocitat descarrega



Boxplot: Velocitat contractada

