

Variabes Estudiantines (examen de contrôle continu depuis 2003)

A.B. Dufour & J.R. Lobry

Une initiation de niveau L3 visant à familiariser les étudiants aux problèmes concrets rencontrés lors d'une analyse statistique dans la vraie vie : vous allez analyser vos propres données (Yes!). On part de la mesure et de la saisie des données sur la promotion actuelle, puis, via l'importation des données des *maudites* promotions précédentes, on rédige un rapport **reproductible** sur une problématique donnée. Le choix de la problématique est libre à ce niveau, l'originalité est encouragée (si elle est reproductible).

1 Le fichier de départ

```
library(MASS)
data(survey)
str(survey)
'data.frame':      237 obs. of  12 variables:
 $ Sex      : Factor w/ 2 levels "Female","Male": 1 2 2 2 2 1 2 1 2 2 ...
 $ Wr.Hnd   : num  18.5 19.5 18 18.8 20 18 17.7 17 20 18.5 ...
 $ NW.Hnd   : num  18 20.5 13.3 18.9 20 17.7 17.7 17.3 19.5 18.5 ...
 $ W.Hnd    : Factor w/ 2 levels "Left","Right": 2 1 2 2 2 2 2 2 2 ...
 $ Fold     : Factor w/ 3 levels "L on R","Neither",...: 3 3 1 3 2 1 1 3 3 3 ...
 $ Pulse    : int   92 104 87 NA 35 64 83 74 72 90 ...
 $ Clap     : Factor w/ 3 levels "Left","Neither",...: 1 1 2 2 3 3 3 3 3 3 ...
 $ Exer     : Factor w/ 3 levels "Freq","None",...: 3 2 2 2 3 3 1 1 3 3 ...
 $ Smoke    : Factor w/ 4 levels "Heavy","Never",...: 2 4 3 2 2 2 2 2 2 ...
 $ Height   : num  173 178 NA 160 165 ...
 $ M.I      : Factor w/ 2 levels "Imperial","Metric": 2 1 NA 2 2 1 1 2 2 2 ...
 $ Age      : num  18.2 17.6 16.9 20.3 23.7 ...

summary(survey)
      Sex      Wr.Hnd      NW.Hnd      W.Hnd      Fold
Female:118  Min.   :13.00  Min.   :12.50  Left  : 18  L on R : 99
Male  :118  1st Qu.:17.50  1st Qu.:17.50  Right:218  Neither: 18
NA's  : 1    Median :18.50  Median :18.50  NA's  : 1   R on L :120
      Mean   :18.67  Mean   :18.58
      3rd Qu.:19.80  3rd Qu.:19.73
      Max.   :23.20  Max.   :23.50
      NA's   : 1.00  NA's   : 1.00

      Pulse      Clap      Exer      Smoke      Height
Min.   : 35.00  Left  : 39  Freq:115  Heavy: 11  Min.   :150.0
1st Qu.: 66.00  Neither: 50  None: 24  Never:189  1st Qu.:165.0
Median : 72.50  Right  :147  Some: 98  Occas: 19  Median :171.0
Mean   : 74.15  NA's   : 1    Regul: 17  Mean   :172.4
3rd Qu.: 80.00  NA's   : 1    NA's  : 1  3rd Qu.:180.0
Max.   :104.00  NA's   : 1    Max.   :200.0
```

```

NA's : 45.00
M.I      Age
Imperial: 68  Min. :16.75
Metric :141  1st Qu.:17.67
NA's : 28   Median :18.58
          Mean  :20.37
          3rd Qu.:20.17
          Max. :73.00
NA's : 28.0

```



Adelaide, Australie

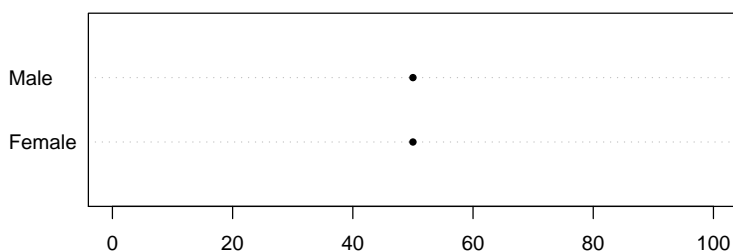
Ce jeu de données `survey` ([1]) contient les réponses à 12 questions de 237 étudiant(e)s en statistiques à l'Université d'Adelaide en Australie :

Sex Le sexe des étudiant(e)s. Variable qualitative nominale à deux modalités Female et Male.

```

attach(survey)
par(mar = c(2, 4.1, 0.1, 0.1))
dotchart(100 * table(Sex)/sum(table(Sex)), xlim = c(0, 100), pch = 20)

```

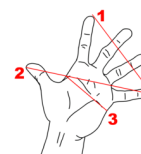
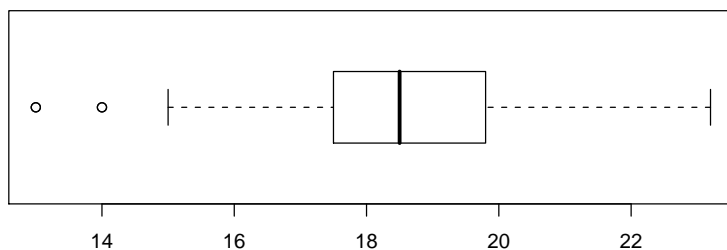


Wr.Hnd L'empan (distance entre l'extrémité du pouce et l'extrémité de l'auriculaire, doigts écartés au maximum, le poignet bien à plat sur le plan de mesure) de la main d'écriture, en centimètres.

```

par(mar = c(2, 0.1, 0.1, 0.1))
boxplot(Wr.Hnd, horizontal = TRUE)

```



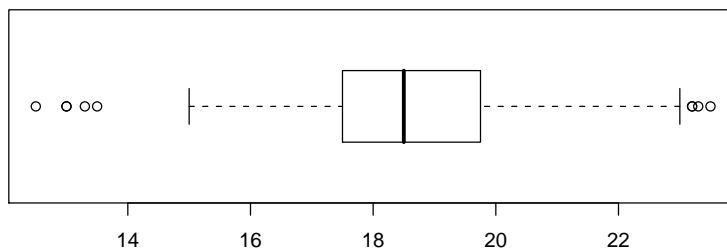
Les mesures de la main : 1) Palme 2) Empan 3) Paume (source : Wikipedia)

NW.Hnd Empan de l'autre main.

```

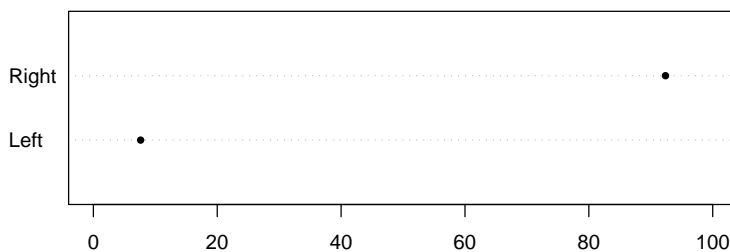
par(mar = c(2, 0.1, 0.1, 0.1))
boxplot(NW.Hnd, horizontal = TRUE)

```



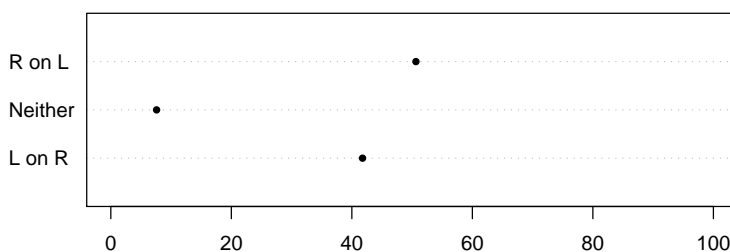
W.Hnd Main d'écriture. Variable qualitative nominale à deux modalités Left et Right.

```
par(mar = c(2, 4.1, 0.1, 0.1))
dotchart(100 * table(W.Hnd)/sum(table(W.Hnd)), xlim = c(0, 100),
pch = 20)
```



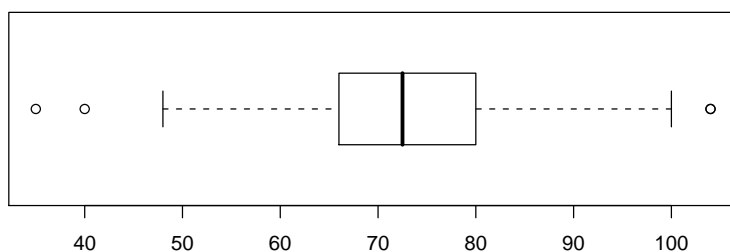
Fold Croisez vos bras ! Lequel est au-dessus ? Variable qualitative à trois modalités R on L, L on R, Neither.

```
par(mar = c(2, 4.1, 0.1, 0.1))
dotchart(100 * table(Fold)/sum(table(Fold)), xlim = c(0, 100), pch = 20)
```



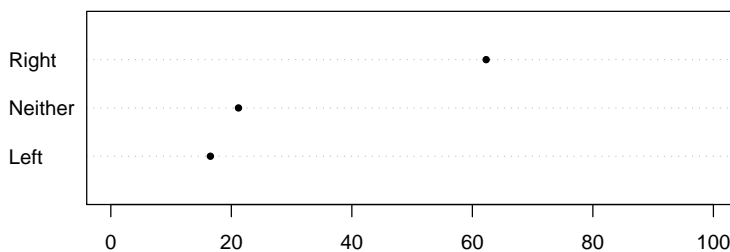
Pulse Rythme cardiaque au repos en pulsations par minute.

```
par(mar = c(2, 0.1, 0.1, 0.1))
boxplot(Pulse, horizontal = TRUE)
```



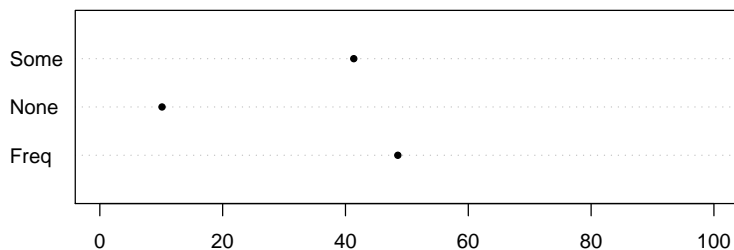
Clap Applaudissez ! Quelle main est au-dessus ? Variable qualitative nominale à trois modalités Right, Left, Neither.

```
par(mar = c(2, 4.1, 0.1, 0.1))
dotchart(100 * table(Clap)/sum(table(Clap)), xlim = c(0, 100), pch = 20)
```



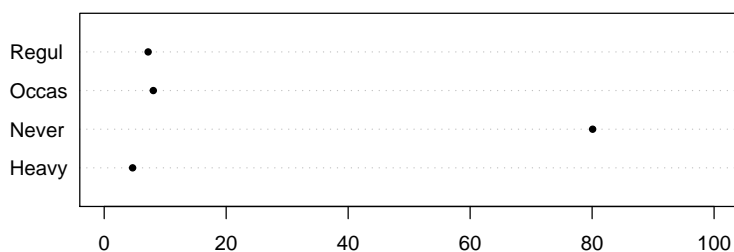
Exer Fréquence des activités physiques et sportives. Variable qualitative ordonnée à trois modalités Freq (fréquemment), Some, None.

```
par(mar = c(2, 4.1, 0.1, 0.1))
dotchart(100 * table(Exer)/sum(table(Exer)), xlim = c(0, 100), pch = 20)
```



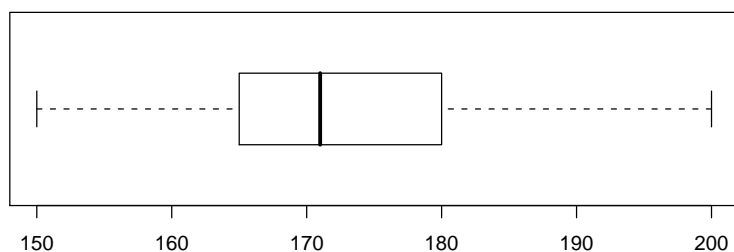
Smoke Intensité de l'addiction au tabac. Variable qualitative ordonnée à quatre modalités Heavy, Regul, Occas, Never.

```
par(mar = c(2, 4.1, 0.1, 0.1))
dotchart(100 * table(Smoke)/sum(table(Smoke)), xlim = c(0, 100),
pch = 20)
```



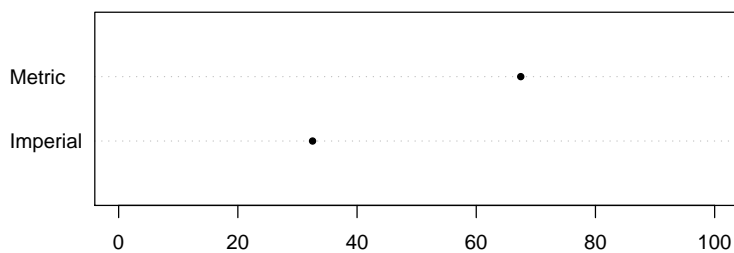
Height Taille en centimètres.

```
par(mar = c(2, 0.1, 0.1, 0.1))
boxplot(Height, horizontal = TRUE)
```



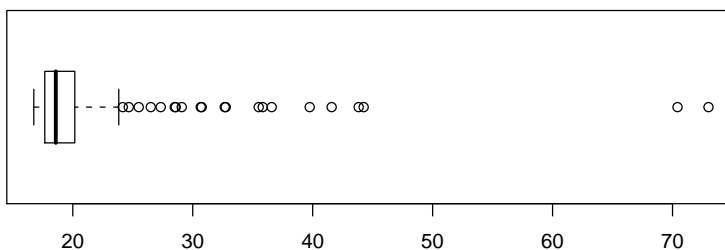
M.I Indique si l'étudiant(e) exprime sa taille en unités vernaculaires (pieds/pouces) ou universelles (mètres). Variable qualitative nominale à deux modalités Metric, Imperial.

```
par(mar = c(2, 4.1, 0.1, 0.1))
dotchart(100 * table(M.I)/sum(table(M.I)), xlim = c(0, 100), pch = 20)
```



Age Âge des étudiant(e)s en années.

```
par(mar = c(2, 0.1, 0.1, 0.1))
boxplot(Age, horizontal = TRUE)
detach(survey)
```



2 Acquisition de données

Complétez le jeu de données précédent en incluant les réponses de tous les étudiants de votre promotion. Attention, la version électronique de ce jeu de données ne doit en aucun cas, et en aucune façon, contenir des informations nominatives (il doit être impossible d'identifier les individus). Une version au format texte de ce jeu de données sera disponible sur le site du pbil : <http://pbil.univ-lyon1.fr/R/donnees/>. Vous disposez également des données des promotions des années précédentes à partir de 2003 :

```
s2003 <- read.table("http://pbil.univ-lyon1.fr/R/donnees/survey2003.txt",
  header = TRUE, dec = ",")
head(s2003)
```

	sexe	WrHnd	NWHnd	WHnd	Fold	Pulse	Clap	Exer	Smoke	Height	MI	Age
1	F	20.0	20.5	R	RonL	72	R	Some	Never	170	M	24.100
2	M	23.0	22.0	R	RonL	86	neither	None	Heavy	182	M	20.900
3	M	23.0	22.0	R	LonR	74	neither	None	Heavy	180	M	24.990
4	F	20.2	20.2	R	LonR	68	R	Some	Never	172	M	21.006
5	M	23.0	23.0	R	RonL	76	neither	Freq	Never	194	M	21.000
6	F	19.5	19.3	R	LonR	80	neither	Some	Never	173	M	20.500

Vous devrez donc créer un data frame comportant les données de `survey` et celles de Lyon, avec en plus une variable qualitative indiquant l'origine australienne ou française de l'étudiant(e).

3 Consignes

Tous documents, papiers, informatiques, calculatrices, neurones, etc... autorisés. Les instructions détaillées vous seront données en séance.

Références

- [1] W.N. Venables and B.D. Ripley. *Modern applied statistics with S-PLUS. Third edition.* Springer-Verlag, New York, 1997.