

# Questionnaire étudiant

*Sebri, JcB*

*19/02/2015*

## Contents

<b>Questionnaire étudiant</b>	<b>1</b>
Etablissements participant: . . . . .	1
Age . . . . .	2
Sexe . . . . .	3
Age et sexe . . . . .	3
Q1- Pour ce cours, vous avez pris des notes . . . . .	5
Q2- Pendant ce cours, vous avez complété la prise de notes par (plusieurs réponses possibles) . . .	5
Q3- Quels sont les outils numériques que vous aviez avec vous pendant ce cours? (plusieurs réponses possibles) . . . . .	6
Q4- Pendant ce cours (en dehors des temps de pause éventuels), vous avez utilisé votre téléphone pour (plusieurs réponses possibles): . . . . .	7
<b>Information de session</b>	<b>8</b>

## Questionnaire étudiant

```
## [1] "Etab"      "Etud"      "Q1"        "Q2.1"      "Q2.2"      "Q2.3"      "Q2.4"
## [8] "Q2.5"      "Q2.6"      "Q2.7"      "Q3.1tpc"   "Q3.2sp"    "Q3.3tab"   "Q3.4ord"
## [15] "Q4.1"      "Q4.2"      "Q4.3"      "Q4.4"      "Q4.5"      "Q4.6"      "Q4.7"
## [22] "Q4.8"      "Q4.9"      "Q4.10"     "Q4.11"     "Q4.12"     "Q4.13"     "Q4.14"
## [29] "Q4.15"     "Q4.16"     "Q5"        "Q6"        "Q7.1"      "Q7.2"      "Q7.3"
## [36] "Q7.4"      "Q7.5"      "Q7.6"      "Q7.7"      "Q7.8"      "Q7.9"      "Q7.10"
## [43] "Q7.11"     "Q7.12"     "Q7.13"     "Q7.14"     "Q7.15"     "Q7.16"     "Q8"
## [50] "Q9"        "Q10"       "Q11"
```

```
## Loading required package: foreign
## Loading required package: survival
## Loading required package: MASS
## Loading required package: nnet
```

Le fichier comporte:

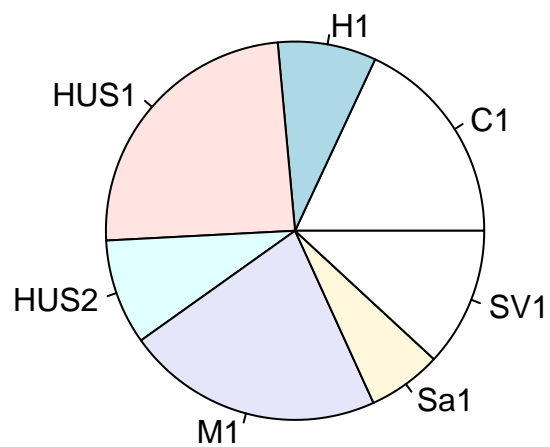
- 665 lignes
- 52 variables

**Etablissements participant:**

```
s <- summary(as.factor(d1$Etab))
s
```

```
##      C1      H1 HUS1 HUS2      M1      Sa1      SV1
##  120    56   162    60   146    42    79
```

```
pie(s)
```



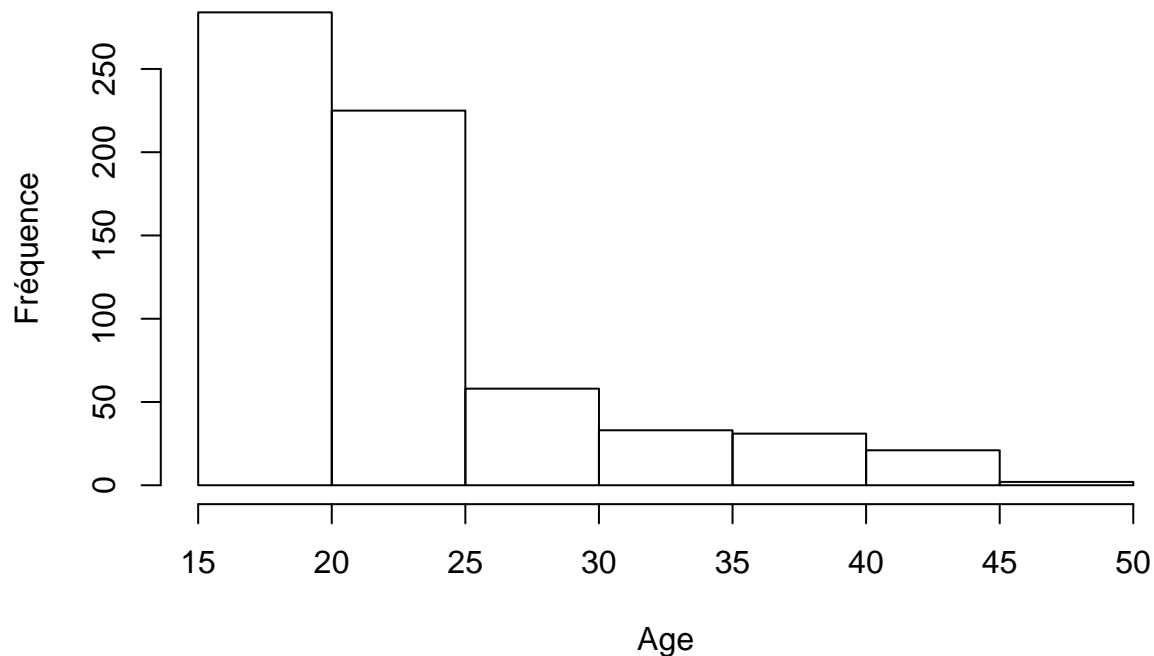
Age

```
summary(d1$Q11)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
##  17.00   20.00   21.00   23.48   25.00   48.00     11
```

```
hist(d1$Q11, main = "Histogramme de l'age", xlab = "Age", ylab = "Fréquence")
```

## Histogramme de l'age



## Sexe

```
summary(as.factor(d1$Q10))
```

```
##      F      H   NR NA's  
##  549  108     3     5
```

## Age et sexe

L'âge des hommes et des femmes sont-ils identiques ? On part de l'hypothèse qu'il n'y a à priori de différence d'âge entre les hommes et les femmes (on appelle cela l'hypothèse nulle ou  $H_0$ ). Si cette hypothèse est vraie, la différence des moyennes des âges entre les hommes et les femmes devrait être nulle. En pratique cette différence est rarement exactement égale à 0 et le problème est de savoir si le chiffre obtenu est assimilable à 0 ou si au contraire il est trop important pour qu'on puisse se livrer à cette assimilation, auquel cas on est obligé de renoncer à l'hypothèse nulle et accepter l'hypothèse alternative: l'âge des hommes est en moyenne différent de celui des femmes. Pour répondre à la question, on pratique un test statistique pour lequel on définit un écart par rapport à 0. Si le résultat du test tombe dans l'intervalle on admet que la différence de moyenne est assimilable à 0 et on accepte l'hypothèse nulle: pas de différence entre les groupes. Sinon on la rejette. Bien sûr, plus on définit un intervalle important, plus on augmente le risque de se tromper en affirmant qu'il n'y a pas de différence entre les moyennes. C'est ce qu'on appelle le risque de première espèce ou  $\alpha$ . Dans les sciences de la santé, ce risque est fixé consensuellement (et arbitrairement) à  $5\% = 5/100 = 0.05$  et généralement rapporté sous la forme  $p = 0.05$ . C'est à dire que j'admet  $H_0$  (pas de différence) en prenant le risque consenti de me tromper dans 5% des cas. En pratique les logiciels calculent la probabilité exacte d'observer par hasard une telle différence entre les deux groupes. Si cette probabilité est supérieure à 0.05 (c'est-à-dire comprise entre 0.05 et 1) on considère que la différence entre les moyennes est un artefact lié à la fluctuation d'échantillonnage et qu'en réalité il n'y a pas de différence entre les groupes. Si au contraire, la

probabilité exacte est inférieure à 0.05, on admet qu'elle n'est pas due au hasard et on est obligé d'admettre qu'il y a bien une différence entre les deux groupes. On voit par là le côté arbitraire du petit p, mais il est considéré dans toutes les publications comme un chiffre magique...

Il existe de nombreux tests statistiques. Pour répondre à la question posée, on utilise le test t de Student qui s'applique si:

- on ne compare que 2 groupes (c'est le cas)
- la variable d'intérêt (ici l'âge) suit une loi normale (on va admettre que oui) dans les 2 groupes
- la variance (moyenne des écarts à la moyenne) des 2 groupes est égale (si ce n'est pas le cas, on peut utiliser une variante de test de Student appelée test de Welch).

La colonne sexe (Q10) comporte 3 valeurs: H, F et NR. Il faut éliminer les NR en les transformant en NA pour rendre le test possible

```
d1$Q10[d1$Q10 == "NR"] <- NA
```

Puis faire le test:

```
t <- t.test(d1$Q11 ~ d1$Q10, var.equal = TRUE)
t
```

```
##
## Two Sample t-test
##
## data: d1$Q11 by d1$Q10
## t = -2.7498, df = 650, p-value = 0.006129
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.1159677 -0.5197362
## sample estimates:
## mean in group F mean in group H
##      23.15385      24.97170
```

On voit que la probabilité exacte d'observer par hasard une telle différence entre les moyennes est égale à 0.0061288. Cette probabilité est très inférieure à 0.05 et donc on rejette l'hypothèse d'égalité des âges. En moyenne, pour cet échantillon, les étudiants hommes sont plus âgés que les étudiantes et cette différence est statistiquement significative.

Comme on peut avoir un doute sérieux sur la normalité de l'âge (voir le graphique des âges ci-dessus), on réalise un test non paramétrique, c'est à dire qui ne fait pas d'hypothèse sur la façon dont la variable est distribuée. Dans le cas particulier on utilise le test de Wilcoxon qui est l'équivalent non paramétrique du test de Student:

```
wilcox.test(d1$Q11 ~ d1$Q10)
```

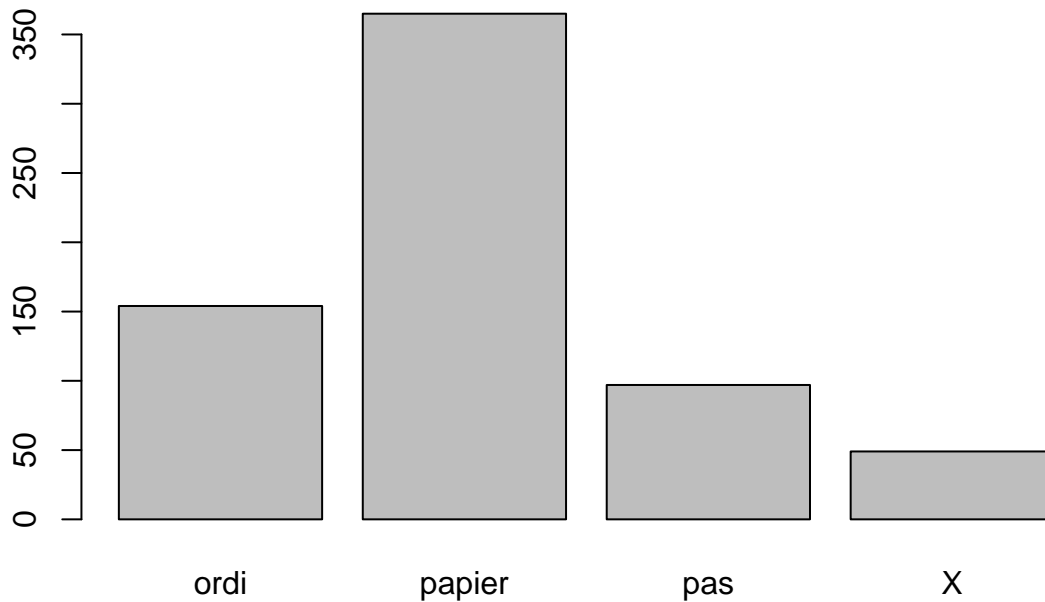
```
##
## Wilcoxon rank sum test with continuity correction
##
## data: d1$Q11 by d1$Q10
## W = 21322.5, p-value = 1.487e-05
## alternative hypothesis: true location shift is not equal to 0
```

On arrive à la même conclusion.

**Q1- Pour ce cours, vous avez pris des notes**

ordi	papier	pas	X
154	365	97	49

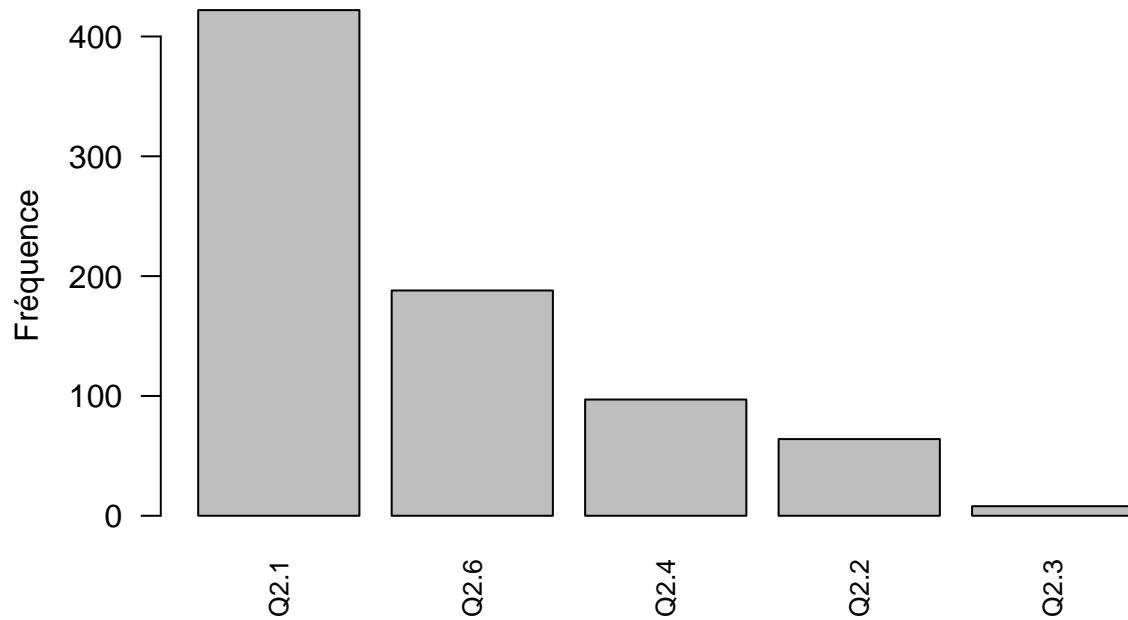
**Support de notes utilisé par l'étudiant**



**Q2- Pendant ce cours, vous avez complété la prise de notes par (plusieurs réponses possibles)**

La variable Q2.5 est anormale. Il ne peut y avoir dans la même colonne du texte et des nombres. La colonne ne peut contenir que 1 ou NA. Créer une colonne supplémentaire pour le texte. Par ex. Q2-7.

### Compléments de la prise de notes

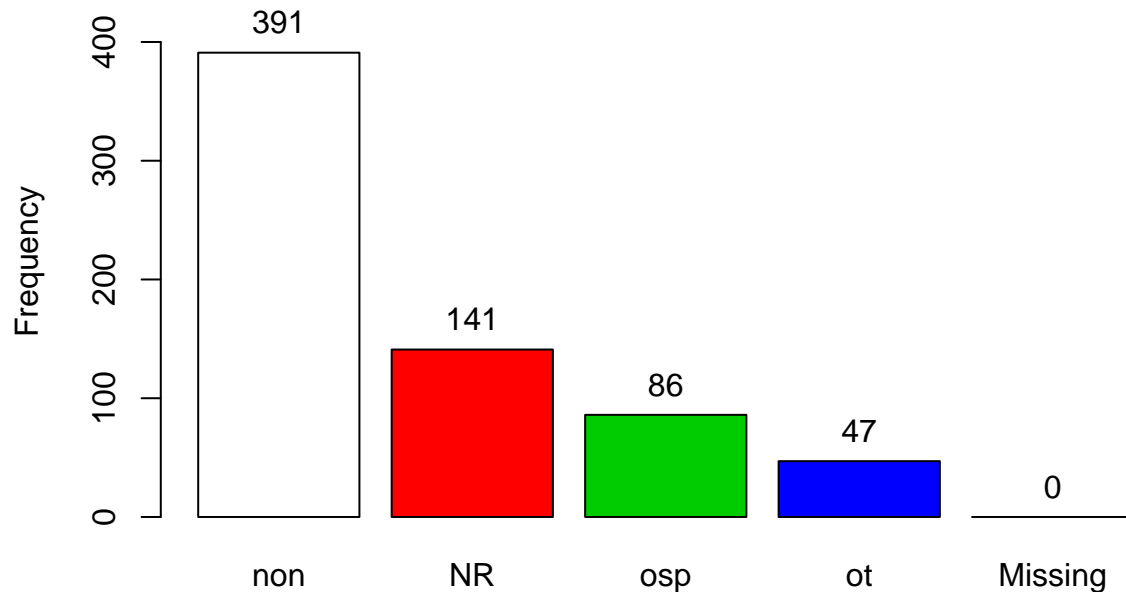


**Q3- Quels sont les outils numériques que vous aviez avec vous pendant ce cours?**  
(plusieurs réponses possibles)

Colonne 11 à 14

**téléphone portable classique** colonnes 10: ( ):non, oui sur la table= ot, oui dans mon sac ou ma poche= osp

### Téléphone portable classique, ylab = Fréquence



```
d1$Q3.1tpc :
      Frequency Percent Cum. percent
non           391    58.8         58.8
NR            141    21.2         80.0
osp            86    12.9         92.9
ot             47     7.1        100.0
Total         665   100.0        100.0
```

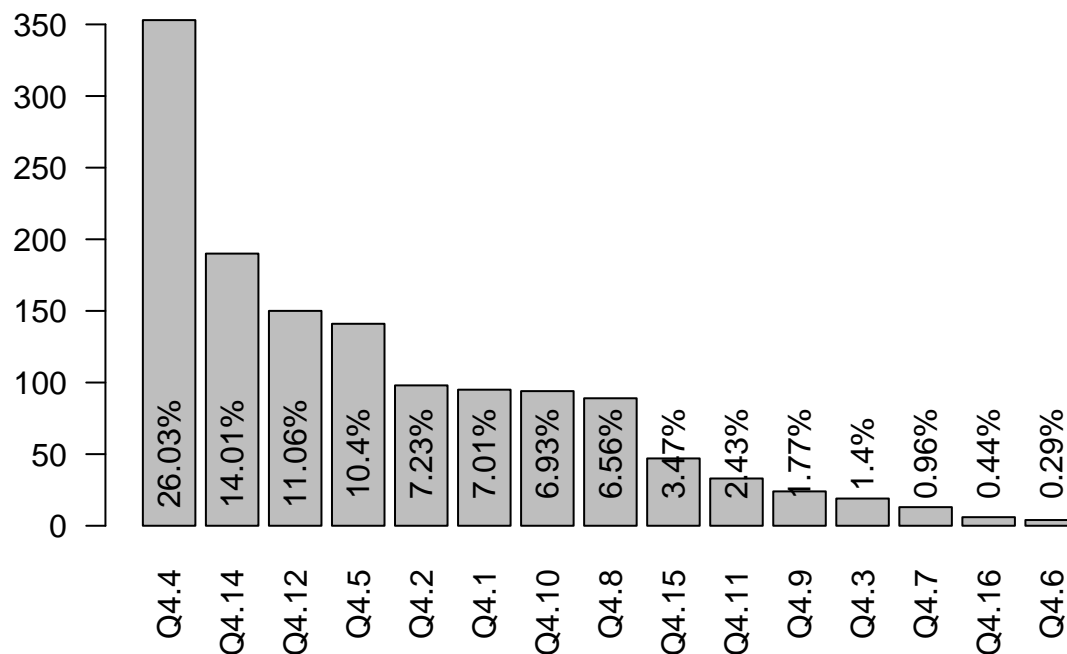
**Q4- Pendant ce cours (en dehors des temps de pause éventuels), vous avez utilisé votre téléphone pour (plusieurs réponses possibles):**

question 15 à 30

```
d1 <- read.csv(paste0(path, file1), skip = 1, stringsAsFactors = FALSE)

q4 <- d1[, c(15:26, 28:30)]
q4 <- as.data.frame(sapply(q4,gsub,pattern="NR",replacement="NA"), , stringsAsFactors = FALSE)
q4 <- as.data.frame(sapply(q4, as.integer))
a <- apply(q4,2,sum, na.rm = TRUE)
x <- barplot(sort(a, decreasing = TRUE), las = 2, main = "Utilisation du téléphone pendant le cours")
v <- paste0(sort(round(a*100/sum(a), 2), decreasing = TRUE), "%")
text(x, 50, v, srt=90)
```

### Utilisation du téléphone pendant le cours



Combien d'actions simultanément:

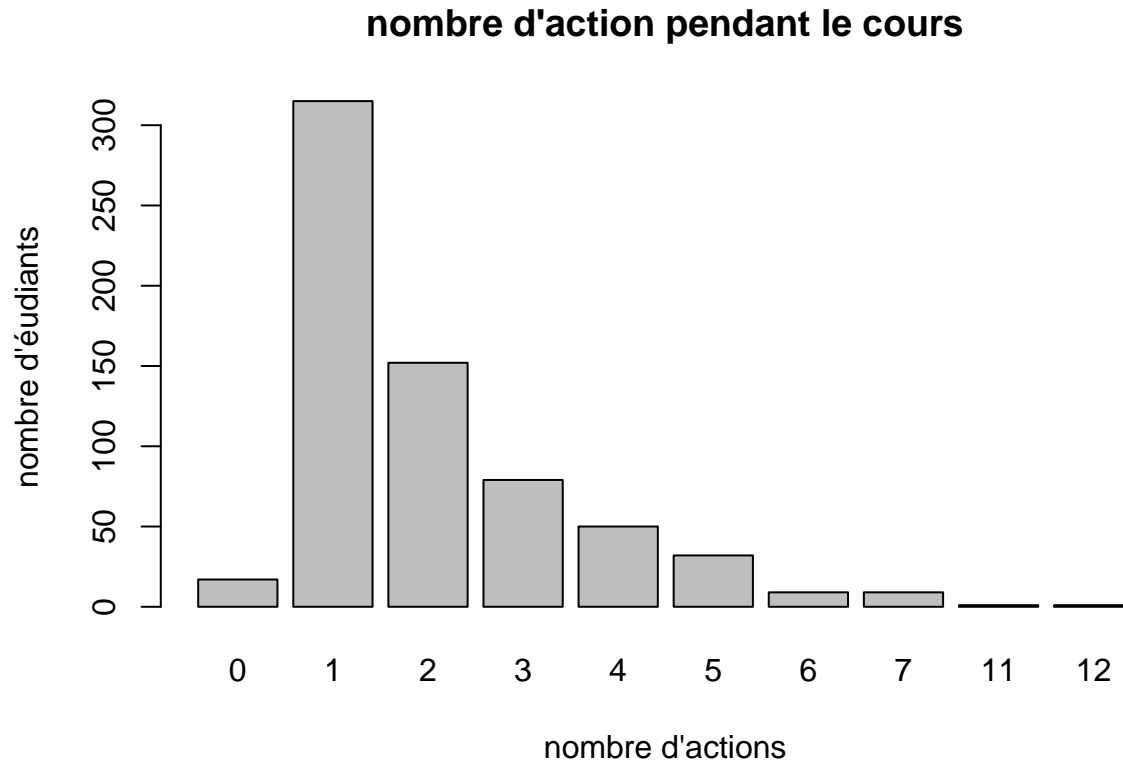
```
a <- apply(q4,1,sum, na.rm = TRUE)
summary(a)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.000   1.000   2.000   2.039   3.000   12.000
```

```
summary(as.factor(a))
```

```
##  0  1  2  3  4  5  6  7 11 12
## 17 315 152 79 50 32 9 9 1 1
```

```
barplot(summary(as.factor(a)), main = "nombre d'action pendant le cours", ylab = "nombre d'éludiants", xlab = "nombre d'actions")
```



## Information de session

```
## R version 3.1.3 (2015-03-09)
## Platform: x86_64-pc-linux-gnu (64-bit)
## Running under: Ubuntu 14.04.2 LTS
##
## locale:
##  [1] LC_CTYPE=fr_FR.UTF-8      LC_NUMERIC=C
##  [3] LC_TIME=fr_FR.UTF-8      LC_COLLATE=fr_FR.UTF-8
##  [5] LC_MONETARY=fr_FR.UTF-8  LC_MESSAGES=fr_FR.UTF-8
##  [7] LC_PAPER=fr_FR.UTF-8     LC_NAME=C
##  [9] LC_ADDRESS=C             LC_TELEPHONE=C
## [11] LC_MEASUREMENT=fr_FR.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
## [1] epicalc_2.15.1.0 nnet_7.3-9      MASS_7.3-39     survival_2.38-1
```



```
## [5] foreign_0.8-63
##
## loaded via a namespace (and not attached):
## [1] digest_0.6.4    evaluate_0.5.5  formatR_1.0     htmltools_0.2.6
## [5] knitr_1.9       rmarkdown_0.5.1 splines_3.1.3   stringr_0.6.2
## [9] tools_3.1.3     yaml_2.1.13
```