

# Devoir maison du mercredi 8 janvier 2014 : Énoncé

Taper sous R

```
library(car)
data(Mroz)
Mroz
fix(Mroz)
help(Mroz)
```

qui charge les données Mroz, les affiche et affiche leur descriptif. À partir de là, produire un court rapport (format PDF) contenant :

## 1. Une analyse de l'influence du niveau d'études des femmes sur la variable `inc`

Vous commencerez par retirer du jeu de données la valeur de `inc` qui est aberrante (pour cela, il faut regarder la correction du devoir 3 : toute erreur sur cette partie sera lourdement sanctionner dans la notation).

L'analyse contiendra des descriptifs numériques et graphiques, tous commentés.

## 2. Une analyse de l'influence du nombre d'enfants des femmes (nombre total) sur la variable `inc`

L'analyse contiendra des descriptifs numériques et graphiques, tous commentés.

## 3. Les scripts ayant permis ces analyses

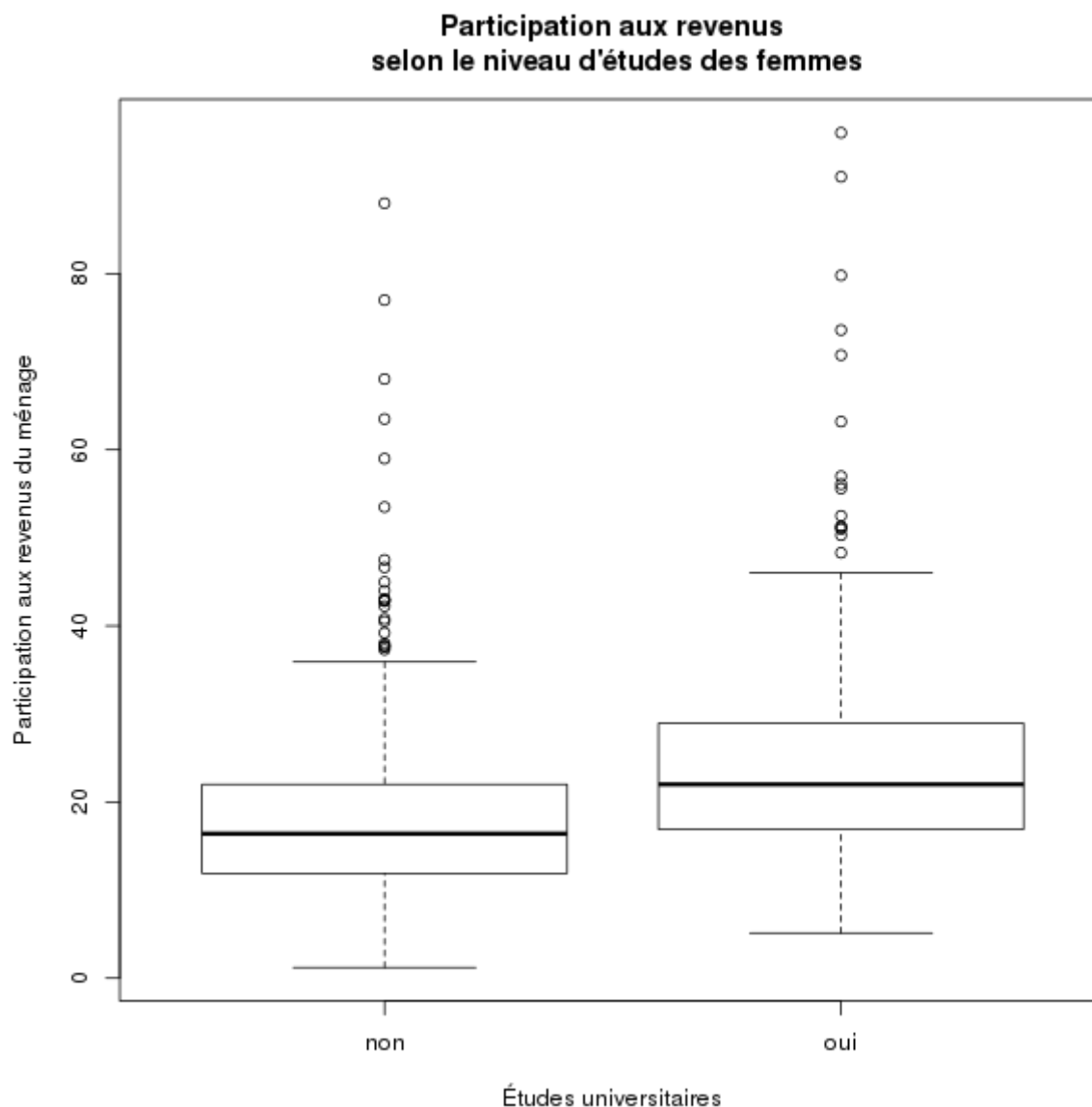
*Remarque : Les graphiques produits peuvent être exportés dans le répertoire courant avec*  
`dev.print(png, file="graphique.png", width=600)`

*Remarque : Les tableaux produits peuvent être exportés dans le répertoire courant avec*  
`write.csv2(monTableau, file="mon_tableau.csv")`

# Devoir maison du mercredi 8 janvier 2014 : Corrigé

## 1 Analyse de l'influence du niveau d'études sur la variable *inc*

La distribution de la participation aux revenus du ménage selon le niveau d'étude est donnée dans la figure ci-dessous :



Par ailleurs, le tableau ci-dessous résume les principales caractéristiques numériques de la variable « participation aux revenus du ménage » en fonction du niveau d'études :

	Minimum	Q1	Médiane	Moyenne	Q3	Maximum	Écart type
<b>Études universitaires</b>	5,12	16,89	22	25,40	28,95	96	14,0
<b>Pas d'études universitaires</b>	0,20	11,90	16,40	18,11	22	88	9,8

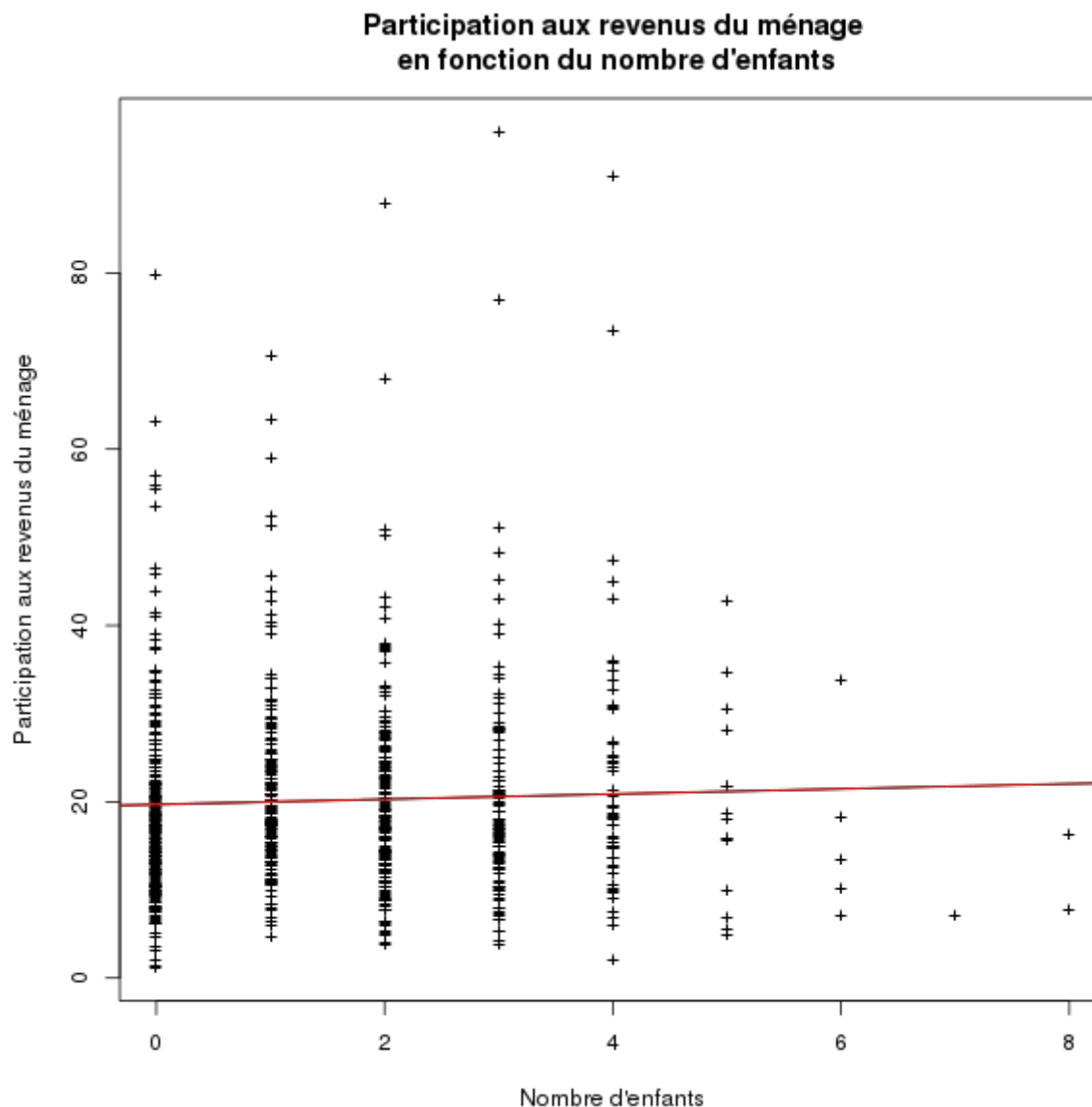
Les femmes ayant suivi des études universitaires ont donc une participation aux revenus du ménage

qui est plus forte que les femmes n'ayant pas suivi d'études universitaires. En moyenne, les premières participent à hauteur de 25,40 % aux revenus du ménage contre 18,11 % pour les secondes. La moitié des femmes ayant suivi des études universitaires participent pour plus de 22 % aux revenus du ménage. Toutefois la distribution de la participation aux revenus du ménage est fortement dissymétrique dans les deux populations, avec un fort étalement vers les grandes valeurs (le maximum est à 96 % pour les femmes ayant suivi des études universitaires et à 88 % pour les autres) : les deux distributions ont une variabilité assez importante qui est plus marquée pour les femmes ayant suivi des études universitaires que pour les autres.

Le rapport de corrélation est égal à 0,282 environ, ce qui est le signe d'une influence plutôt faible du niveau d'études sur la participation aux revenus du ménage.

## 2 Analyse de la corrélation entre participation aux revenus du ménage de la femme et nombre d'enfants

Le nuage de points de la participation aux revenus du ménage des femmes en fonction du nombre de leurs enfants est donné dans la figure ci-dessous avec la droite de régression linéaire :



Aucune tendance nette ne semble se dessiner sur l'influence du nombre d'enfants sur la participation des femmes aux revenus du ménage, si ce n'est que les femmes ayant beaucoup d'enfants (5 ou plus) participent toutes de manière minoritaire aux revenus du ménage. L'équation de la droite de

régression est :

$$inc = 0,293 \times nb. d'enfants + 19,690$$

Le  $R^2$  est égal à 0,037 ce qui confirme une faible corrélation et donc l'inadéquation de la droite de régression pour expliquer les valeur de la participation des femmes aux revenus du ménage.

### 3 Script

Les lignes de code suivantes ont permis d'extraire les résultats commentés au-dessus :

```
# pré-traitements
Mroz$nbenfants <- Mroz$k5+Mroz$k618
Mroz <- subset(Mroz, inc>0)

## niveau d'études et revenu
# graphique
plot(inc~wc, data=Mroz,
     main="Participation aux revenus\n selon le niveau d'études des femmes",
     xlab="Études universitaires", names=c("non","oui"),
     ylab="Participation aux revenus du ménage")
dev.print(png, file="boxplot-wc-inc.png",width=600)
# statistiques conditionnelles
by(Mroz$inc, Mroz$wc, summary)
tapply(Mroz$inc, Mroz$wc, sd)
# rapport de corrélation
res <- anova(lm(inc~wc, data=Mroz))
sqrt(res[1,"Sum Sq"]/sum(res[,"Sum Sq"]))

## nombre d'enfants et revenus
# graphique
plot(Mroz$nbenfants, Mroz$inc,
     main="Participation aux revenus du ménage\n en fonction du nombre
d'enfants",
     xlab="Nombre d'enfants", ylab="Participation aux revenus du ménage",
     pch="+")
resLM <- lm(inc~nbenfants, data=Mroz)
abline(resLM$coefficients, col="red",lwd=1.5)
dev.print(png, file="scatterplot-inc-k.png",width=600)
# droite de régression
resLM$coefficients
# coefficient de corrélation linéaire
cor(Mroz$nbenfants,Mroz$inc)
```

## Devoir maison du mercredi 8 janvier 2014 : Barème

### **1 inc VS wc**

Graphique : \_\_\_\_\_ / 1

Statistiques : \_\_\_\_\_ / 2 (-50 % si il manque l'écart type)

Rapport de corrélation : \_\_\_\_\_ / 1

Commentaires : \_\_\_\_\_ / 3 (central, dispersion, rapport de corrélation)

### **2 inc VS nombre d'enfants**

Graphique : \_\_\_\_\_ / 1

Analyse de la régression : \_\_\_\_\_ / 1

Commentaires : \_\_\_\_\_ / 1

### **3 Script**

Script : \_\_\_\_\_ / 2

**Total : \_\_\_\_\_ / 12**