

# Devoir maison du mercredi 13 novembre 2013 : 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 la variable quantitative discrète « nombre total d'enfant »

Il faudra tout d'abord créer cette variable à partir des variables `k5` et `k618`. Ensuite, l'analyse contiendra des résumés numériques et graphiques de la variable ainsi qu'un court commentaire sur les résultats produits.

## 2. Une analyse de la variable quantitative continue `inc`

Cette analyse contiendra des résumés numériques et graphiques de la variable ainsi qu'un court commentaire. En particulier, on étudiera la concentration des revenus et on repérera la présence d'une variable aberrante et on effectuera l'analyse avec et sans cette valeur aberrante.

Aide : La fonction `subset` permet d'extraire des données correspondant à une condition. Par exemple, `subset(airquality, Temp>80)` extrait du jeu de données `airquality`, les observations pour lesquelles la variable (colonne) `Temp` a une valeur supérieure à 80.

## 3. Les scripts ayant permis ces analyses

*Remarque : Les graphiques produits peuvent être exportés dans le répertoire courant avec*

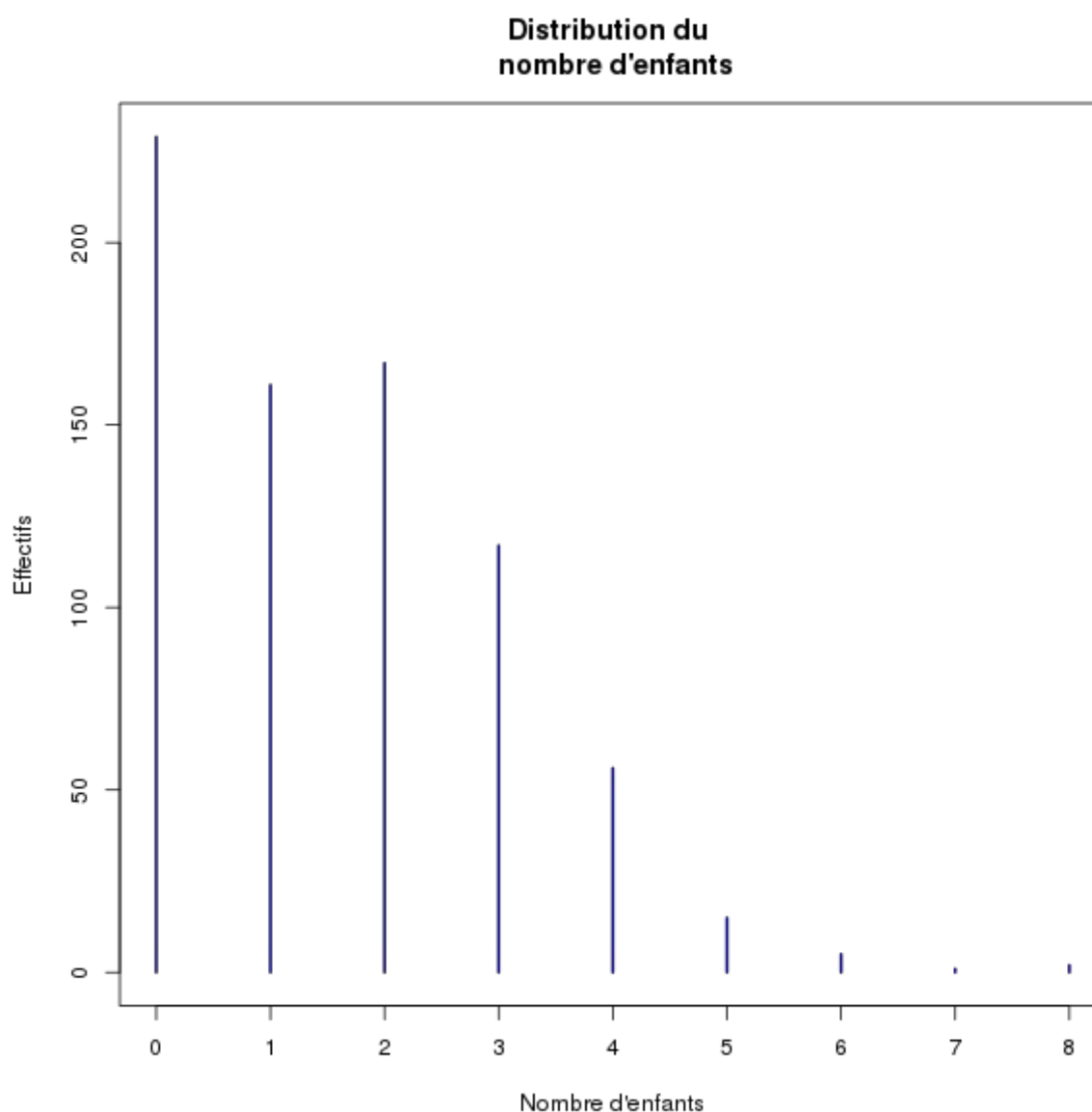
```
dev.print(png, file="graphique.png", width=600)
```

## Devoir maison du mercredi 13 novembre 2013 : Corrigé

### 1 Analyse de la variable « nombre d'enfants »

La distribution du nombre d'enfants par femme est donné dans le tableau et la figure ci-dessous :

Nb d'enfants	0	1	2	3	4	5	6	7	8
Effectifs	229	161	167	117	56	15	5	1	2
Fréquences	30,41 %	21,38 %	22,18 %	15,54 %	7,44 %	1,99 %	0,66 %	0,13 %	0,27 %
Fréquences cumulées	30,41 %	51,79 %	73,97 %	89,51 %	96,95 %	98,94 %	99,60 %	99,73 %	100 %



La majorité des femmes (plus de 51%) ont un enfant ou aucun enfant et près des trois quarts des femmes ont moins de deux enfants. Cependant, un faible nombre de femmes (environ 1 % de la population) ont 5 enfants ou plus.

## 2 Analyse des revenus du ménage provenant de la femme

Les principales caractéristiques numériques des revenus du ménage provenant de la femme sont répertoriées dans le tableau ci-dessous :

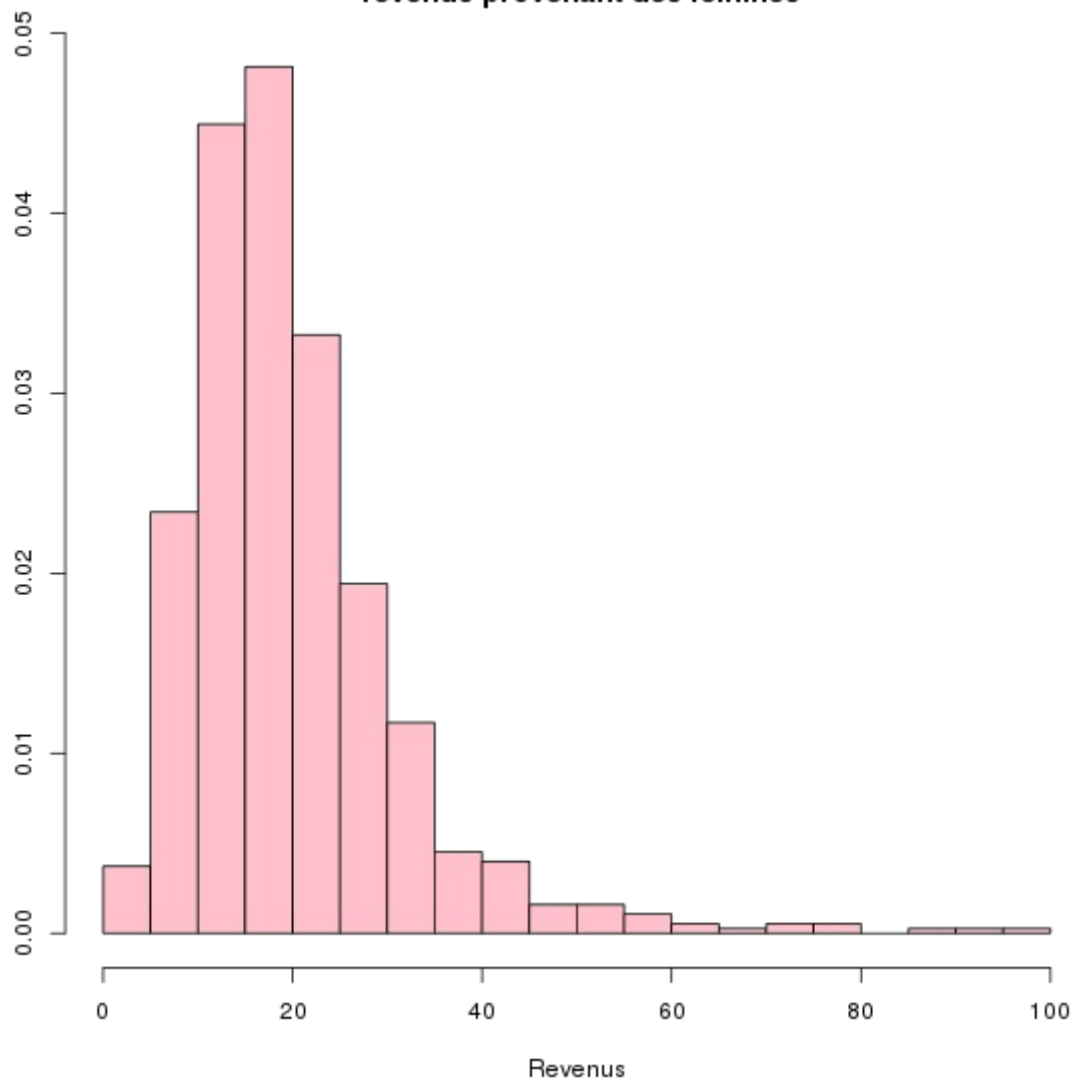
Caractéristique numérique	Valeur
Minimum	-0,029
Maximum	96
Moyenne	20,13
Écart type	11,63
Coefficient de variation	57,8 %
Q1	13,02
Médiane	17,70
Q3	24,47
Coefficient d'asymétrie	2,21
Coefficient d'aplatissement	11,38
Indice de Gini	0,578

Une des valeurs de revenu est négative, ce qui semble étonnant pour cette variable. Par précaution, nous retirons la valeur négative de l'analyse. Les valeurs des principales caractéristiques numériques deviennent alors :

Caractéristique numérique	Valeur
Minimum	1,20
Maximum	96
Moyenne	20,13
Écart type	11,62
Coefficient de variation	57,6 %
Q1	13,06
Médiane	17,74
Q3	24,47
Coefficient d'asymétrie	2,22
Coefficient d'aplatissement	11,43
Indice de Gini	0,286

Comme le confirme l'histogramme ci-dessous

### Distribution des revenus provenant des femmes

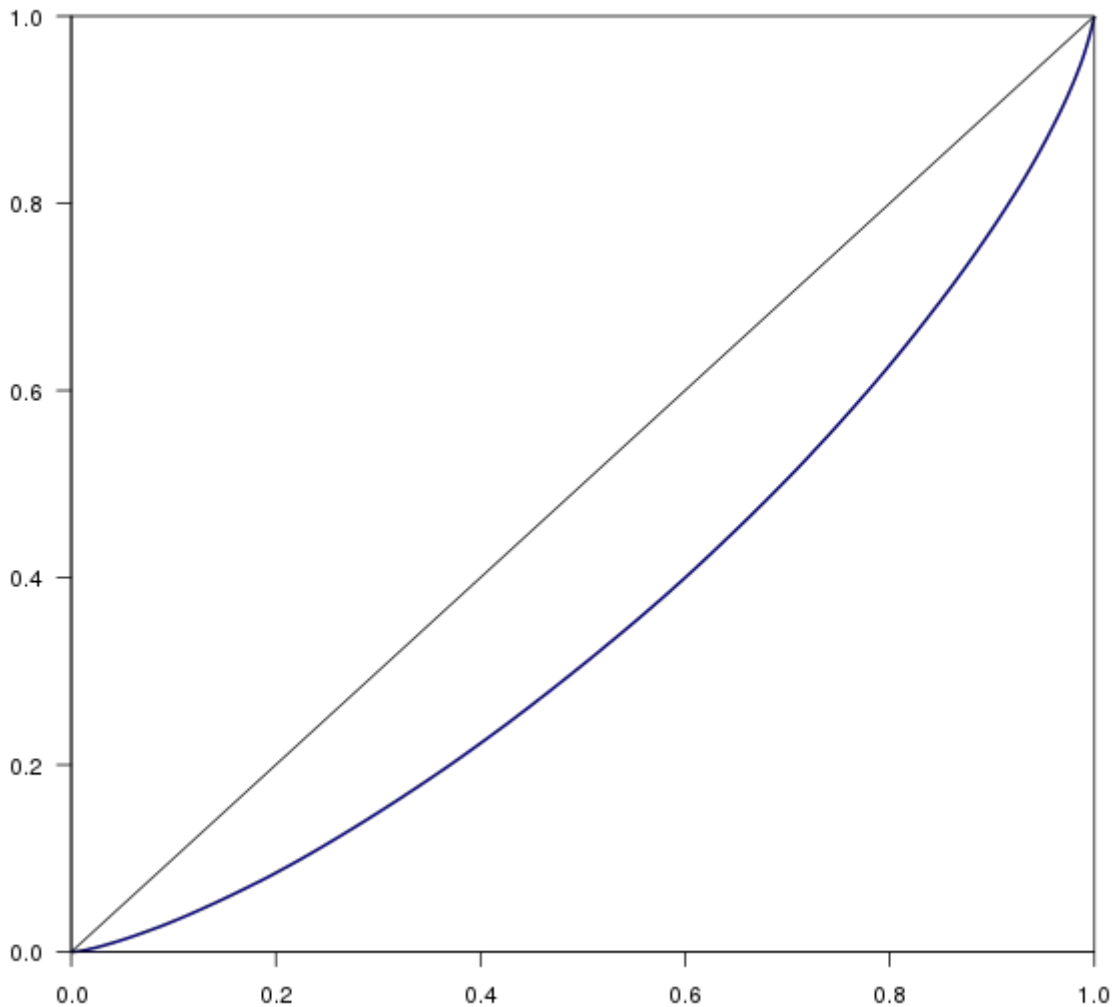


les revenus des femmes sont centrés autour de la valeur 20 avec un étalement vers les fortes valeurs confirmés par le coefficient d'asymétrie élevé et le fait que la médiane est inférieure à la moyenne. La distribution est assez concentrée autour des valeurs 10 à 30, ce qui est confirmé par un coefficient d'aplatissement élevé.

La concentration des revenus est néanmoins modérée avec un coefficient de Gini égal à 0,286 (l'indice de Gini des revenus en France est 0,289 à comparer à celui des USA, par exemple, supérieur à 0,45 en 2009<sup>1</sup>). Ce fait est confirmé par la courbe de Lorenz ci-dessous :

1 Wikipedia France [http://fr.wikipedia.org/wiki/Coefficient\\_de\\_Gini](http://fr.wikipedia.org/wiki/Coefficient_de_Gini)

**Courbe de Lorenz**



### 3 Script

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

```
# nombre d'enfants
## Création de la variable
Mroz$nbchildren <- Mroz$k5 + Mroz$k618
## Effectifs
table.nbchildren <- table(Mroz$nbchildren)
table.nbchildren
## Fréquences
freq.nbchildren <- round(table.nbchildren/sum(table.nbchildren)*100,2)
freq.nbchildren
## Fréquences cumulées
cumsum(freq.nbchildren)
## Diagramme en bâtons
plot(table.nbchildren, main="Distribution du \n nombre d'enfants" , xlab=
      "Nombre d'enfants", ylab="Effectifs" , col= "darkblue")
dev.print(png, file="nbchildren-batons.png", width=600)

# inc
```

```
## Caractéristiques numériques
summary(Mroz$inc)
sd(Mroz$inc)
sd(Mroz$inc)/mean(Mroz$inc)
library(e1071)
skewness(Mroz$inc, type=1)
kurtosis(Mroz$inc, type=1)+3
library(ineq)
ineq(Mroz$inc, type="Gini")
## Extraction d'un sous-ensemble
Mroz=subset(Mroz , inc>0)
summary(Mroz$inc)
sd(Mroz$inc)
sd(Mroz$inc)/mean(Mroz$inc)
skewness(Mroz$inc, type=1)
kurtosis(Mroz$inc, type=1)+3
ineq(Mroz$inc, type="Gini")
## Histogramme
hist(Mroz$inc,breaks=20,col="pink",xlab="Revenus",ylab="",freq=FALSE,main="Dist
ribution des \n revenus provenant des femmes")
dev.print(png,file="revenus-hist.png",width=600)
## Courbe de Lorenz
plot(Lc(Mroz$inc),main="Courbe de Lorenz",xlab="",ylab="",col="darkblue")
dev.print(png,file="revenus-lorenz.png",width=600)
```

# Devoir maison du mercredi 13 novembre 2013 : Barème

## 1 Variable « nombre d'enfants »

Effectifs et/ou fréquences : \_\_\_\_\_/1

Effectifs et/ou fréquences cumulées : \_\_\_\_\_ / 1

Diagramme en bâtons : \_\_\_\_\_ /1

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

## 2 Variable inc

Toutes les caractéristiques numériques de base : \_\_\_\_\_ / 2

Aplatissement, asymétrie, coefficient de variation... : \_\_\_\_\_ / 1 (si au moins 2)

Traitement de la valeur aberrante : \_\_\_\_\_ / 1

Histogramme ou boîte à moustaches : \_\_\_\_\_ /1

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

Indice de Gini, courbe de Lorenz : \_\_\_\_\_ / 1

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

## 3 Script

Script : \_\_\_\_\_ / 3 (nombre d'enfants, basique pour inc, avancé pour inc)

**Total : \_\_\_\_\_ / 15**