

## Devoir maison du mercredi 2 octobre 2012 : 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. Description des données

Un bref descriptif des données, rédigé en français, et contenant (a minima) la population, sa taille, les variables, leurs types.

### 2. Script de préparation des données

Un script de préparation des données contenant :

- les lignes de code permettant de recoder les variables `lfp`, `wc` et `hc` en français ;
- les lignes de code permettant de créer une nouvelle variable contenant le nombre d'enfants.

Enfin, sauver les données :

```
save(Mroz, file="Mroz.rda")
```

et tenir à ma disposition ce fichier également.

# Devoir maison du mercredi 2 octobre 2012 : Corrigé

## 1 Description des données

Les données Mroz contiennent des informations collectées sur une population de 753 femmes mariées et publiées dans l'article (Mroz, 1987). Les mesures de 8 variables sont disponibles :

- `lfp` est une variable qualitative nominale à deux niveaux (« oui » et « non ») qui indique si la femme est active (ie : travaille) ;
- `k5` est une variable quantitative discrète qui compte le nombre d'enfants de moins de 5 ans de la femme ;
- `k618` est une variable quantitative discrète qui compte le nombre d'enfants âgés de 6 à 18 ans de la femme ;
- `age` est une variable quantitative discrète qui indique l'âge de la femme ;
- `wc` est une variable qualitative nominale à deux niveaux (« oui » et « non ») qui indique si la femme a suivi une formation universitaire ;
- `hc` est une variable qualitative nominale à deux niveaux (« oui » et « non ») qui indique si son mari a suivi une formation universitaire ;
- `lwg` est une variable quantitative continue qui contient le logarithme du taux de rémunération de la femme. Si la femme ne travaille pas, cette variable est estimée à partir par une régression à partir des autres variables ;
- `inc` est une variable quantitative continue qui contient le revenu familial à l'exception des revenus provenant de la femme.

## 2 Nettoyage des données

Les lignes de code suivantes permettent de recoder les variables `lfp`, `wc` et `hc` en français : de manière plus précise, « yes » est recodé en « oui » et « no » en « non » :

```
Mroz$lfp <- factor(Mroz$lfp, labels=c("non","oui"))
Mroz$wc <- factor(Mroz$wc, labels=c("non","oui"))
Mroz$hc <- factor(Mroz$hc, labels=c("non","oui"))
```

La ligne de code suivante permet de créer une nouvelle variable, « `nbchildren` », contenant le nombre total d'enfants de chaque femme (ie : la somme du nombre d'enfants de moins de 5 ans et du nombre d'enfants entre 6 et 18 ans) :

```
Mroz$nbchildren <- Mroz$k5+Mroz$k618
```

## Références

Mroz, T. A. (1987) The sensitivity of an empirical model of married women's hours of work to economic and statistical assumptions. *Econometrica*, 55, 765-799.

## Devoir maison du mercredi 2 octobre 2012 : Barème

### 1 *Description des données*

Phrase rédigée décrivant la population et sa taille : \_\_\_\_\_/1

Phrases rédigées décrivant les variables et leurs types : \_\_\_\_\_ / 3 (-1 par erreur)

### 2 *Nettoyage des données*

Recodage des données : \_\_\_\_\_ / 2 (0,5 seulement si le recodage est fait dans le mauvais sens ; -1 par ligne de code manquante)

Nouvelle variable : \_\_\_\_\_ / 2

Fichier rda correct : \_\_\_\_\_ / 1

**Total : \_\_\_\_\_ / 9**