

# CPU. Corrigé

1. Ecrire le modèle correspondant à l’instruction R précédente (on donnera la taille des objets).

$\mathbf{Y} \in \mathbb{R}^n$ ,  $n = 206$ ,  $\mathbf{X} \in \mathcal{M}_{206,7}(\mathbb{R})$  dont la première colonne ne vaut que des 1. Les autres colonnes correspondent aux covariables.  $\beta \in \mathbb{R}^7$ .  $\varepsilon$  gaussien de taille  $n = 206$ .

$$\mathbf{Y} = \mathbf{X}\beta + \varepsilon$$

2. Pourquoi pensez-vous que la variable `estperf` n’a pas été incluse dans le modèle. On pourra utiliser les corrélations entre les variables représentées dans la figure des corrélations.

Elle est parfaitement corrélée à  $\mathbf{Y}$ , non intéressant

3. Les hypothèses sur les résidus sont-elles respectées ? Indiquer le graphique utilisé pour chaque hypothèse et expliquer pour quelles raisons chaque hypothèse est ou n’est pas respectée.

4 postulats : [P1] résidus centrés / [P2] de variance constante / [P3] indépendants / [P4] gaussien.  
 Pour [P1] : points autour d’une droite horizontale graphe en haut à gauche  
 Pour [P2] : pas de tendance dans le graphe en bas à gauche  
 Pour [P3] : non-vérifiable. Seulement sur l’échantillonnage.  
 Pour [P4] : QQ plot.

4. Donner les hypothèses du test du modèle global de `mod1`. Rappeler l’expression de la statistique de test, sa loi sous  $H_0$ . Donner la valeur de sa réalisation ici. Conclure

- $\mathcal{H}_0 : \beta_1 = \beta_2 = \dots = \beta_7 = 0$  versus  $\mathcal{H}_1 = \exists k = 1 \dots 7$  such that  $\beta_k \neq 0$ .
- $F = \frac{(SCR_0 - SCR)/(p+1-1)}{SCR/(n-(p+1))} \sim \mathcal{F}_{p,n-(p+1)} = \mathcal{F}_{6,199}$
- $F = 235.1$ . p-value  $< 2.10^{-16}$ . On rejette  $\mathcal{H}_0$ .

5. Quel est le test fait sur la ligne `mmin` du `summary`

- $\mathcal{H}_0 : \beta_{mmin} = 0$  versus  $\mathcal{H}_1 : \beta_{mmin} \neq 0$

6. Quel modèle correspond à l’instruction suivante?

```
res_lm_0 = lm(perf ~ 1, data=data.cpu)
```

$$Y_i = \beta_0 + \varepsilon_i$$

7. On cherche à faire de la sélection de variables. Commentez les sorties R correspondantes. Quel critère a été utilisé?

Quel algorithme a été utilisé pour parcourir les modèles (rappeler brièvement son principe). . On ajoute et on enlève une covariable à chaque étape: Quelles sont les variables sélectionnées? `mmin`, `mmax`, `cach`, `chmin` L’intercepte n’est pas testé. On ne le sélectionne pas.