

TD 2

Etude de simulation sur les intervalles de confiance

On travaille toujours sur l'exemple de **Tension Age**. On va simuler 100 jeux de données selon le modèle estimé. Les données simulées sont considérées comme des observations des différentes expériences, c'est à dire des nouvelles observations $(X_1, Y_1^0), \dots, (X_5, Y_5^0)$. La méthode des moindres carrés est appliquée à chaque jeu pour obtenir les paramètres $\hat{\mu}^0$ et $\hat{\beta}^0$. On obtiendra ainsi 100 lignes droites de régression. On compare les intervalles de confiance obtenues à partir du premier jeu de données `age=c(35,45,55,65,75)`, `tension=c(114,124,143,158,166)` et les résultats de simulation.

```
age=c(35,45,55,65,75)
tension=c(114,124,143,158,166)
reg=lm(tension~age)
# Extraction des paramètres
mu = 65.1; beta = 1.38; sig = 3.246
```

Taper les codes suivants, qu'obtenez vous ?

```
(mu = reg$coef[1]); (beta = reg$coef[2]); (sig = summary(reg)[[6]])
```

Exercice 1

Simuler d'abord un nouveau jeu de données `y` de taille 5 pour les mêmes ages en utilisant `mu`, `beta` et `sig`. Tracer le nuage de points (`tension`, `age`). Ajouter ensuite le nuage de points des données simulées (`y`, `age`) en couleur verte. Faire une régression linéaire sur (`y`, `age`) et sauvegarder les résultats sous le nom `reg2`. Ajouter dans la figure précédente la droite de régression de `reg` en rouge et la droite de régression de `reg2` en bleu. Utiliser l'option `lwd` dans les commandes de graphique pour augmenter l'épaisseur de ligne.

Exercice 2

Les codes ci dessous sont pour simuler 100 jeux de données de même façon de l'exercice 1 et ajouter chaque fois la nouvelle ligne de régression dans la figure précédente. Calculer les intervalles d'estimation et prédiction à partir du premier jeu de données (`tension`, `age`) en utilisant la fonction `predict`. Tracer ensuite dans la même figure ces intervalles.

```
for (i in 1:100){
  y=mu+beta*age+rnorm(5,0,sig)
  reg2=lm(y~age)
  points(age,y,col="green")
  abline(reg2,col="cyan")
}
```

Exercice 3

Que fait le programme suivant ?

```

n = 10000; m = rep(0, n); b = rep(0, n); s = rep(0, n)

for (i in 1:n){
y = mu+beta*age+rnorm(5,0,sig)
reg2 = lm(y~age)
m[i] = reg2$coef[1]; b[i] = reg2$coef[2]; s[i] = summary(reg2)[[6]]
}

mus = summary(reg)[[4]][3]; betas = summary(reg)[[4]][4];
hist((m-mu)/mus, freq = F, breaks = 100)
y = seq(-4, 4, length.out = 100)
lines(y, dnorm(y), lwd=3)
hist((b-beta)/betas, freq = F, breaks = 100)
lines(y, dnorm(y), lwd=3)

```