

## TP 4 : Modèles linéaires

Pour ce TP nous allons reprendre les données sur les crabes dormeurs qui sont fournies dans le fichier `crabs.data`. Représentez la taille avant mue en fonction de la taille après mue. Que remarquez-vous ?

Nous remarquons qu'il y a une présence de crabes juvéniles. Ceci est embêtant puisqu'ils muent tous les deux mois contrairement aux crabes adultes qui muent une fois par an. Il convient donc de les écarter. Un crabe est dit « juvénile » lorsque sa taille après mue est inférieure à 100mm. Gardez que les crabes adultes.

Une fois cette sélection faite, « attachez » à votre session R les données. A quoi sert le graphique suivant ?

```
plot(postsz, presz, col = 1 + (lf == 1))
```

## 1 Premières analyses

Selon vous l'utilisation d'un modèle linéaire est-elle justifiée ? Si oui dites pourquoi.

Le coefficient de corrélation linéaire peut-être calculé à l'aide de la fonction `cor`. Calculez ce coefficient. Puis à l'aide de la fonction `sd` calculer l'estimateur des moindres carrés de la droite prédisant la taille avant mue en fonction de celle après mue. Comparez vos résultats à la sortie suivante

```
fit <- lm(presz ~ postsz)
fit
```

Vous venez d'ajuster votre premier modèle linéaire avec R. La formule `presz ~ postsz` indique que l'on veut prédire la taille avant mue en fonction de la taille après mue.

Commentez les sorties suivantes

```
summary(fit)
anova(fit)
```

Qu'en concluez-vous ?

Afin de savoir si un modèle (linéaire ou non) est correct, l'analyse des résidus est primordiale. Rappelez les hypothèses sous-jacentes au modèle linéaire et essayez de comprendre la sortie graphique suivante. *Remarque : la notion derrière le quatrième graphique est un peu complexe mais votre enseignant se fera un plaisir de vous commenter l'intérêt de tel graphique.*

```
par(mfrow = c(2, 2))
plot(fit)
```

La fonction `predict` permet de prédire la taille avant mue selon notre modèle. Interprétez la sortie graphique suivante

```
plot(predict(fit), presz, xlab = "Taille avant mue predites", ylab =
      "Taille avant mue observees")
abline(0, 1)
```

Avec cette même fonction, il est également possible de prédire pour des tailles autres que celle dont nous disposons. Par exemple

```
predict(fit, data.frame(postsz = c(125, 132, 144)), interval = "prediction")
```

nous donne les prédictions selon notre modèle pour des tailles après mues de 125, 132 et 144mm ainsi que leur intervalles de confiance à 95%.

Remarquez que nous avons dû écrire `postsz = c(125,132,144)` pour que le logiciel sache que ces données correspondent à des tailles après mue.

## 2 Analyse plus poussée

On se pose la question de savoir si la taille avant mue est fonction de la taille après mue mais également du fait d'être ou non un crabe de laboratoire. Écrivez le modèle linéaire que vous considèreriez pour répondre à cette question.

A quoi correspond la sortie suivante ?

```
fit2 <- lm(presz ~ postsz + factor(lf))
fit2
```

Selon vous le facteur « crabe de laboratoire » vous semble-t-il pertinent ? Comment feriez vous pour tester si tel est vraiment le cas ? Et avec l'aide de R ?

A partir de la sortie suivante, essayez de retrouver comment à été calculé le nombre 0.392.

```
summary(fit2)
```

## 3 Pour les plus motivés

S'il vous reste encore du temps vous pouvez recommencer la feuille de TP en essayant de prédire l'accroissement de la taille en fonction de la taille avant mue.