

TP III: ANOVA

Nous utiliserons le langage R pour ce TP. Le corrigé sera à faire sous forme de `.Rmd` ou `.rnw`.

Nous allons travailler sur des données de qualité de l'air mesurées en Bretagne. Elles sont contenues dans le fichier `ozone.txt` à télécharger sur l'espace pédagogique.

Exercice 1. Analyse de covariance

1. Lire le fichier contenant les données:

```
ozone <- read.table("ozone.txt", header=T)
```

Quelles sont les deux variables qualitatives? Combien de modalités comportent-elles?

2. On va chercher à expliquer la concentration maximale en ozone (`maxO3`) en fonction de la température à midi (`T12`) et de la direction du vent.

- (a) Créer une variable `ozEst` contenant les données observées lorsque le vent vient de l'est.

Utiliser cela pour estimer un modèle de régression simple de `maxO3` en fonction de `T12` pour chaque modalité de la variable `vent` (modèles stratifiés).

- (b) Tracer sur un même graphe les points d'observation en changeant la marque en fonction de la modalité de la variable `vent` (e.g. adapter la valeur de `pch` et `col`). Enfin, rajouter les 4 droites de régression estimées (e.g. avec la fonction `abline`). Commenter.

- (c) On veut maintenant écrire le modèle complet :

```
mod1 <- lm(maxO3~1+vent+T12:vent, ozone)
```

Le terme `T12:vent` correspond à l'interaction entre `T12` et le vent (lire la sortie de `?formula`). Regarder les estimateurs des coefficients. A quoi correspondent-ils?

On aurait également pu estimer le modèle avec la constante:

```
mod1b <- lm(maxO3~vent+T12:vent, ozone)
```

Qu'est-ce qui est modifié ?

- (d) On peut ensuite tester l'égalité des pentes et des ordonnées à l'origine suivant la direction du vent. Le modèle avec une seule pente s'écrit:

```
mod2 <- lm(maxO3~1+vent+T12, ozone)
```

Pour comparer les deux modèles, on tape :

```
anova(mod1, mod2)
```

Quelle est la conclusion ?

Faire de même pour tester l'égalité des ordonnées à l'origine. Peut-on conclure que la direction du vent n'est pas utile pour estimer `maxO3` en fonction de `T12`?

Exercice 2. Analyse de variance à deux facteurs On va maintenant chercher à expliquer la concentration en ozone uniquement par les deux variables qualitatives du modèle: `vent` et `pluie`.

1. Calculer la moyenne de `maxO3` lorsque `vent=="Est"` et `pluie=="Sec"`. Faire de même pour les 7 autres couples de modalités. Représenter graphiquement ces moyennes.
2. Estimer le modèle complet, avec interaction. Quelles sont les contraintes sur les paramètres?
3. Comparer avec le modèle sans interaction. L'interaction entre vent et pluie est-elle significative?

4. Si elle n'est pas significative, tester la significativité du facteur pluie, puis du facteur vent.
5. Quel modèle retenir finalement? Cela confirme-t-il les résultats de l'analyse de covariance?