

Acquisition et analyse de données (BING-F4002): travaux pratiques

Marius Gilbert & Marc Dufrêne

Année académique 2015-2016

Séance VI. ANOVA à 1 et à 2 facteurs

- Chargez le jeu de données PlantGrowth
- Affichez les boîtes de dispersion de la variable weight en fonction de la variable de groupement.
- La fonction ci-dessous réalise une analyse de variance et stocke les résultats de celle-ci dans un objet appelé myAOV :

```
myAOV = aov(weight ~ group, data = PlantGrowth)
```

- La fonction summary(myAOV) vous permet d'afficher les résultats de l'analyse variance. Qu'est-ce qui a été testé ici ? Précisez les hypothèses H0 et H1. Quelles sont les conclusions que vous pouvez tirer de ce test ?
- La fonction residuals(myAOV) vous renvoie les résidus de cette analyse. A l'aide des fonctions hist(), qqnorm(), et shapiro.test(), vérifiez la normalité des résidus de l'ANOVA.
- La fonction bartlett.test(weight ~ group, data = PlantGrowth), va nous permettre de vérifier l'homogénéité des variances. Quelles sont les hypothèses H0 et H1 de ce test ? Dans ce cas ci, les variances peuvent-elles être considérées comme égales ?
- Chargez le jeu de données InsectSprays et en utilisant les mêmes codes que dans l'exemple précédent, testez l'effet des insecticides sur les nombres d'insectes observés. Vérifiez la normalité des résidus et l'homoscédasticité. Avant même de procéder au test de Bartlett, les variances semblent-elles égales ?
- Essayez de tester l'effet d'une transformation des variables (transformation racine, et log), ceci améliore-t-il la normalité des résidus ? Et l'homoscédasticité ?
- Chargez le package MASS en tapant :

```
library(MASS)
```

- Chargez à présent le jeu de données « cabbages ». Quelle est l'expérience réalisée ici ? Décrivez ce jeu de données (statistiques et/ou graphes descriptifs). Sur cette base, pensez-vous qu'il y ait un effet du cultivar et de la date sur le poids des choux et sur leur contenance en vitamines C ? Avant de procéder aux analyses, pensez-vous qu'il y ait une interaction entre la date et le cultivar dans leur effet sur les poids de choux, et sur le taux de vitamines C ? Pourquoi ?
- Faites une ANOVA à deux facteurs (modèle complet avec terme d'interaction) pour comparer les moyennes de poids (variable HeadWt) en fonction du cultivar et de la date. Que pouvez-vous constater sur base des résultats ? Pour rappel, un modèle complet à deux facteurs s'écrit :

```
myAOV = aov(VarQuantitative ~ Facteur1 + Facteur2 + Facteur1:Facteur2, data = JeuxdeDonnées)
```

- Vérifiez les conditions d'application de l'ANOVA.
- A l'aide de la fonction TukeyHSD, réalisez les tests de comparaison multiple sur cette analyse de variance. Comment interprétez-vous ces résultats ?
- Répétez l'analyse (points 3 à 5), mais en étudiant cette fois la contenance en vitamines C (VitC). Les résultats sont-ils comparables ?
- Quel cultivar et quelle date de plantation recommanderiez-vous en fonction de ces analyses ?