

Tableaux de contingence et ANOVA

29 septembre 2021

1. Sélection florale d'une espèce de bourdon

En suivant l'activité de butinage de bourdons fébriles (*Bombus impatiens*) sur un site, vous notez le nombre de visites des bourdons sur quatre genres de plantes, ainsi que la proportion des fleurs de chaque genre sur le site.

Genre	Nombre de visites	Proportions des fleurs du site
<i>Rubus</i>	8	0.12
<i>Solidago</i>	8	0.24
<i>Trifolium</i>	18	0.33
<i>Vaccinium</i>	11	0.31

L'hypothèse nulle pour cette étude est que *B. impatiens* visite chaque genre proportionnellement à sa prévalence sur le site.

- Selon l'hypothèse nulle, quelles sont les fréquences attendues pour les visites à chaque genre de plante?
- Testez l'hypothèse nulle avec la fonction `chisq.test` dans R, avec un seuil de signification de 5%. Si l'hypothèse nulle est rejetée, quel(s) genre(s) sont plus ou moins visités que prévus?

2. Butinage de trois espèces de bourdon

Sur le même site que l'exercice précédent, vous observez l'activité de butinage de deux autres espèces de bourdons (*B. affinis* et *B. ternarius*). Voici le tableau de contingence montrant le nombre de visites par espèce de bourdon et par genre de plante.

	<i>Rubus</i>	<i>Solidago</i>	<i>Trifolium</i>	<i>Vaccinium</i>
<i>B. affinis</i>	10	9	15	8
<i>B. impatiens</i>	8	8	18	11
<i>B. ternarius</i>	20	4	6	5

- Quelle hypothèse nulle pouvez-vous tester à partir de ce tableau? Quelle est l'hypothèse alternative?
- Créez une matrice représentant ce tableau dans R, puis testez l'hypothèse nulle mentionnée en (a) avec la fonction `chisq.test`, avec un seuil de signification de 5%.
- D'après les résultats du test en (b), quel est le nombre de degrés de liberté du χ^2 ? Comment cette valeur est-elle calculée?
- Comment pouvez-vous consulter les fréquences attendues selon l'hypothèse nulle, ainsi que les résidus?
- Si l'hypothèse nulle est rejetée, quelle paire bourdon-plante a le résidu le plus positif, et laquelle a le résidu le plus négatif? Comment interprétez-vous ces résidus?

3. Caractéristiques de choux plantés à différentes dates

Le jeu de données `cabbages` inclus dans le package `MASS` présente le poids en kg (*HeadWt*) et le nombre d'unités de vitamine C (*VitC*) de choux selon la variété (cultivar *Cult*) et la date de plantation. Il y a 10 réplicats pour chacune des six combinaisons de variété et de date.

```
library(MASS)
str(cabbages)
```

```
## 'data.frame': 60 obs. of 4 variables:
## $ Cult : Factor w/ 2 levels "c39","c52": 1 1 1 1 1 1 1 1 1 1 ...
## $ Date : Factor w/ 3 levels "d16","d20","d21": 1 1 1 1 1 1 1 1 1 1 ...
## $ HeadWt: num 2.5 2.2 3.1 4.3 2.5 4.3 3.8 4.3 1.7 3.1 ...
## $ VitC : int 51 55 45 42 53 50 50 52 56 49 ...
```

- Choisissez le sous-ensemble des données correspondant à la variété `c52`. Pour chacune des deux variables numériques (*HeadWt* et *VitC*), produisez un graphique de boîtes à moustaches montrant la distribution de cette variable selon la date de plantation. Avant même de réaliser l'ANOVA, croyez-vous que les suppositions du modèle (en particulier l'égalité des variances) seront respectées dans chaque cas?
- Choisissez l'une des deux variables (*HeadWt* ou *VitC*) qui correspond le mieux au modèle d'ANOVA d'après votre résultat en (a). Réalisez une ANOVA à un facteur et déterminez si la moyenne de cette variable varie significativement ($\alpha = 0.05$) selon la date de plantation.
- Selon un test des étendues de Tukey, entre quelles dates retrouve-t-on une différence significative ($\alpha = 0.05$)? Quel est l'estimé de chacune des différences significatives?