

Contingency tables

September 26, 2018

Answers for this lab must be submitted before **October 3rd at 5pm on Moodle**. In your answer for each question, please include a copy of the R code used (if applicable) and the results obtained.

1. Floral selection of a bumblebee species

By following the foraging activity of bumblebees *Bombus impatiens* on a site, you note the number of bumblebee visits on four genera of plants, as well as the proportion of flowers of each genus on the site.

Genus	Number of visits	Proportions of the flowers of the site
<i>Rubus</i>	6	0.12
<i>Solidago</i>	8	0.24
<i>Trifolium</i>	20	0.33
<i>Vaccinium</i>	11	0.31

The null hypothesis for this study is that *B. impatiens* visits each genus in proportion to its prevalence on the site.

- According to the null hypothesis, what are the expected frequencies for visits to each genus of plant?
- Test the null hypothesis with the `chisq.test` function in R, with a significance level of 5%. If the null hypothesis is rejected, which genus or genera are more or less visited than expected?

2. Foraging activity of three bumblebee species

On the same site as the previous exercise, you observe the foraging activity of two other bumble bee species (*B. affinis* and *B. ternarius*). Here is the contingency table showing the number of visits by bumblebee species and plant genus.

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

- What null hypothesis can you test from this table? What is the alternative hypothesis?
- Create a matrix representing this table in R, then test the null hypothesis mentioned in (a) with the `chisq.test` function, with a significance level of 5%.
- Based on the test results in (b), what is the number of degrees of freedom of the χ^2 distribution? How is this value calculated?
- How can you check the expected frequencies according to the null hypothesis, as well as the residuals?
- If the null hypothesis is rejected, which bumblebee-plant pair has the most positive residual, and which has the most negative residual? How do you interpret these residuals?

3. Applications of the chi-squared test

Here are three questions related to urban forestry. For which of these questions is the chi-squared test most appropriate? Justify your answer.

- a) Does the mean age of trees in different neighborhoods of a city increase with the mean age of houses in the neighborhood?
- b) Does the presence of a large tree affect the energy cost of a home?
- c) Does the distribution of planted tree species differ between neighborhoods?