# Home assignment 4

## Exercise 4.1

Consider the macropod study. A test was conducted to determine if kangaroos and wallabies differ with respect to how many animals reveal M. paratuberculosis infection.

> "*Mycobacterium avium subspecies paratuberculosis (M. paratuberculosis) is the causal agent of Johnes Disease (JD) or paratuberculosis, which is a serious wasting disease predominantly of cattle, sheep and goats…. A study was performed to determine whether M. paratuberculosis infection was present in macropods[1] grazing with infected sheep on Kangaroo Island, South Australia in 2001–2002*
> *No significant difference was found between the proportions of infected animals of the two species; 2/55 kangaroos (3.6%) and 6/454 wallabies (1.3%) had positive tissue cultures (P = 0.21).*"

[1]The marsupial family Macropodidae, includes kangaroos, wallabies, tree-kangaroos, pademelons, and several others

a) Determine which test was used in the study above and verify the results by running the test on the data.

## Exercise 4.2

From an article[1][2] on hens and caging systems the following data and results are extracted: The results showed that the proportion of ascarid eggs in free-range laying hens were high, both for hens that live indoors and outdoors. For hens living in cages the occurrence of ascarid eggs was significantly lower. A significant increase of ascarid eggs was seen between 2004 and 2008 (P=0,001). Geographically the infected flocks were spread over the country.

Data table:

**Table 3.** *Occurrence of ascarid eggs in faecal samples from Swedish laying hen flocks in the autumn of 2004 and 2008.*

| | 2004 | | | 2008[a] | | |
|---|---|---|---|---|---|---|
| Housing system | Flocks (*n*) | Infected flocks (*n*) | Prevalence (%) | Flocks (*n*) | Infected flocks (*n*) | Prevalence (%) |
| Cages[b] | 42 | 1 | 2.4 | 46 | 2 | 4.3 |
| Single-tiered floor system | 72 | 20 | 27.8 | 42 | 12 | 28.6 |
| Multi-tiered aviary system | 36 | 6 | 16.7 | 46 | 24 | 52.2 |
| Free-range, including organic hens[c] | 35 | 17 | 48.6 | 35 | 26 | 77.1 |
| Unknown | 1 | 0 | 0 | 0 | 0 | 0 |
| Total | 186 | 44 | 23.5 | 169 | 64 | 38.3 |

From: Ascarid infections in laying hens kept in different housing systems, D Jansson, A Nyman, I Vågsholm, D Christensson, M Göransson, O Fossum & J Höglund, Avian Pathology, Volume 39, Issue 6, 2010.

Do the following computations even if the assumption for one or several are not fulfilled. Discuss the assumption in exercise d).

a) Calculate an appropriate test **by hand** to verify that there is a significant increase in the occurrence of ascarid eggs between 2004 and 2008. Use the total number in the last line, where you can find the estimated proportion as prevalence and the number of observations as number of flocks(n). Conduct the test and draw a conclusion.
b) Redo this test in **R**[3] and present code and results as well as comments. Especially present the p-value gained from this test and interpret it.
c) **By hand** compute a confidence interval for: year 2008, free-range, including organic hens. Draw conclusions.
d) Which assumptions need to be fulfilled for the analysis in a) and c)? Are they fullfilled here?
e) In this study the number of infected flocks are presented in tables and analysis. From the article you can learn that 4 samples were used from each of the flocks. Why do the authors not use all the data available, i.e. all 4 observations from each flock?

[1] "Spolmasken Ascaridia galli ökar hos svenska värphöns" (Svensk veterinärtidning, Nummer 8-9, 2011, in Swedish)

[2] Ascarid infections in laying hens kept in different housing systems, D Jansson, A Nyman, I Vågsholm, D Christensson, M Göransson, O Fossum & J Höglund, Avian Pathology, Volume 39, Issue 6, 2010.

[3] In R you will note that the computed value for the test statistics is called X-squared. This is actually a chi-square statistics and should get this value by computing $z^2$, where z is the value for the test statistics in a).

## Exercise 4.3

Read the methane.txt data. The dataset contains:

lactation_week…..the lactation week

DM… Dry matter intake of the cow

milk… milk production of the cow

breed… two different breeds Hol (Holstein) and SRB

methane…methane production of the cow

methaneclass… methane production as categorical variable (low, medium, high)

Do the following:

a) Determine if the methane production is significantly different for Holstein and SRB breeds. Assume equal variances and normal distribution. Note the p-value and the degrees of freedom.
b) Compute the variances for these two groups of methane productions. They should be quite different. Redo (or look at) the test with assuming unequal variances. Note again the p-value and the degrees of freedom and compare to a).
c) Make a categorical variable for lactation week with classes about 6 weeks long.
d) Use an appropriate test to test if the categorical variable lactation week (created in c) and methaneclass are independent or not. Interpret.
e) In the test above you have probably gotten a warning message together with the result. Why?
f) Compute the expected values in the test in d) **by hand** and what do they tell you?
g) Redo c) and d), but this time with classes about 9 weeks long (this will give you only two classes). Do you still get a warning message? Why or why not?
h) Create a subset of the data only containing the SRB breed. Test if the milk production is significantly higher than 30.