Home assignment 2

Exercise 2.1

In the fact sheet "Mycobacterium paratuberculosis and kangaroos" it is stated that:

> "*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) What is meant by saying that no **significant** difference was found? How does this differ from only saying that no difference was found?
b) For this statistical test conducted: what is/are the population(s), how large is/are the sample(s).

Exercise 2.2

Hand calculations: The prevalence of *M. paratuberculosis* infection, i.e. the risk that a randomly selected macropod is infected, is believed to be around 2%. Assume that a random sample of 30 macropods is examined.

a) What is the probability that exactly one macropod will be infected?
b) What is the probability that at least one macropod will be infected? (Hint: Think about what is the opposite of "at least one")
c) Assume that all 30 macropods were examined, how many infected animal would you in average expect in the sample?

If we want to compute probabilities for large groups, we usually use a normal approximation to the distribution above.

d) Compute the probability to observe at least 20 infected macropods among 600 if the infection rate is again 2%.

Exercise 2.3

The number of coliforms (on a log-transformed scale) were counted in milk samples.

3.9
5.3
6.1
4.9
9.1
2.8
3.5
3.2
2.6
5.9

a) **Hand calculation:** Compute a 95% confidence interval for the mean of (log-transformed) counts of coliforms assuming normal distribution for the observations. Interpret the results.

b) Outliers can be a problem in statistical analysis. A value 9.1 is observed here which is much larger than other observations. What would the effect or removing this observation be on the analysis? Consider first and then redo the confidence interval with the observation 9.1 removed (use SAS, R or hand calculations). (Remember that in reality we should never remove observations unless we have a good motivation to do so).

Exercise 2.4

Hand-calculations: We use again the milk data, where we have counted the number of coliforms (on a log-transformed scale) in milk samples. This time the data is heat treated.

With heat treatment

3.4
4.1
4.7
4.2
7.6
1.5
2.3
2.6
2.0
5.2

a) We have before seen that the number of coliforms in untreated milk is about 4.7. Conduct a one-sample statistical test to determine if the mean number of coliforms in treated milk is significantly lower than 4.7. Assume again normal distribution for the (log-transformed) observations. Interpret.
b) Test the same with a Wilcoxon signed rank test **in R** and draw conclusions.