Dirk Metzler SS 2021

STATISTICS FOR EES — EXERCISE SHEET 7

1. To measure the effect of three different fertilizers, a field was subdivided into 18 equal plots and the plots were randomly assigned to three groups A, B, C in which different fertilizers were applied. The following table shows the logarithmized yields.

fertilizer	yield						
A	2.89	2.81	2.78	2.89	2.77		
B	2.73	2.88	2.89	2.82	2.90	2.85	
C	2.84	2.81	2.80	2.66	2.83	2.58	2.80

(a) Calculate the corresponding anova table without using the R commands like anova, lm, drop1 and aov. Compute the F statistic and test the null hypothesis that fertilizers do not differ in their effect on the yield.

(b) Write down the mathematical model underlying your anova and all its underlying assumptions in precise mathematical terms.

(c) Use the R commands lm and anova to check your results from (a).

2. In file `songtime.cvs` in column "song.t" you find the total times (in minutes) that male bird of a certain species spent singing within a period of 15 Minutes under various conditions: presence or absence of a female (column "fem"), presence or absence of another male (column "mal"), presence of (a visual model of) a potential predator (cat, dog, hawk, none; column "pred"). The measurement were taked with 25 different individual, column "ind", for each individual repeatedly on different days.

(a) Fit an anova model to the data and test whether the different factors influence the average total song time. Account for possible individual effects. Use the packages `lme4` to fit the model.

(b) In particular, test whether the effect of the presence of a male depends on the presence of a female.

(c) Predict with the best-fitting model the total time that an average male will spend singing in a 15 min interval if a female and another male is present but no predator (model).

(d) Specify the fitted models with and without interaction term for the effects of the presence of a female an of anther male in precise mathematical terms. Specify all assumptions of these models.

3. (Optional; for those who watched the videos on Wilcoxon tests) Catfish were caught in three different lakes to examine whether the three populations differ in the sizes of the fish. The file `catfish.txt` contains the weight of the fish in kg.

(a) Visualize the data in an appropriate way.

(b) Apply a Kruskal-Wallis test to the data step-by-step, i.e. without using the `kruskal.test()` command.

(c) Check your results with the R command `kruskal.test()`.

(d) Apply an anova to the data and compare the results to those of the Kruskal-Wallis test. How can the differences be explained?

(e) Which conclusions do you draw for the catfish?

4. (Inspired by Ouattara et al., 2009, Animal Behaviour 78:35–44) Females of Campbell's mona monkeys can produce several different alarm calles, the main types being "wak-oos", "hoks" and "trill" calls. In a research project over several months you present three different visual predator models (eagle, leopard and snake) several times per day in randomized orderto a group of monkeys, and you record for two females (A and B) how often which alarm call was used in which case.

(a) In 87 cases female A was the one who gave the warning call. The following table shows how often she used the different calles in the different cases:

	wak-oos	hoks	trill
eagle	16	5	7
leopard	6	8	20
snake	7	15	3

Perform a statistical test to check whether A's choice of the warning call depends on the species of the predator model. Summarize the result of the test in a sentence that refers explicitly to the application context.

(b) How much did the case eagle/hoks contribute to the test statistic?

(c) Explain the number of degrees of freedom and find in the quantile table the threshold value for significance on the 5% level.

(d) (Advanced!) In 72 cases female B was the one who called, and the the following table shows how often she used which call in which case:

	wak-oos	hoks	trill
eagle	8	4	7
leopard	5	11	15
snake	4	10	8

Test for differences between females A and B in how the distribution of call types depends on the predator species. Summarize the result of the test in a sentence that refers explicitly to the application context.

5. In a study[1] about the alkaline-phosphatase gene, three distinguishable alleles "S", "I" and "F" were found. The following numbers of genotypes were observed in 332 persons: SS: 141, SF: 111, FF: 28, SI: 32, FI: 15, II: 5.

(a) Compute the relative allele frequencies for S, I and F.

(b) Use the allele frequencies to compute the expectation values for the genotypes in a sample of 332 persons, assuming a Hardy-Weinberg equilibrium for this gene.

(c) Is the observed deviation from Hardy-Weinberg equilibrium significant?

[1] Harris (1966) Enzyme polymorphism in Man. *Proc. Roy. Soc. B* **164**:1153-64