Dirk Metzler SS 2026

STATISTICS FOR EES — EXERCISE SHEET 6

1. Gene expression levels of 100 genes were measured in tissue samples from 20 different fly lines after cold shock and were tested for association with cold resistance, proxied by chill coma recovery time of flies from the same lines. The second column in the file `gene_p_values.csv` contains the p-values from these 100 tests. We aim to obtain a list of genes that are presumably associated with cold resistance.

(a) Compile a list that contains as many genes as possible, but in a way such that genes that are **not** associated with cold resistance must have 95% probability to be **not** on the list.

(b) Compile a list that contains as many genes as possible, but in a way such that the probability that the list contains only genes that are in fact associated to cold resistance is $\approx 95\%$.

(c) Compile a list that contains as many genes as possible, but in a way such that $\approx 95\%$ of the genes on your list are indeed associated with cold resistance.

2. Three groups of chicks were fed with different diets and weighed after three weeks. The results in gram were as follows.

Diet A	215	205	157	175	
Diet B	331	175	167	192	233
Diet C	322	237	238	264	

(a) Visualize the data in an appropriate way.

(b) Perform an ANOVA to test whether there is a difference in the three diets. Do not use the special R commands for anova, and use the quantile table to assess significance. In addition, use the R command `pf` to calculate the *p* value.

(c) Perform the ANOVA with the special R commands.

(d) Perform pairwise comparisons to check which diets do actually differ from each other in their effects. Don't forget to correct your p-values for multiple testing!

3. 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).

4. In file `songtime.csv` 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.

5. (Optional; for those who watched the video on the Kruskal–Wallis test) 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?