Dirk Metzler SS 2016

STATISTICS FOR EES — EXERCISE SHEET 9

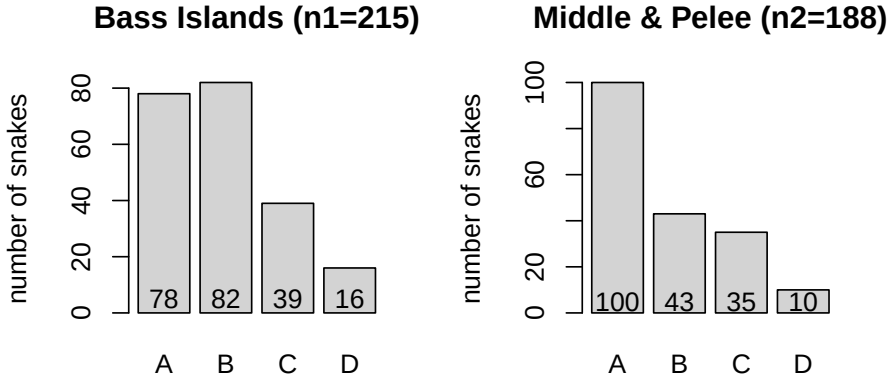
1. It may be possible that chlorinated water (as in swimming pools) damages the dental enamel. The enamel of 100 swimmers who swam less than 6 hours per week and 100 swimmers who swam more than 6 hours per week was surveyed.

	enamel damaged		
Swimming per week	Yes	No	Total
more than 6 h	29	71	100
less than 6 h	19	81	100
total	48	152	200

a) Do these observations support the hypothesis above? Perform a χ^2-test with paper, pen and pocket calculator. Which alternative test could you perform?

b) Now assume that the same study had been performed with double sample sizes of 200 swimmers of each group. Assume further that 58 of the frequent swimmers and 38 of the others had damaged dental enamel. Which value do you then get for the X^2 statistic and what is your conclusion?

2. The markings of the water snake *Nerodia sipedon* in Lake Erie can be grouped into four classes. On the mainland all snakes have strong markings (A). On the islands many snakes have no markings (A) or only weak markings(B,C). Frequencies of the four markings in samples from the Bass islands and from Middle Island and Pelee Island are as follows.



(a) Let H_0 be the hypothesis that the distribution of marking classes were the same for all islands. What is the expected number of snakes of class A sampled from the Bass Islands if we assume that H_0 is true?

(b) We apply a Chi-square test with null hypothesis H_o. How much does the table entry (Bass Islands : class A) contribute to the test statistic X^2?

(c) If we sum $(O_i - E_i)^2/E_i$ for all table entries we find $X^2 = 14.74$. How many entries does the table have? How many degrees of freedom does X^2 have?

(d) Using the quantile tables on the website of the course, what can you say about the p-value in the snake example?

3. 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?

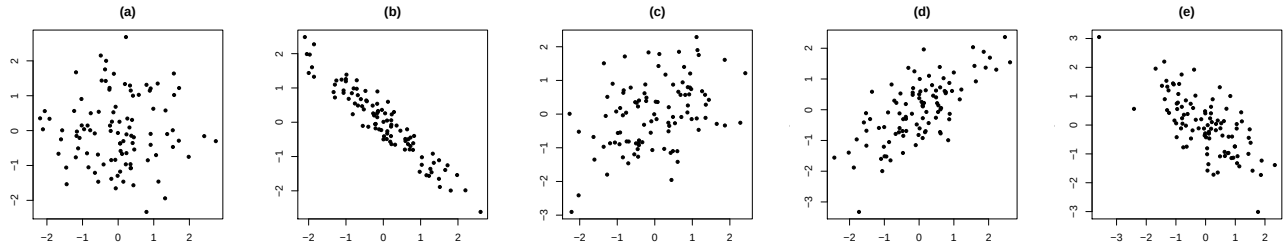
4. A simple medical test indicates influenza for 98% of the infected and for 3% of the non-infected. Currently, 1% of those that have the symptoms of a cold are infected with influenza. I worry because of my running nose, so I decide to make a test. If the test indicates influenza, what is the probability that I am actually infected with it?

5. In a survey about the efficacy of a hypnotic the additional hours of sleep per night when taking the hypnotic was measured for 10 test persons: :

1.9 0.8 1.1 0.1 -0.1 4.4 5.5 1.6 4.6 3.4

The mean value of the additional hours of sleep is $\mu = 2.33$ with variance $s^2 = 4.01$. Is the increase in sleep time significant? (You may assume that the data are normally distributed.) Don't forget to state the null hypothesis .

6. a) The scatterplots (a)–(e) below show datasets with various correlation coefficients. Assign the figures to the correlation coefficients.



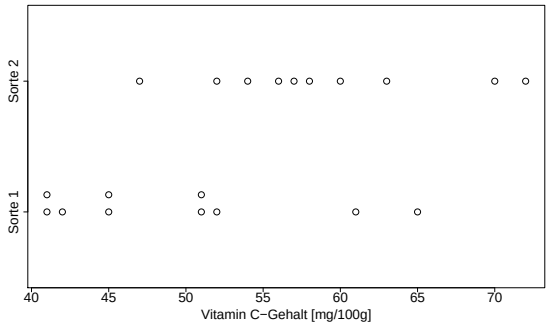
Correlation coefficient	-0.95	-0.7	0	0.4	0.7
plot					

b) Compute for the following mini dataset the mean and the sample variance of the x and the y values and the covariance and correlation of x and y:

x	1	3	2
y	0	2	-2

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

7. The vitamin C contents of two types of cabbage were surveyed in 10 samples per type:



For the samples of type 1 a mean content of $\mu_1 = 49.4$ mg per 100g with standard deviation $s_1 = 8.33$ was found, for type 2 the values were $\mu_2 = 58.9$ and $s_2 = 7.74$. Test the hypothesis that the mean vitamin C contents is the same for both types with a t-test. You may assume that the true variances are equal.