

## Exercise Sheet 10

1. The concentration of a certain substance is measured in a river on a daily basis. After a clean-up of the river, the number of days on which the environmental limit are breached in the subsequent 6 months are 2, 3, 1, 5, 6 and 3.

a) Compute a 95% confidence interval for the mean number of breaches of the environmental limit per month. (Rule of thumb: The mean of a distribution,  $\pm 2$  standard deviations, contains around 95% of its probability.)

Hint: Assume that the number of breaches in the limit per month follows a Poisson distribution, and that the numbers from each month are independent.

b) Before the clean-up process, the long-term average was 6 breaches of the environmental limit per month. Has the subsequent decrease been significant?

Perform a one-sided test at level 2.5%.

2. A certain proportion of pigeons in the city of Zurich are afflicted with disease A. To gain an estimate of the prevalence of this condition, the municipal hunter shoots 50 pigeons, which are then tested for disease A. The tests reveal that 12 of these pigeons had contracted the disease.

a) Estimate the probability  $p$  that a given pigeon has disease A. Which model best describes the outcome of the shooting experiment?

b) Compute the standard error (empirical standard deviation) of  $\hat{p}$ .

c) Rumour has it that a full 40% of all pigeons have disease A. Test this rumour at level 5% using a confidence interval. Furthermore, formulate the null and alternative hypotheses.

d) The following probabilities are correct for for binomial distribution with sample size  $n = 50$ :

- when  $p = 0.105$ ,  $P[\text{number of diseased pigeons} \geq 12] = 0.005$ ,

- when  $p = 0.425$ ,  $P[\text{number of diseased pigeons} \leq 12] = 0.005$ .

What do you conclude from this?

3. The following data give us the (scaled) concentration of a certain chemical substance in 10 docked broad bean plants, and in 10 plants with roots attached.

Data:	docked plants	53	58	48	18	55	42	50	47	51	45
	plants with roots	36	33	40	43	25	38	41	46	34	29
Statistics:	docked plants	$\bar{x} = 46.7, s_x = 11.14$									
	plants with roots	$\bar{y} = 36.5, s_y = 6.45$									

a) Are these samples connected (paired) or unconnected?

- b) Perform a  $t$ -test to investigate whether the observed difference in means is significantly different from 0 at level 5%. Write down the null and alternative hypotheses, the test statistic and the rejection set.
- c) In part b), the computer returns the p-value 0.022. The p-value from the Wilcoxon (rank sum) test is 0.0039. Interpret these p-values (you can also use the following normal QQ plots for this).

