Statistics for the Doctoral School in Biomolecular Sciences
Academic year 2014-2015

Exercises after Lecture 4

Confidence interavls and tests of hypotheses

[Some computation require the use a calculator for computing sample means, variances; a statistical program, like R, can be necessary to obtain quantiles of the appropriate distributions; apart from this and unless otherwise remarked, test s can be done by hand. Using also R, or other software, for everything is however advised.]

EXERCISE 4.1 Burch and Chao (1999) have propagated 5 lines from a clone of the RNA bacteriophage f6. After 100 days they have computed the fitness of every line relatively to the original population obtaining the levels

$$0.063 \ -0.062 \ 0.064 \ -0.043 \ 1.34$$

Can we say fitness has changed? Compute a 95% confidence interval?

EXERCISE 4.2 Explain whether in a test for differences in the mean of the two groups, the following examples should be treated as paired or independent sample?

- weight of 14 patients before and after heart surgery
- number of cigarettes smoked by 14 men who were affected by a stroke, and 14 men who were not
- basal metabolism of 7 chimpanzees vs. 7 gorillas
- photosynthesic rate of leaves near the top of 10 Sitka spruces with that of leaves near the base of the same trees
- photosynthesic rate of leaves of 10 Sitka spruces with that of 10 Douglas firs randomly chosen
- photosynthesic rate of leaves of 10 Sitka spruces with that of the Douglas fir growing nearby

EXERCISE 4.3 A pharmaceutical company studied the efficacy of an antiviral to prevent infection from a flu virus. Experiments on a sample of ferrets yielded the following results

Treatment	No infection	Mild symptoms	Severe symptoms
Antiviral	12	19	4
Placebo	7	12	16

Study whether we can say, at a significance level of 5%, that the antiviral is effective.

What would be the result if we grouped together all infected animals (whether the symptoms are mild or severe). Comment the results.

[It is enough to present the procedure (in both cases, with or without the grouping) without performing the computations, and explaining with which number we should compare the test statistics.]

EXERCISE 4.4 In a study on the importance of cicadas as nutrient source in the forest, some researchers measured the fungin phospholipidic fatty acids (PLFA) in 44 plots of 1 square metre each. 22 of these plots were then used as control, while in the other 22 120 cicada corpses were added ("cicada added"). 28 days later PLFA were measured again in the same 44 plots, obtaining the following results:

Treatment	average PLFA	Std. dev. PLFA	number of plots
Control: before	7.16	3.39	22
Control: after	4.50	2.3	22
cicada added: before	6.97	3.29	22
cicada added: after	6.31	2.12	22

1. Compute a 95% confidence interval for the value of mean PLFA in each of the 4 groups.
2. Test for a difference in mean PLFA between control plots and"cicada added" plots, separately for before and after the experiment.
3. Test for a difference in mean PLFA between before and after the experiment, separately for control plots and"cicada added" plots.
4. Which are the assumptions used in these test? From the data do they seem verified? What could be suggested?

Note that from the data shown, it is not possible to answer to all questions. If information is missing, please specifty what would be needed.

EXERCISE 4.5 A cross between two pure varieties of peas has yielded, at the F_2 generation, the following result

DD	Dr	rD	rr	Totale
366	106	96	32	600

where D and r represent dominant and recessive genes at two loci.

Check whether results are compatible with Mendel laws which predict frequencies of 9/16, 3/16, 3/16, 1/16 for the four groups.

[It is enough to present the procedure (in both cases, with or without the grouping) without performing the computations, and explaining with which number we should compare the test statistics.]

EXERCISE 4.6 In a study on the effect of Nifedipine (an antidepressive) on arterial pressure, Nifedipine has been given to a sample of 10 dogs, and a placebo to another sample of 12 dogs. Later, the arterial pressure has been measured in the two groups: in the sample treated with Nifedipine the sample mean was 102.45 with sample variance 281.87, while in the sample treated with the placebo the mean was 128.91 and the variance498.69.

Are these information enough to perform a test on the difference in mean arterial pressure in the two groups? Which assumptions would be used?

Under these assumptions, test (at a significance level of 1%) the null-hypothesis that Nifedipine has no influence on arterial pressure.

EXERCISE 4.7 In order to test whether American coffee has an effect on heart rate, this was measured in 10 volunteers before and after having drunk half a litre of American coffee. The results are:

Volunteer	Before	After	Volunteer	Before	After
1	60	66	6	88	89
2	63	63	7	69	67
3	86	92	8	72	69
4	72	79	9	75	78
5	69	70	10	80	84

Conclude whether we can reject the null hypothesis that American coffee has no effect on heart rate. *[Write down the exact procedure to be followed, but do not perform the tests by hand.]*

Explain how the hypothesis can be tested in a non-parametric way, and which are advantages and disadvantages of that? Would it be useful in this case?

EXERCISE 4.8 A certain number of skulls of two species of *Austrolapithecus*, A e B., have been collected The widths (in mm) of the cranium are the following:

Species A	82	70	75	81	75	88
Species B	69	74	71	63	78	

Can we conclude, at the significance level of 10%, that the mean cranial width was different in the two species?

EXERCISE 4.9 In a hospital there were 932 births in a period of 20 consecutive weeks. Of these, 216 births occured in the weekends? From these data, can we infer that birth rates in the week-ends are different from those in work days?

Write down the null and the alternative hypotheses of an appropriate test. Perform it and state clearly the conclusions.

EXERCISE 4.10 Rodgers and Doughty (2001) examined data on US families, obtaining the following data on the sex of children in a random sample of 2,444 families with 2 children:

Male number	Frequency
0	530
1	1332
2	582
Total	2444

Conclude if these data are compatible with the hypothesis that the number of male children follows a binomial distribution:

a) with $p = 1/2$;

b) with p estimated from data.

EXERCISE 4.11 In a sociological study the on-the-job injuries in a group of 647 workers in a bullet factory were monitored over a 5 week period. The results are in the table below. Examine whether these data are compatible with a Poisson distribution. Interpret the result

Number of injuries per worker	Frequency
0	447
1	132
2	42
3	21
4	3
5	2
Total	647