NTNU

Faculty of Medicine
Department of Cancer Research and Molecular Medicine

EXAM
KLH3004 Medisinsk statistikk (Medical statistics)
KLMED8004 Medisinsk statistikk, del I
(Medical Statistics Part I)

Wednesday December 16, 2009, 09.00-13.00

ECTS credits: 7.5
Supporting materials: Calculator and all written and printed material permitted.
Number of pages (including front page): 5

Contact person during exam:
Stian Lydersen. Tel 72575428 / mobile 92632393

Examination results: 20 January 2010
Examination results are announced on
http://studweb.ntnu.no/

Problem 1: Probability and illness

We are studying a non-infectious disease and observe families consisting of one child and one adult. We assume that

- the probability that the adult gets the disease is 0.3, and
- the probability that the child gets the disease is 0.1.

a)
Let us first assume that the members of the family get the disease independently of each other.

What is the probability that both the adult and the child in a family get the disease?
Calculate the probability that at least one of the family members gets the disease.

b)
Let us now assume that the families we are studying have common environmental and genetical factors, such that the members of the family do not get the disease independently of each other. Assume that the probability that both the adult and the child get the disease is 0.08.

Calculate the probability that at least one of the family members gets the disease.
Compare this answer with the answer in a) and comment on your findings.

Problem 2: HPV-vaccine

In Norway 12 year old girls (7^{th} grade) are offered, without charge, vaccine against Human papillomavirus (HPV-vaccine). One hopes to achieve a vaccine coverage of 80%. That is, 80% of the those given the offer to take the vaccine accepts the offer and 20% does not accept the offer.

a)
At one primary school there are ten girls enrolled in the 7^{th} grade. All the girls at this grade are invited to be take part in the vaccination programme.

If we assume a vaccine coverage of 80%, how many girls do we expect to not accept the vaccination offer?
What is the probability that all the ten girls accept the vaccination offer?
State the assumptions you have made for calculating this probability.

b)
It turns out that at the primary school in question a total of eight of the ten girls do not accept the offer to take the vaccine.

Given that we believe there is a 80% vaccination coverage in Norway, calculate the probability that we in a 7^{th} grade with ten girls observe that at least eight girls do not accept the offer to take the vaccine.
What can you conclude from this proabability calculation?

Problem 3: The normal distribution with multiple choice questions

a)
Assume that the weight of a newborn boy is normally distributed with expected value (mean) 3.60 kg and standard deviation 0.50 kg. The probability that a newborn boy will weigh at least four kg is then equal to:

i) 0.055
ii) 0.212
iii) 0.309
iv) 0.345
v) 0.788

b)
The vital capacity is a measure of a persons lung function. It is the maximum amount of air a person can expel from the lungs after a maximum inspiration. Assume that for healthy 12 years old boys the vital capacity is normally distributed with expected value (mean) 3.0 litre and standard deviation 0.4 litre. When a physician examines the vital capacity of a patient she/he is interested in knowing if the vital capacity is substantially lower than normal. This may be a sign of illness of the lungs and should lead to further investigations. A lower threshold (limit) for a normal vital capacity is given as the value where 97.5% of the healthy population has a vital capacity above this threshold. For 12 years old boys this threshold is:

i) 1.0 litre
ii) 2.0 litres
iii) 2.2 litres
iv) 2.4 litres
v) 2.6 litres

Problem 4: Training in a randomized controlled trial

In a randomized controlled trial, patients were randomized to two types of training aimed at enhancing their endurance. After completing the training program, the endurance measured in treadmill time in seconds were as follows:

Group 1:
1020, 840, 1000, 1030, 800, 1045, 650, 1160
Average: 943.13. Sample standard deviation: 165.247.

Group 2:
775, 550, 725, 1415, 590, 680, 735, 870, 595, 540
Average: 747.50. Sample standard deviation: 257.480.

We shall study possible differences between the groups.

a) Calculate the median in each group.

b) Figure 1 shows a box plot of the treadmill times. What does the bottom of the box, the top of the box, and the horizontal line in a box represent? What does the asterisk (*) above the box for group 2 mean?

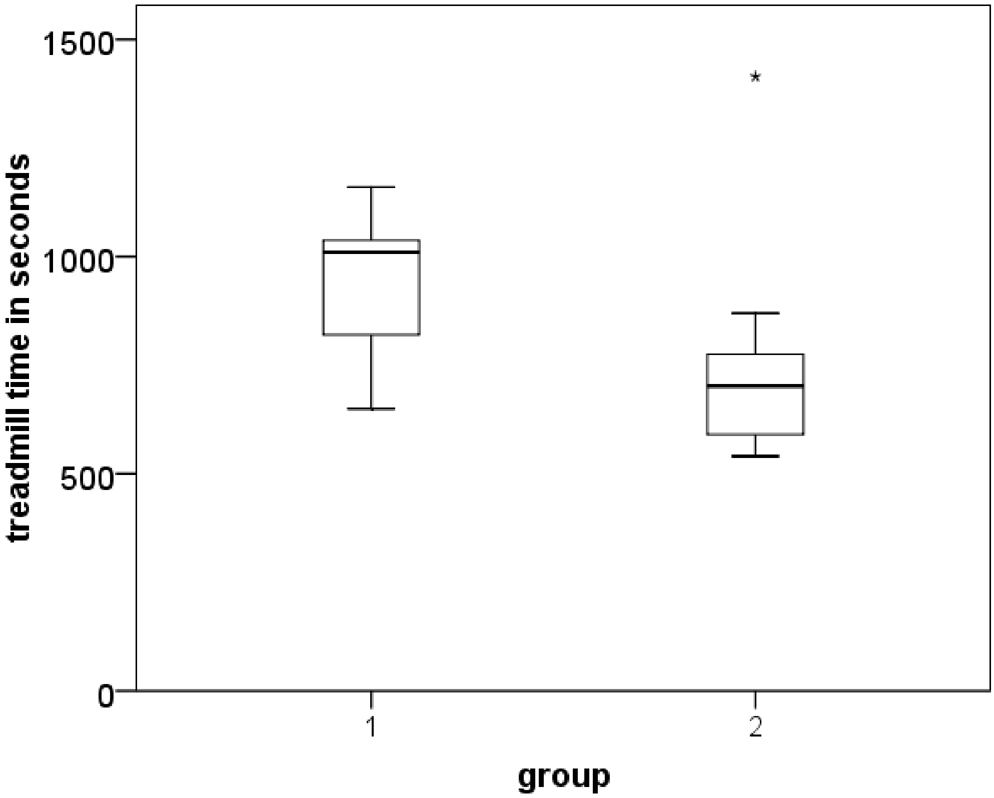

Figure 1.

Under c) and d), the times are assumed to be normally distributed, and the two groups are assumed to have the same (unknown) variance.

c)
Does the expected time differ between the groups? State the null hypothesis, the alternative hypothesis, and carry out a test at significance level 5%. What is the conclusion?

d)
Calculate a 95% confidence interval for the difference in expected time between the two groups. Is this in accordance with the conclusion in c)? Why or why not?

e)
The Wilcoxon Rank-Sum test (also called the Mann-Whitney test) is an alternative to the test used in c). Calculate the rank sum for Group 1. What is the expected rank sum under the null hypothesis? The corresponding p-value is 0.026 (you shall not calculate it). What can you conclude from these results?

f)
The test in c) yields a p-value of 0.082, and the p-value from e) is 0.026.
Compare c) and e) with respect to assumptions made and conclusions of the tests. Which method will you recommend here?

Problem 5: Power and sample size

You are planning a new randomized controlled trial to compare two training programs, and you have carried out the following sample size calculation:

With expected endurance in the two groups of 950 and 750 seconds and common standard deviation 200 seconds, to achieve power 80% at significance level 5% you need $2 \cdot 17 = 34$ participants.

If you want to achieve power 90% instead of 80%, what can you say about the needed sample size?

i) $2 \cdot 17$
ii) Smaller than $2 \cdot 17$
iii) Larger than $2 \cdot 17$
iv) Cannot say without computing the sample size