Norwegian University of Science and Technology
Department of Cancer Research and Molecular Medicine

Contact person during exam:
Øyvind Salvesen (93 64 87 72)

EXAM
KLMED8004 Medical statistics I
KLH3004 Medisinsk statistics I

June 10, 2010
09:00–13:00

ECTS credits: 7.5

Supporting materials:

- All written and printed material permitted.
- Calculator.

Examination results: July 2, 2010.
Examination results are announced on `http://studweb.ntnu.no/`.

ENGLISH

Problem 1 Hay fever and eczema

In a population of 11 year old children the prevalence of hay fever and eczema has been studied.

Define two events:

- E= a randomly chosen child from the popuation has eczema.
- H= a randomly chosen child from the popuation has hay fever.

Assume that in this population we have the following probabilities:

$$\begin{aligned} P(E) &= 0.04 \\ P(H) &= 0.07 \\ P(E \cap H) &= 0.009 \end{aligned}$$

a) Draw a Venn diagram depicturing the to events.
Explain in your own words what $P(E \cap H) = 0.009$ means.
Explain in your own words what independence between E and H means.
Are the events E and H independent?

b) Among the children with hay fever in the populations, we randomly select one child. What is the probability that this child has eczema?

Among the children that do not have eczema in the population, we randomly select one child. What is the probability that this child does not have hay fever?

Problem 2 Smoking habits

a) We study a population of women between the age of 35 and 50 years. Assume that the probability that a randomly chosen woman from this population never has smoked is 0.5.

We draw a random sample of size 10 from this population.

In this sample, what is the expected number of women that has never smoked?
Find the probability that at least two women in this sample have never smoked.

b) Let us now assume that we study a population consisting of women between the age for 35 and 50, but that the proportion of women that have never smoked in this population is unknown. We want to use a cohort study to estimate this proportion.

We have sampled a cohort of 24505 women, and find that among them 11823 had never smoked.

Find a point estimate and a 95% confidence interval for the proportion of women between 35 and 50 years that have never smoked. How can you interpret this confidence interval?

Problem 3 IQ

An IQ test is constructed so that the IQ score of a randomly chosen person will be normally distributed with mean score 100 and standard deviation 15.

a) What is the probability that a randomly chosen person will have an IQ score of at least 110?

b) Define a number x such that there is a probability of 90% that a randomly chosen person will have an IQ score lower than x. What is x?

Problem 4 LDL Cholesterol

In a clinical study the aim was to examine the effect of a low carbohydrate diet on the level of LDL (low density lipoprotein) cholesterol in blood (mmol/litre).

The participants in the study were women between 18 and 30 years of age, and had a BMI (body mass index) between 24.5 and 27.5 (kg/m^2). A total of 13 women participated in the study.

For each participant in the study the effect of diet on the LDL cholesterol level in blood was calculated as the difference between the LDL cholesterol level after the diet and before the diet. A negative value of this difference means that the LDL cholesterol level after the diet is lower than the LDL cholesterol level before the diet.

Data from the study are presented in Table 1 and Figure 1.

Person	Measured effect (mmol/litre)
1	0.020
2	0.575
3	0.130
4	-0.175
5	-0.540
6	-0.190
7	0.455
8	-0.340
9	-0.460
10	-0.790
11	0.090
12	-0.300
13	-1.303
average	-0.218
sample standard deviation	0.504

Table 1: Table of measured effect of diet on LDL cholesterol in blood.

a) Consider the data presented in Table 1. Calculate the median measured effect of diet on LDL cholesterol in blood.

Look at the boxplot in Figure 1. What does the bottom of the box, the horizontal line in a box, and the top of the box represent?

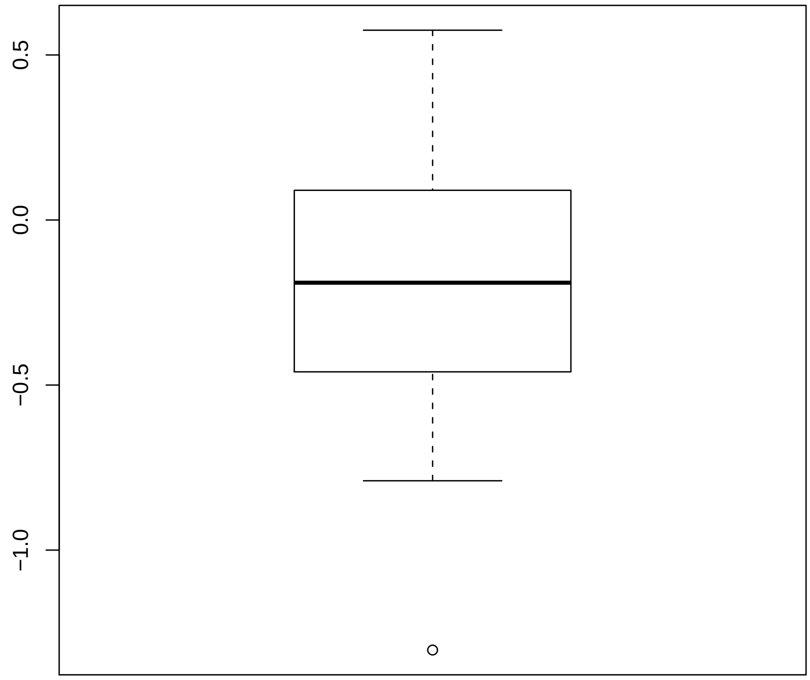

Figure 1: Boxplot of the measured effect of diet on LDL cholesterol in blood.

b) Assume that the measured effect of diet on LDL cholesterol in blood is normally distributed.
Is the LDL cholesterol level in blood changed because of the diet?
State the null hypothesis, the alternative hypothesis, and carry out a test at significance level 5%.
What is the conclusion?

c) The hypothesis in b) can also be tested by a non-parametric test. What are the reasons to use a non-parametric test?

The Wilcoxons Signed-Rank test was conducted on the data set presented in Table 1. The test gave a p-value of 0.17. Which null and alternative hypothesis is underlying the Wilcoxon Signed-Rank test? What will you conclude from this?