Norwegian University of Science and Technology
Department of Mathematical Sciences

English

Contact during examination:

Ingelin Steinsland	926 63 096
Ola Diserud	932 18 823
Arvid Næss	995 38 350

TMA4245 Statistics

Saturday 26 May 2012 9:00–13:00

Permitted aids: Yellow A5 sheet with your own handwritten notes (stamped by the Department of Mathematical Sciences), *Tabeller og formler i statistikk* (Tapir forlag), *Matematisk formelsamling* (K. Rottmann), calculator HP 30s or Citizen SR-270X

Grades to be announced: 18 June 2012

Problem 1 Snow density

The density (mass density) of snow is at most 1 kg/dm^3 (very wet snow). Assume that the probability density of the snow density in kg/dm^3 of a randomly chosen snow sample is given by $f(x) = \beta(\beta+1)x(1-x)^{\beta-1}$, $0 \le x \le 1$, where β is a positive parameter.

a) Assume, only here, that $\beta = 2$.

What is the probability that the density of a randomly chosen snow sample is between 0.5 and 0.9 kg/dm^3?

b) Find the maximum likelihood estimator of β based on a random sample $X_1, X_2, \ldots, X_n$ of snow densities.

What is the estimate if $n = 100$ and $\sum_{i=1}^{n} \ln(1-x_i) = -104.0$?

Problem 2 Temperature in March og April

This year (2012) many places in Norway were colder in April than in March.

Let X be the average temperature in March and Y the average temperature in April at Værnes a randomly chosen year, both measured in °C. Assume that X has the normal distribution with mean (expected value) μ_{m} and variance σ^2, and that Y has the normal distribution with mean μ_{a} and variance σ^2.

a) Assume that we have data in the form of values of X and Y for a random sample of years. Explain how you graphically can assess whether the assumption of normal distribution is satisfied. How can you graphically investigate whether X and Y are independent?

Give som rough sketches showing how such graphs may look like both when the assumption of normal distribution is satisfied and when it is not, and also when X and Y are independent and when they are not.

The average temperature in °C at Værnes for the years 2001–2012 were as follows:

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
x_i (March)	−2.5	0.5	3.3	2.6	−0.7	−4.6	3.3	0.8	1.9	−0.5	1.2	3.8
y_i (April)	4.1	7.2	5.0	7.9	5.8	4.9	5.0	5.9	6.9	4.8	6.7	3.2

It is given that $\sum_{i=1}^{12} x_i = 9{,}10$, $\sum_{i=1}^{12} y_i = 67{,}40$, $\sum_{i=1}^{12} x_i^2 = 77{,}07$, $\sum_{i=1}^{12} y_i^2 = 399{,}30$, $\sum_{i=1}^{12}(x_i - y_i)^2 = 364{,}53$, where $i = 1$ stands for year 2001, $i = 2$ for 2002 etc.

b) Assume that the March temperatures are independent from year to year. Find a 99% confidence interval for the expected average temperature in March.

We want to use hypothesis testing to try to show that the difference between expected average temperatures in April and March is less than 5 °C. Also assume that the April temperatures are independent from year to year.

c) Set up the hypotheses.

We can either use a two-sample test or a paired test. Which would you choose? Argue for your choice, and perform the test you choose. Use significance level $\alpha = 0.05$.

Problem 3 Chemical factory

A machine at a factory performs a procedure to make a chemical. A poisonous by-product is formed in an amount of X grams every time the machine performs the procedure, where X has the normal distribution with mean 20 and standard deviation 4. If more than 25 g of the by-product is formed, a warning lamp lights up and stays lit until the procedure is finished. The machine can be set so that it performs the procedure multiple times, but it cannot be stopped until all is finished. The amount of by-product is independent each time the procedure is performed.

a) What is the probability that the lamp lights up when the machine performs the procedure once?

The machine performs the procedure three times. What is the probability that the lamp lights up at least once?

The machine performs the procedure 100 times. Find an approximate probability that the lamp lights up 15 times or more.

b) The pollution authorities require that the probability that the machine produces 500 g or more by-product in one day, should be 0.01 or less. How many times can the machine perform the procedure during of one day for this requirement to be satisfied?

Problem 4 Barbecue tonight?

A party of students wants to take a well-deserved evening off exam preparations after the statistics exam. They want to barbecue, but subject to weather conditions: It should not rain, and not be too cold (below 15 °C).

a) Let A be the event that it rains, B the event that it is cold (below 15 °C) and C the event that it is weather for barbecue (15 °C or more, and not rain).

Illustrate the events in a Venn diagram.

Further, we know that $P(A) = P(B) = 0.4$, and $P(A \cap B) = 0.2$.

Are A and B disjoint? Are A and B independent?

Find $P(C)$.

Are A and C disjoint? Find $P(C \mid A')$.

(Justify your answers briefly.)

They postpone the decision to immediately after the exam, at 13:00, and want to consider whether there will be a barbecue based on the weather then. To make a best possible choice, they make a simple linear regression model, with temperature at 13:00, x, as regressor (independent variable) and temperature at 20:00, Y, as response (dependent variable), both measured in °C: $Y = \alpha + \beta x + \epsilon$, where ϵ has the normal distribution with mean 0 and variance σ^2.

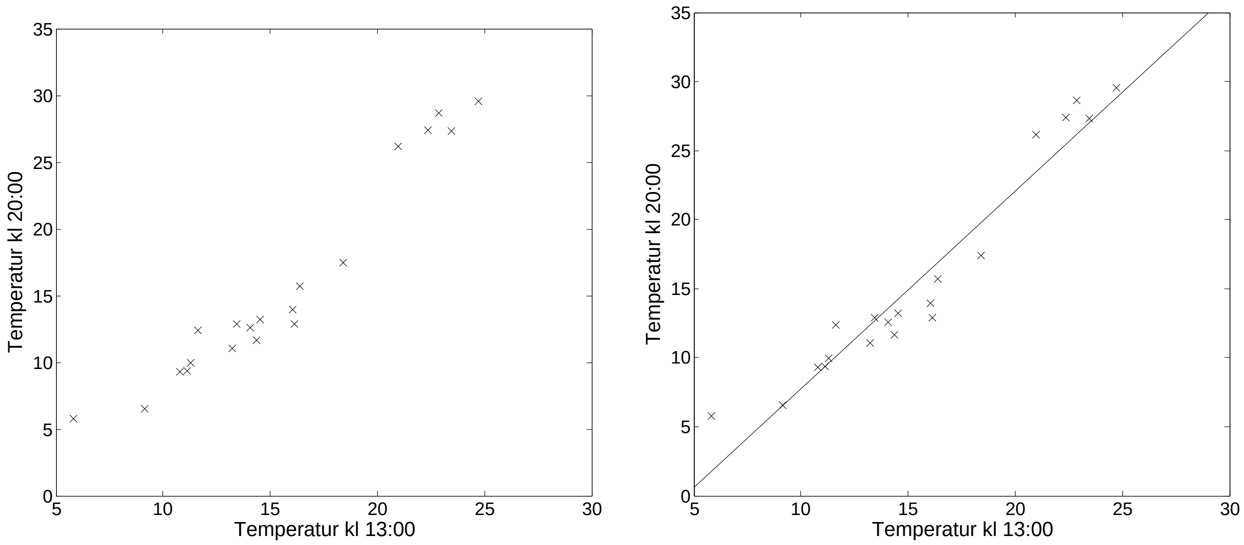

Figure 1: Temperature at 13:00 and 20:00 on 26 May in the years 1992–2011. The estimated regression line is shown to the right.

Let x_i be the temperature at 13:00 and y_i the temperature at 20:00 on 26 May in the years 1992–2011, where $i = 1$ stands for year 1992, $i = 2$ for 1993 etc. ($n = 20$ years). The landlord of one of the students has data available (Figure 1, left). We assume that the data are a random sample from the mentioned regression model.

Let $\hat{\alpha}$ and $\hat{\beta}$ be the least squares estimators of α and β. Below, you may without proof use the estimators given in *Tabeller og formler i statistikk*, and that $\hat{\beta}$ is unbiased, normally distributed, has variance $\sigma^2 / \sum_{i=1}^n (x_i - \bar{x})^2$ and is independent of $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$.

b) Find the probability distribution, expected value and variance of $\hat{\alpha}$. Briefly justify your answers and steps of the computations.

List the assumptions of the regression model. If possible, assess from Figure 1 whether these are satisfied.

The estimates based on the data are $\hat{\alpha} = -6.57$, $\hat{\beta} = 1.43$ and $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{\alpha} - \hat{\beta} x_i)^2 = 2.04^2$. Further, $\bar{x} = 15.5$ and $\sum_{i=1}^n (x_i - \bar{x})^2 = 510.7$. The estimated regression line is drawn in Figure 1 (right).

c) Find an estimate of the expected temperature at 20:00 if the temperature at 13:00 is 15 °C.

Find a 95% prediction interval for the temperature at 20:00 if the temperature at 13:00 is 15 °C.