Norwegian University of Science and Technology
Department of Mathematical Sciences

English

Contact during exam:
Ingelin Steinsland 926 63 096
Øyvind Bakke 990 41 673
Ola Diserud 932 18 823

EXAM IN TMA4245 STATISTICS

June 3rd 2011
Hours: 09:00–13:00

Permitted aids: *Tabeller og formler i statistikk*, Tapir Forlag
K. Rottmann: *Matematisk formelsamling*
Calculator HP30S / CITIZEN SR-270X
Yellow, stamped A5-sheet with your own handwritten notes.

Examination results are due: 24. juni 2011

Problem 1 Residence in Hawaii

From 1832 to 1950, the volcano Mauna Loa in Hawaii had 37 eruptions, yielding an average rate (intensity) of 0.026 eruptions per month. Over short geological periods we assume that a volcano's eruptions follow a Poisson-prosess in time. Due to studies at the University of Hawaii, you get a very tempting offer for a five year tenancy agreement for a house at the base of Mauna Loa.

a) Describe shortly the properties that have to be satisfied for a volcano's eruptions to be a Poisson process in time.

What is the probability that there will be at least one eruption during the tenancy period of five years? Apply a rate (intensity) as from 1832 to 1950.

You are told that Mauna Loa had an eruption six months before the starting date of the tenancy agreement. What is the probability that the next eruption will occur more than three years after the starting date?

Problem 2 The housing market in Trondheim

Petter just accepted a new job offer in Trondheim, and is searching for a new apartment. As his new job is located in Midtbyen (the city center) he mainly searches for an apartment there. For a sample space consisting of all apartments for sale in Trondheim at `finn.no`, Petter draws one at random (all apartments are equally likely to be drawn). We define these events:

M: The apartment is in Midtbyen
T: The apartment has a suggested price below 2 mill. kr

a) Draw the events in a Venn diagram.

There are 381 apartments for sale in Trondheim. Out of these, 94 are in Midtbyen and 190 have a suggested price below 2 mill. Out of the apartments in Midtbyen, 50 have a suggested price below 2 mill.

Are the events M and T disjoint?
Are the events M and T independent?

Justify and comment on your answers.

Each apartment is sold after a bidding round, where the highest bidder gets the apartment. We assume that the selling price per square meter for apartments in Midtbyen can be modeled by the regression model

$$Y = \beta_1 x + \epsilon(x),$$

where Y is the selling price (the actual price paid) (kr per m^2), x is the suggested price (kr per m^2) and $\epsilon(x)$ is normally distributed (Gaussian distributed) with expected value $E(\epsilon) = 0$ and variance $\text{Var}(\epsilon) = \tau^2 x^2$ where $\tau = 0.1$ is a known parameter.

b) Assume in this part only that $\beta_1 = 1.1$.

Explain the meaning of β_1 being greater than 1.

An apartment of size 60 m^2 is put out for sale with suggested price 1.8 mill. kr. Define W as the selling price (in million kroner), i.e. what is payed for this apartment.

Show that W is normally distributed with expected value 1.98 and standard deviation 0.18.

Find the probability of the selling price being larger than 2 mill. kr for this apartment.

Petter wants to examine the ratio between the suggested price and the selling price, and observes the suggested and selling price (in 1000 kr/m^2) for $N = 30$ apartments in Midtbyen. The data found are plotted in figure 1a, and $\sum_{i=1}^{N} \frac{y_i}{x_i} = 32.98$.

c) If we assume independent observations, the maximum likelihood estimator of β_1 is (not to be shown)

$$\hat{\beta}_1 = \frac{1}{N}\sum_{i=1}^{N}\frac{Y_i}{x_i}.$$

Show that $\hat{\beta}_1$ is normally distributed with expected value β_1 and variance $\frac{\tau^2}{N}$.

Derive a 95% confidence interval for β_1 and find the interval numerically based on the observations from Midtbyen.

d) For Petter, Tyholt is also a suitable place to live. We assume the regression model

$$V = \beta_2 x + \epsilon(x)$$

for the selling price per square metre for apartments at Tyholt, where ϵ is the same as in (b), i.e. $\epsilon(x)$ is normally distributed with expected value 0 and variance $\tau^2 x^2$ where $\tau = 0.1$. He wants to test whether Midtbyen and Tyholt has the same ratio between suggested price and expected selling price or not.

State this as a hypothesis test.

Data for $M = 50$ apartments at Tyholt are given in figure 1b, and $\sum_{i=1}^{M}\frac{v_i}{x_i} = 56.66$.

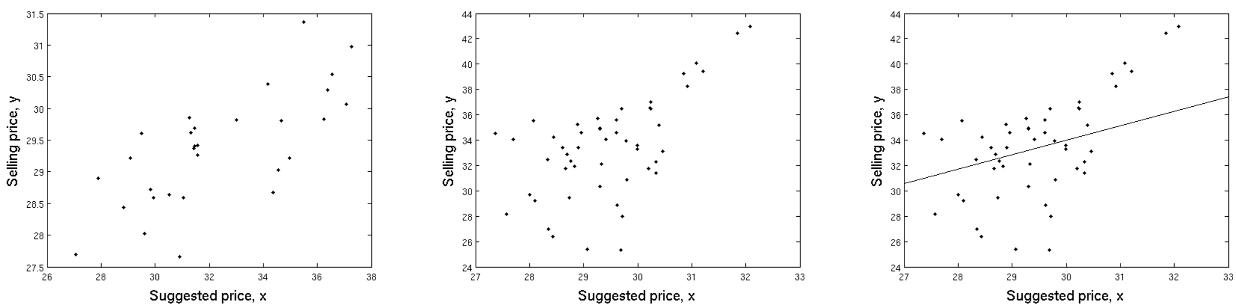
Perform the test based on the data from Midtbyen and Tyholt. What is the conclusion of the test using a 0.05 level of significance?

e) For Tyholt, β_2 is estimated to be $\hat{\beta}_2 = 1.13$ based on the data, which resulting regression line is given in figure 1c.

List the model assumptions that have been made.

From the figure, do the model assumptions seem to hold? Explain.

Explain how to further examine the assumptions graphically.



(a) Midtbyen (b) Tyholt (c) Tyholt

Figure 1: Data for suggested price and selling price for (a) $N = 30$ apartments in Midtbyen, (b) and (c) $M = 50$ apartments at Tyholt. In (c), the regression line $v = \hat{\beta}_2 x$ is included.

Problem 3 The prefab house factory

The factory House-in-Time produces prefab house modules. Production time, i.e., the number of weeks from start of production of a house until the house is ready for transportation, has the exponential distribution with expected value μ, i.e., with a probability density given by $\frac{1}{\mu}e^{-x/\mu}$ for $x > 0$.

a) Let $\mu = 2$ (weeks) in this part only. What is the probability that production time is less than 1 week?

What is the probability that among 5 houses there is no production time less than 1 week? Assume that production times are mutually independent.

The factory does not find average production time satisfactory, and engage the consulting firm GoodDearAdvice. GoodDearAdvice suggests an alternative production process, which according to the consultants will give an average production time that is c times shorter than that of the old production process, i.e., production time for the new production process has a probability density given by $\frac{c}{\mu}e^{-cy/\mu}$ for $y > 0$.

Further, assume that μ is unknown. House-in-Time wishes to estimate μ based on n_1 production times $X_1, X_2, \ldots, X_{n_1}$ using the old production process and n_2 production times $Y_1, Y_2, \ldots, Y_{n_2}$ using the new one. They utilize the estimator $\tilde{\mu} = \alpha\bar{X} + \beta\bar{Y}$, where $\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$ and $\bar{Y} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$. Assume that all production times are mutually independent, and that c is known.

b) Assume, in this part only, that $c = 2$, $\alpha = \frac{1}{2}$ and $\beta = 1$.

Calculate the expected value and the variance of $\tilde{\mu}$. Is $\tilde{\mu}$ unbiased?

Derive a 95% confidence interval for μ based on $\tilde{\mu}$, and find its numerical value when $n_1 = 30$, $n_2 = 20$, $\bar{x} = 2.07$ and $\bar{y} = 0.59$. State which approximations or assumptions you make, if any.

c) Determine α and β so that $\tilde{\mu}$ is unbiased and have the least possible variance (among unbiased estimators of this form).

What is the variance?

As a newly-hired engineer at House-in-Time in charge of quality and delivery realibility, you have a suspicion that the value stated by GoodDearAdvice for c is not correct. Based on X_1, $X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ you want to estimate c along with μ.

d) Find the maximum likelihood estimator $(\hat{\mu}, \hat{c})$ of (μ, c)

Also determine the maximum likelihood estimator μ^* of μ when we consider c known, and compare with the estimator $\tilde{\mu}$ of (b).

(You are not required to argue that critical points are maxima.)