Norwegian University of Science and Technology
Department of Mathematical Sciences

Engelsk

Contact under exam:
Ingelin Steinsland 73 55 02 39/ 92 66 30 96
Jo Eidsvik 73 59 01 53/ 90 12 74 72

EXAM IN COURSE TMA4240 STATISTICS

December 1st 2008
Time: 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: 22. December 2008

Problem 1 Bicycle routes

Solan and Fabian live in the same house. They bike to Campus every day, but use different routes. We assume that the time (in minutes) that each of them use is Normal distributed. For Solan $X \sim N(\mu_1, \sigma^2)$ and for Fabian $Y \sim N(\mu_2, \sigma^2)$. Note that we assume a common variance.

a) Assume, for now, that $X \sim N(6, 1^2)$ and $Y \sim N(7, 1^2)$.

- What is the probability that Fabian uses more than six minutes?
- What is the probability that Solan uses less than seven minutes, given that he uses more than eight?
- One day they start at the same time. What is the probability that at least one of them reach Campus within six minutes?

b) At a party last weekend Solan made a bet on his route being faster than the one Fabian uses. The day after they decide to collect data from both routes over a weeks time, keeping notes every time they bike to Campus. Based on this they will carry out a hypothesis test with significance level $\alpha = 0.05$. Data are summarized in Table 1. Assume that the observations are independent and that $\sigma^2 = 1^2$.

x_1	x_2	x_3	x_4	x_5	x_6	x_7	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8
7.0	6.8	5.4	7.3	5.9	8.4	6.2	7.1	4.1	6.4	6.9	6.6	6.7	6.0	6.7

Table 1: *Observed times for Solan* $(x_1, x_2, \ldots, x_7)$ *and Fabian* $(y_1, y_2, \ldots, y_8)$. *We get* $\bar{x} = \frac{1}{7}\sum_{i=1}^{7} x_i = 6.31$, $\bar{y} = \frac{1}{8}\sum_{i=1}^{8} y_i = 6.81$, $\sum_{i=1}^{7}(x_i - \bar{x})^2 = 5.44$ *and* $\sum_{i=1}^{8}(y_i - \bar{y})^2 = 6.91$

Write up the hypothesis mathematically, and test it from the observed data.

What is the power of the test if the true probability distributions of X and Y are as in a).

c) Let us study closer the assumption that $\sigma^2 = 1$. Based on data in b), estimate σ^2, and construct a 95% confidence interval. Could you, from this confidence interval, indicate if the variance is significantly different from 1?

d) Because of traffic, Fabian is lead to believe that the bike trip takes longer time the later he starts in the morning. He collects more data the next week, keeping track of his starting time (t_i minutes after $7:00$) and the bike time (y_i), see Table 2.

t_i	0	15	30	45	60
y_i	5.6	5.5	6.1	7.5	7.4

Table 2: *Start time* t_i *in minutes after* $7:00$ *and bike time* y_i *in minutes. We get* $\bar{t} = 30$, $\bar{y} = 6.42$, $\sum_{i=1}^{5}(t_i - \bar{t})^2 = 2250$ *and* $\sum_{i=1}^{5}(t_i - \bar{t})y_i = 84$

Construct a simple linear regression model for the bike time. Specify your assumptions. Write down the least-squares estimator for the parameters of the model (do not show how they appear by a mathematical proof). Use, without proof, that these estimators are unbiased. Assume, in the rest of the exercise, that the regression noise terms ϵ are normal distributed with expectation 0 and known variance 0.5^2.

Formulate Fabians theory as a hypothesis test. Do the test at significance level 1%.

e) One day Fabian starts at $8:30$. Use the regression model in d) to predict Fabians bike time. Calculate a 95% prediction interval for Fabians bike time. Discuss the result.

Problem 2 Rockslide associated with road constuction

The surprising rockslide near Steinkjer caused a stop in the railway traffic and a redirection of traffic, with costs at about 1 million per day. In this exercise we will study a similar case, where the probability of rock slide associated with road construction is known.

A company is improving a road, and as part of their work they use dynamite to get rid of rock mass. There is a risk of rock slide. From basic knowledge about the geology they assume that the probability of rock slide is $p = P(X = 1) = 0.15$. In this notation with stochastic variable X, a rock slide takes place if $X = 1$, while we set $X = 0$ if no slide occurs.

a) Assume, for now, that there are 4 possible locations for rock slide along the road. They are treated as independent.

Let S be the number of rock slides out of the four possible slides. Argue why is it natural to assume that S is binomial with $n = 4$ and $p = 0.15$.

What is the probability that no rock slide occurs?

Given that there is at least one rock slide, what is the probability that there is more than one rock slide?

We will in the rest of this exercise study just one of the places where a rock slide might occur. If there is a rock slide here, the company must redirect traffic, cover damages, and so on. This has a cost of 40 million kroner. If there is no rock slide, no damage is done, and the cost is 0 kroner. The company has another option which entails a controlled redirection of traffic over the time when they use explosives. This has a fixed cost of 7 million kroner. A rock slide will not cause additional expenses when using this option.
Strategy A: Use explosives without redirecting traffic
Strategy B: Redirect traffic during the critical time.

b) Compute the expected cost Z under *Strategy A*. Compute the standard deviation of this cost. Should the traffic be redirected? Discuss.

At the extra cost of 5 million kroner the company can choose to hire geologists to do a thorough experiment with complex equipment. The experiment can tell, for sure, whether there will be a rock slide or not. In this way the company will know which one of strategy A or B to choose after the experiment has been done.

What is the expected cost if the company chooses to buy the experiment? Should the company buy the experiment? Discuss.

The geologists can, for a less expensive price of 1 million kroner, do a simpler experiment which gives a qualified guess of whether the rock mass will slide ($Y = 1$) or not ($Y = 0$).

This simpler experiment is not certain, and we assume that it has probability distribution $P(Y = 1|X = 1) = P(Y = 0|X = 0) = \gamma > 0.5$.

We first want to estimate the parameter γ. We do this based on data where we know that there was / was not a slide, i.e. we know the outcome of X. The company has used the geologists in 15 similar, independent, cases before. In seven out of those a rock slide did take place, so $X_i = 1$, $i = 1, \ldots, 7$. In these cases the geologists said: $Y_1 = 0$, $Y_2 = 1$, $Y_3 = 0$, $Y_4 = 1$, $Y_5 = 1$, $Y_6 = 1$, $Y_7 = 1$. In eight of the 15 cases no rock slide took place, so $X_i = 0$, $i = 8, \ldots, 15$. In these cases the geologists said: $Y_8 = 1$, $Y_9 = 0$, $Y_{10} = 0$, $Y_{11} = 0$, $Y_{12} = 1$, $Y_{13} = 0$, $Y_{14} = 1$, $Y_{15} = 0$.

An estimator for γ is:

$$\hat{\gamma} = \frac{\sum_{i=1}^{15} I_i}{15}$$

where $I_i = 1$ if $Y_i = X_i$, and $I_i = 0$ if $Y_i \neq X_i$.
We assume that I_i, $i = 1, \ldots, 15$ are independent.

c) Is $\hat{\gamma}$ the maximum likelihood estimator for γ? Reply to this question by computing the maximum likelihood estimator for γ.

Compute the estimate $\hat{\gamma}$ above.

d) Compute the expectation and variance of $\hat{\gamma}$.

The company is now wondering if it is worth buying the simple geologist data.

e) Use in this exercise the estimated γ from c).

Use Bayes formula to find the probability of a rock slide when the geologists say no rock slide will take place. Similarly, compute the probability of a rock slide when the geologists say a rock slide will take place.

The company wants to compute the expected cost before they ask the geologists to do the simple experiment. Why can the expected cost in this case be written like:

$$C = 1 + \sum_{y=0}^{1} \min[7, E(Z|Y = y)]P(Y = y),$$

where Z is cost using *strategy A*. Further, $\min[a, b] = a$ if $a < b$, and $\min[a, b] = b$, if $a > b$.

What is the expected cost C? Should the company buy the simple experiment from geologists? Discuss.