Norges teknisk–
naturvitenskapelige universitet
Mathematical Sciences

Contact:
Jo Eidsvik 90127472
Ola Diserud 93218823

EXAM IN TMA4240 STATISTICS

Monday 12 December 2011

Time: 09:00–13:00

Permitted material:
Yellow A5-sheet of paper with handwritten notes.
Tabeller og formler i statistikk (Tapir Forlag).
K. Rottmann: *Matematisk formelsamling.*
Calculator: HP30S.

ENGLISH

Results: 5.jan 2012.

Oppgave 1 Oil exploration

a) Let A and B be two events defined in a sample space. We have $P(A) = 0.3$, $P(B) = 0.3$ and $P(A \cup B) = 0.6$.

Are the events A and B disjoint? Are A and B independent?

The petroleum companies drilled for oil at the gigantic Aldous Major / Avaldsnes field because of other, smaller discoveries nearby. Dependence in the model for oil presence made it more likely to discover oil at Aldous Major / Avaldsnes.

Consider two oil fields. Define event A = oil at field 1, with complement A^c = no oil at field 1. Likewise, B = oil at field 2, while B^c = no oil at field 2. We know that $P(A \cap B) = 0.05$, $P(A^c \cap B) = 0.1$, $P(A \cap B^c) = 0.15$ and $P(A^c \cap B^c) = 0.7$.

b) Draw a Venn diagram of the events.

Compute the probability of oil at field 1.

Assume we discovered oil at field 2. What is now the probability of oil at field 1?

Assume that we know that field 2 does not contain oil. What is now the probability of oil at field 1?

Are the events A and B independent?

We consider a cost $K = 100$ million associated with oil exploration. This cost is paid no matter the outcome (oil or no oil). If one discovers oil, there is a reward, that exceeds the cost of exploration. Assume the profit when discovering oil at field 1 is $R_1 = 500 - K = 400$ million, while the profit when discovering oil at field 2 is $R_2 = 1100 - K = 1000$ million.

c) Compute the expected profit by oil exploration at field 1.

Assume we discovered oil at field 2. What is now the expected profit by oil exploration at field 1?

Consider the situation where we can choose between the following strategies: Not exploring for oil, exploring at field 1, or exploring at field 2. If we decide to explore at 1 or 2, we can decide to continue exploring at the other field, or to stop. This decision might depend on the outcome when exploring the first field selected. Decisions are based on expected profit.

Which exploration strategy gives the largest expected profit? And what is the associated expected profit?

Oppgave 2 Saleries

Let random (stochastic) variable Y be a persons annual salary in 2011 in kNOK (1 kNOK is 1000 kroners). Paretos law claims that

$$P(Y \geq y) = \left(\frac{k}{y}\right)^{\theta},$$

where $\theta > 2$, $y \geq k$, and k is the minimum salary in the population.

a) Assume Paretos law holds. Show that the probability density of Y is

$$f(y) = \theta k^{\theta} \left(\frac{1}{y}\right)^{\theta+1}, \quad y \geq k, \;\; \theta > 2.$$

Compute the expected value and variance of Y.

Further assume that $k = 214.9$ kNOK, i.e. the first annual salary level according to the Norwegian regulations in 2011. From a random sample of 30 annual salaries we get $\sum_{i=1}^{30} y_i = 13611$ og $\sum_{i=1}^{30} \ln y_i = 174.7$.

b) Show that the maximum likelihood estimator for θ, from a random population of size n, becomes

$$\hat{\theta} = \frac{n}{\sum_{i=1}^{n} \ln y_i - n \ln k}.$$

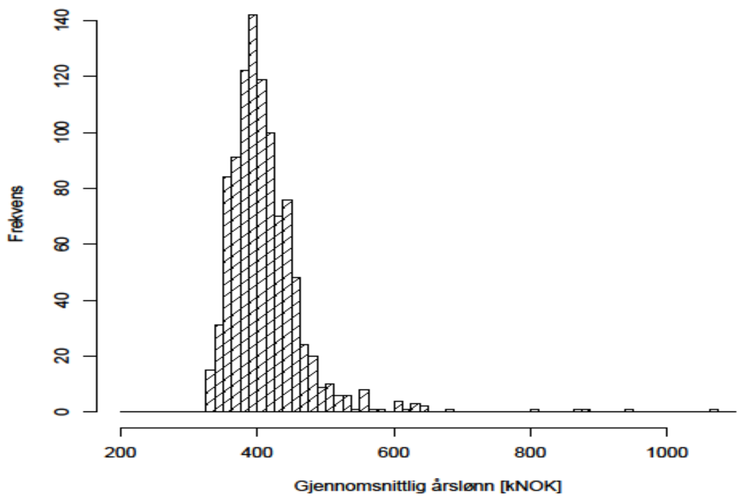
What is the probability that a random person has an annual salary below 472.2 kNOK, which is the level for top-tax in Norway, when we use the point estimate of θ computed from the data above.

c) In 2010 the average annual salary of all employed Norwegians was 435 kNOK, with a standard deviation of 516 kNOK. The average had then increased by 3.6 % compared with 2009. We will now test if the expected annual salary for 2011 is also 3.6 % above the year before (2010), or if the financial crisis has slowed the salary development. We assume the standard deviation is known, and set to the empirical level of 2010.

Formulate the hypotheses involved. Which assumptions do we have to make to be able to test the hypotheses statistically?

d) Do the test with significance level 0.05. What is the conclusion of the test?

We now draw 1000 random samples, of size 30 each, from the probability density model in a), with $k = 214.9$ and θ set to the point estimated from b). Figure 1 shows a histogram of the mean of the 1000 samples. Does it look like $n = 30$ is sufficient to apply the central limit theorem? Please discuss.



Oppgave 3 Mining

When evaluating mineral resources, one collects different data to assess the potential of development. Here, we look at core samples Y_i = the amount of a mineral per unit volume, $i = 1, \dots, n$. In addition, we know the rock category x_i, $i = 1, \dots, n$. The goal is to fit a linear regression model for the response Y_i, where rock type x_i is used as a covariate. Statistically:

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \qquad i = 1, \dots, n,$$

where ϵ_i, $i = 1, \dots, n$, are independent Gaussian random variables, with expectation 0 and assumed known variance $\sigma^2 = 0.5^2$.

By using the method of least squares, we get estimators for β_0 og β_1. They are:

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}, \quad \hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x}) Y_i}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ and $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$.

a) Show that $\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$.

And show that $\hat{\beta}_1$ is an unbiased estimator for β_1.

A mining company collects $n = 7$ core samples, with responses $y_1, \dots, y_7$, measured for different rock types x_i, $i = 1, \dots, 7$. The result is summarized as follows:

Rock type (x)	1	1	2	2	2	3	3
Core samples (y)	2.7	3.1	3.9	3.2	4.2	4.1	5.5

b) Define a 90 % confidence interval for β_1. Find the actual result of the interval given the numbers above.

The company discusses whether the confidence interval could have been shorter if they instead collected three core samples in both rock types 1 and 3, and only one in rock type 2? Or what if they took one sample in both of rock types 1 and 3, and five samples in type 2? Please discuss.

c) The mining company is going to take a new core sample, where the rock type is $x_0 = 3$.

Construct a 90 % prediction interval for this new core sample, and compute the actual values given the data above.