Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

Contacts during exam:
Mette Langaas 98847649
Ole Petter Lødøen 99484158

EXAM IN COURSE TMA4240 STATISTICS

Thursday December 9th 2004

Time: 09:00–13:00

Permitted aids:
Yellow A5 sheet with your own handwritten notes.
Tabeller og formler i statistikk (Tapir Forlag).
K. Rottmann: *Matematisk formelsamling.*
Calculator: HP30S.

ENGLISH

The examination results are due January 6th 2005.

Problem 1 The Bronze Bolts

Bronze is an alloy where copper and tin are the main components. We study the content of copper in bronze bolts of a given dimension, which is made of a special type of bronze alloy.

The company *Bronsespesialisten* produces bronze bolts, and a sample of $n = 10$ bronze bolts is taken from the production. The copper content, X_i,$i = 1, ..., n$, is measured in each bronze bolt in the sample. We assume that $X_1, X_2, ..., X_n$ are independent and normally distributed random variables with $\mathrm{E}(X_i) = \mu_x$ and $\mathrm{Var}(X_i) = \sigma^2$.

a) We assume that the mean is $\mu_x = 85$ grams and the variance is $\sigma^2 = 1$ grams^2.

What is the probability that the copper content in a randomly chosen bronze bolt is less than 84 grams?

Find a number, k, such that the probability is 0.01 for the copper content in a randomly chosen bronze bolt to be larger than k.

We consider the copper content in two randomly chosen, independent bronze bolts. What is the probability that the copper content in the two bronze bolts deviate more than 1.5 grams?

We assume in the rest of this problem that both μ_x and σ^2 are unknown parameters. First, we look at estimators for σ^2.

b) What properties should a good estimator have?

To possible estimators for σ^2 are $\widehat{\sigma}^2$ and S^2

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Find the mean and the variance of the two estimators, and comment on the results. (Clue: use relation to chi-squared distribution).

The company *Metalleksperten* also produces bronze bolts of the same alloy and dimension as Bronsespesialisten does. At Metalleksperten a sample of $m = 10$ bronze bolts is taken, and the copper content Y_j, $j = 1, ..., m$, is measured in each bolt in the sample. Let $Y_1, Y_2, ..., Y_m$ be independent and normally distributed with $\mathrm{E}(Y_j) = \mu_y$ and $\mathrm{Var}(Y_j) = \sigma^2$, where both μ_y and σ^2 are unknown parameters. We assume that measurements taken of the copper content in bronze bolts at the two different companies are independent. Further, we assume that the measurements have the same (but unknown) variance, i.e. $\mathrm{Var}(X_i) = \mathrm{Var}(Y_j) = \sigma^2$.

The actual measurements in the two samples taken from Bronsespesialisten and Metalleksperten are listed in Table 1.

Bronsespesialisten	84.44	84.77	86.56	85.07	85.13	86.72	85.46	83.73	84.31	84.55
Metalleksperten	85.22	84.36	84.40	84.11	83.44	85.79	84.50	82.03	84.70	83.53

Table 1: Copper content in bronze alloy in samples from Bronsespesialisten and Metalleksperten. It is given that for Bronsespesialisten $\sum_{i=1}^{10} x_i = 850.75$ and $\sum_{i=1}^{10}(x_i - \bar{x})^2 = 8.19$, and for Metalleksperten $\sum_{j=1}^{10} y_j = 842.10$ and $\sum_{j=1}^{10}(y_j - \bar{y})^2 = 9.70$.

Lately, the management in Bronsespesialisten has repeatedly accused Metalleksperten for using a lower copper content in the bronze alloy than Bronsespesialisten does. We want to check if this is true.

c) Phrase this as a hypothesis test by defining the null hypothesis and the alternative hypothesis.

Give a test observator and find the critical region, where we reject the hypothesis. What will your decision be if you use the data in Table 1, given that the level of significance is $\alpha = 0.05$?

Calculate the p-value, by using Table 2.

t	1.82	1.86	1.90	1.94	1.98	2.02	2.06
$\nu = 18$	0.957	0.960	0.963	0.966	0.968	0.971	0.973
$\nu = 19$	0.958	0.961	0.964	0.966	0.969	0.971	0.973
$\nu = 20$	0.958	0.961	0.964	0.967	0.969	0.972	0.974

Table 2: Cumulative distribution in the t-distribution. For a t-distributed statistic T with ν degrees of freedom, the table shows $P(T \leq t)$ for different values of t.

Metalleksperten has been vandalized, and the window at the main office was broken by a bronze bolt thrown at it. The police is investigating the case, and want to know if bronze bolt that broke the window was produced by Metalleksperten or by Bronsespesialisten.

d) Find an interval that with 95% probability will contain a new observation of the copper content in a bronze bolt from Bronsespesialisten. Find numerical values for the interval by using the data in Table 1. Give the corresponding interval for Metalleksperten.

The bronze bolt which was used to break the window at Metalleksperten was measured to have a copper content of 86.30 gram. Can you, based on the intervals you made above, say anything about which company that produced the bronze bolt?

Problem 2 The Pyramid Scheme

Ole Petter has been asked to join a pyramid scheme. All he has to do is to pay a certain amount of money, and then recruit five new persons, which will be placed in the level below him in the pyramid. By doing so, he is assured that he will make heaps of money. According to the person who wanted to recruit Ole Petter, a person who is asked to join the pyramid scheme will join with probability $p = 1/3$, so to get five new persons to join should not be a problem.

Before Ole Petter decides on what to do, he will perform some calculations on probabilities and expected values, to find out how much work the recruiting process will be. He decides to consider the random variable X, which describes the number of persons he has to ask until the *first person* agrees to join the pyramid scheme.

a) What are the assumptions for X to have a geometric distribution? In the rest of this problem, you can assume that X is geometrically distributed with probability distribution

$$f(x) = p(1-p)^{(x-1)} \quad \text{for } x = 1, 2, ...$$

If Ole Petter decides to join the pyramid scheme, what is the expected number of persons he has to ask to get one new person to join the pyramid scheme, when $p = 1/3$?

What is the probability that he has to ask more than five persons to get one person to join the pyramid scheme, when $p = 1/3$?

Ole Petter is a bit sceptical to the information that every third person that is asked will join the pyramid scheme. Therefore, he decides to estimate the probability p that a person that is asked wants to participate. He decides to collect data by counting the number of people he has to ask until the first one agrees to join. He will try to get some friends to help him gather observations. The ith friend observes that he has to ask X_i persons. Through n surveys, Ole Petter will get n independent observations, $X_1, X_2, \ldots, X_n$ of the random variable X.

b) Find the maximum likelihood estimator $\hat{p}$ for p based on the observations $X_1, X_2, \ldots, X_n$.

Ole Petter could not get any of his friends to help him with the survey, so he is left with only the observation he collected himself, X_1. Is the maximum likelihood estimator based on only one observation unbiased?

$$\text{Clue}: \sum_{n=1}^{\infty} \frac{1}{n} a^n = -\ln(1-a) \; ; \; 0 < a < 1$$

Problem 3 Test the nation

Saturday November 27th 2004, the TV show "Test the nation" was aired on NRK1. In this show, 270 participants in the studio were asked questions on different topics. Based on their age and the number of correct answers, they were each given an IQ score.

In the TV show it was said that the test was made such that the IQ score for a randomly chosen person should be normally distributed with mean equal to 100 and standard deviation equal to 15.

The highest registered IQ score among the participants in the TV studio was 122. Two people got an IQ score of 122.

a) What is the probability that a randomly chosen person will get an IQ score of at least 122?

If we test a representative sample of size 270 persons, what is the expected number of persons that will get an IQ score of at least 122?

What is the probability that the maximum IQ score in a random sample of size 270 persons will be at least 122?

The participants in the TV studio were divided into six groups. One of the groups consisted of 42 former "Reality-show-participants". Regarding IQ score we decided, prior to the TV show, to regard these 42 participants as a representative sample of the total population. The Reality-show-participants got an average IQ score of 94.

b) We assume that the IQ score of a randomly chosen person is normally distributed with unknown mean μ, but with know standard deviation equal to 15. Find a 95% confidence interval for the mean IQ score based on the data from the Reality-show-participants.

After the TV show, claims have been made that the questions in the test were too difficult, and thus the registered IQ scores were lower than expected. If we assume that the IQ score of a randomly chosen person is normally distributed with mean equal to 100 and standard deviation equal to 15, what is the probability for getting an average IQ score less than or equal to 94 in a random sample of size 42?