Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4240/TMA4245 Statistics**

Academic contact during examination: Jo Eidsvik

Phone: 90 12 74 72

Examination date: August 15th 2016

Examination time (from–to): 09.00-13.00

Permitted examination support material: *Tabeller og formler i statistikk*, Tapir Forlag, K. Rottmann: *Matematisk formelsamling*, Calculator Casio fx-82ES PLUS, CITIZEN SR-270X, CITIZEN SR-270X College or HP30S, A yellow stamped A5-sheet with your own handwritten notes.

Other information:
Provide justification for all answers and include intermediate calculations.

Language: English

Number of pages: 4

Number of pages enclosed: 0

Checked by:

Date Signature

Problem 1 Dice game

Assume a traditional dice is thrown. Let X_i be a stochastic variable taking the result of dice throw number i, where the possible outcomes are $x_i \in \{1, \ldots, 6\}$, and $P(X_i = l) = \frac{1}{6}$, $l = 1, \ldots, 6$. This implies $E(X_i) = 3.5$ and $\text{Var}(X_i) = 1.71^2$.

a) We consider the sum of 2 independent throws of a dice: $Y_2 = X_1 + X_2$.

With what probability is the sum exactly equal to 12?

With what probability is the sum larger or equal to 10?

Find the mean value of the sum.

In a game the dice is thrown repeatedly, and the goal is to obtain a high score. Let Z_k denote the score after k throws. Before throw number $k+1$ one must choose whether one wants to throw the dice once more, or whether the game should be terminated. If the game is terminated one keeps the current score Z_k. The game starts with score $Z_0 = 0$ at $k = 0$. If the outcome of throw $k+1$ is $x_{k+1} \in \{1, \ldots, 5\}$, this outcome is added to the score ($Z_{k+1} = Z_k + X_{k+1}$). If the outcome becomes $x_{k+1} = 6$, the score is set to zero, $Z_{k+1} = 0$, and the game is terminated.

b) Find the expected score after the first throw.

If a player has a score of 30 and chooses to continue to throw the dice, what is the expected score after the next throw?

A player uses optimal expected value to decide whether to *i*) continue to throw the dice or *ii*) to stop at the current score. At what scores is it optimal to continue to throw the dice?

Problem 2 The price of an hotel room

Harald considers a hotel vacation in Barcelona in the late summer. Assume the price of a hotel room can be described by a normal distribution with mean value μ Euros per night and standard deviation 20 Euros. Assume an exchange rate of 9.2 kroner per Euro.

a) Assume in this item that $\mu = 100$.

What is the distribution for the price in kroner for a hotel room per night?

With what probability does a hotel room cost more that 1000 kroner?

With what probability does a hotel room cost more than 1100 kroner, given that it costs more than 1000 kroner?

Harald thinks the expected price is 100 Euros, whereas a friend of him thinks it has become more expensive. They agree that the standard deviation is known and equal to 20 Euros. Harald collects data to study whether the expected price has become higher than 100 Euros. He makes a number of calls and collects price data, $x_1, \ldots, x_{20}$, for 20 hotel rooms. He gets $\bar{x} = 120$ Euros.

b) Formulate the problem as an hypothesis test.

Use the normal assumption and the given numbers to perform the hypothesis test with significance level 0.05.

c) Now assume that the true mean value equals 110 Euros.

Find the power of the test.

Find how many observations Harald must collect for the power to be 0.95?

Problem 3 Concentration of medicine

In this problem we are going to consider a disease for which the treatment is to inject a medicine into the blood. Let x denote the dose of medicine a patient gets injected. We will assume that this dose can be controlled and therefore we do not consider x to be stochastic. Twenty-four hours after the medicine is injected, the concentration, Y, of the medicine in the blood is measured. We assume the following linear regression model for the relation between x and Y,

$$Y = bx + \varepsilon,$$

where b is an unknown parameter and ε has a normal distribution with zero mean and variance σ^2. In all parts of this problem we assume $\sigma^2 = 2.0^2$ to be known.

Assume we have observed values for $n = 10$ patients and let x_i and Y_i for $i = 1, 2, \ldots, n$ be corresponding values of injected dose and measured concentration in the blood. We assume $Y_1, Y_2, \ldots, Y_n$ to be independent stochastic variables and we assume the regression model above to hold for each of them. Observed values for the n patients are:

patient i	1	2	3	4	5	6	7	8	9	10
x_i	2.0	5.0	2.5	2.5	4.5	5.5	3.0	6.0	3.0	2.5
y_i	2.9	11.5	5.1	7.9	9.7	11.0	8.6	13.8	5.7	2.4

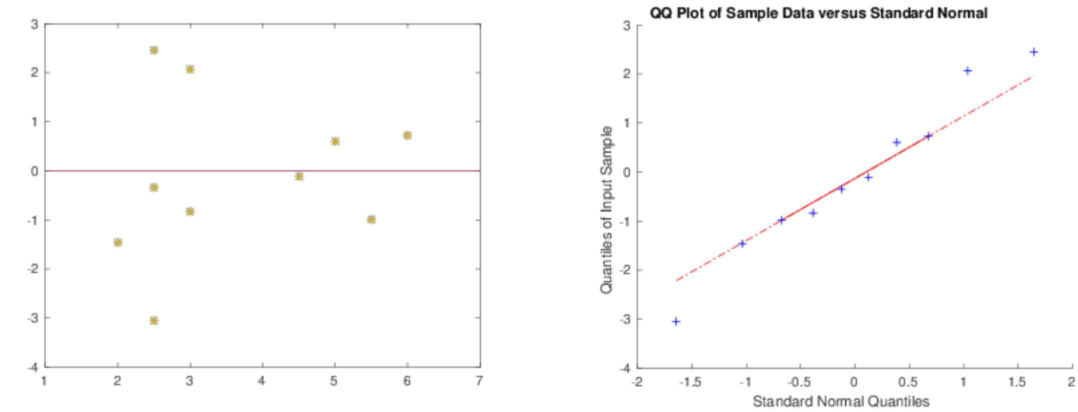

Figure 1: Left: Residual plot. The dose x_i is plotted along the x-axis and estimated value for the residual ε_i is plotted along the y-axis. Right: Normal quantile-quantile plot (QQ-plot) for the estimated residuals.

It is given that $\sum_{i=1}^n x_i = 36.5$, $\sum_{i=1}^n x_i^2 = 152.25$ and $\sum_{i=1}^n x_i y_i = 331.65$.

For the model and the data given above Figure 1 shows the residual plot and a normal quantile-quantile plot (QQ-plot) for the estimated residuals.

a) Describe briefly how a normal quantile-quantile plot (QQ-plot) can be used and how the plot is to be interpreted.

Based on the plots in Figure 1, discuss whether the observed values appear to fit the assumed model.

To estimate the parameter b, the estimators

$$\widetilde{b} = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i} \quad , \quad \widehat{b} = \frac{1}{n}\sum_{i=1}^n Y_i \quad \text{and} \quad \widehat{\widehat{b}} = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}$$

are proposed. It is given that $\widehat{\widehat{b}}$ is unbiased and that $\text{Var}\left[\widehat{\widehat{b}}\right] = \sigma^2 / \sum_{i=1}^n x_i^2$.

b) Which of the three estimators do you prefer when $n = 10$ and the observations are as given above? Give reason for your answer.

c) Write down the likelihood function for b for the situation described above. Use the likelihood function to derive the maximum likelihood estimator (MLE) for b.

Regardless of your answers in parts **b**) and **c**), you should in the rest of this problem base your answers on the estimator $\widehat{\widehat{b}}$ given above.

d) Explain why $\widehat{\widehat{b}}$ has a normal distribution.

Derive a $(1-\alpha)\cdot 100\%$ confidence interval for b expressed by n, $x_1, x_2, \ldots, x_n$, $Y_1, Y_2, \ldots, Y_n$, σ^2 and α.

Find the interval numerically when the observations are as given above and $\alpha = 0.10$.

It is important that the concentration of medicine in the blood is not too high, as this may give serious adverse effects. After having observed the results for the first $n = 10$ patients (given above), the medical doctors receive a new patient and after having examined this patient the medical doctors decide that it is important that the measured concentration of medicine in the blood of this patient does not exceed 10.0.

e) Find the highest dose x_0 this new patient can get injected if one requires a probability of at least 0.95 for the event that the measured concentration of medicine after twenty-four hours does not exceed 10.0.