NTNU – Trondheim
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for TMA4240 Statistics

Academic contact during examination: John Tyssedal, Haakon Bakka.

Phone: John Tyssedal: 41645376. Phone Haakon Bakka: 97955667.

Examination date: 17.12.2014

Examination time (from-to): 09:00-13:00

Permitted examination support material: C

- Tabeller og former I statistikk, Tapir forlag
- K. Rottman: Matematisk formelsamling
- Calculator: Casio fx-82ES PLUS, CITIZEN SR-270X, CITIZEN SR-270X College

  Eller HP30S.
- A stamped yellow A5 piece of paper with your own handwritten formulas and notes.

Other information:

One should give reasons for all answers and all natural calculations should be provided .

Language: English

Number of pages (front page excluded): 4

Number of pages enclosed: 0

Checked by:

Date Signature

Problem 1

During a season of Norwegian premier league football each team plays 30 matches out of which 15 are home matches and 15 are away matches. At the beginning of the season, one of the teams calculates their probabilities of a win (A), a draw (B) and defeat (C) for home matches to P(A) = 0.6, P(B) = 0.2 og P(C) = 0.2. For away matches, the probabilities are P(A) = 0.4, P(B) = 0.2 og P(C) = 0.4. Further assume that the outcomes of all matches are independent.

a) A team is given 3 points for a win, 1 point for a draw and 0 points for defeat. Let X be the number of points the team is given in a home match and let Y be the points the team is given in an away match. Show that the mean (expected value) and variance of X is given by 2 and 1.6 and that the mean and variance of Y is given by 1.4 and 1.84.

Let $X_i,\ i = 1,2,\cdots,15$ be the number of points the team achieves in each of the 15 home matches and let $Y_i,\ i = 1,2,\cdots,15$ be the points the team achieves in each of the 15 away matches. Also let $X_H = \sum_{i=1}^{15} X_i$ and $Y_B = \sum_{i=1}^{15} Y_i$. The team assumes that they will get a medal (first, second or third position) if they achieve at least 60 points in total.

b) Find the mean and variance of X_H and Y_B. Assume that X_H and Y_B can be approximated by normal distributions and use this to find an approximate probability that they get enough points for a medal (at least 60 points).

After a mediocre first part of the season, when 8 matches remain (4 home matches and 4 away matches), the team concludes that they must win at least 7 out of the 8 remaining matches to avoid being relegated (transferred) to the division below. They decide to evaluate the probability that they will avoid relegation using the same probabilities given previously and to fire their coach if this probability is smaller than 0.05.

c) Let N_H and N_B be the number of home and away wins during the last 8 remaining matches. Which probability distribution will N_H and N_B follow? Will the coach be fired?

Problem 2

Les McBurney has always been fascinated by volcanos and volcanic eruptions. After high school he decides to go to Iceland to pursue a 5-year master's degree. Let X_t be the number of volcanic eruptions on Iceland during a time interval of length t. Based on historical data, it is know that X_t is Poisson distributed with parameter λt, where $\lambda = 0.3$ and *t* is the length of the time interval in years. This implies that $P(X_t = x) = \frac{(0.3t)^x e^{-0.3t}}{x!},\ x = 0,1,2,\cdots$.

a) Find the probability that Les will experience at least one eruption during his 5-year stay on Iceland.
Suppose that an eruption occured exactly one year after Les' arrival on Iceland. Given this, find the probability that he will experience at least one additional eruption during his 5-year stay.

b) Let T be the time that passes from Les' arrival until the occurence of the second eruption. Which distribution will T follow? Explain why $P(T \le t) = 1 - P(X_t < 2)$.
Use this to find the time *t* required to make the probability of experiencing at least two eruptions greater than 0.80.

Problem 3

Two manufacturers of electric cars compete within the same market with their respective models A and B. For customers deciding which model to buy, the driving range before the battery must be recharged is an important criteria. The manufacturer of model B accuses the manufacturer of model A of reporting a too long driving range not achieveable under normal driving conditions.To investigate this the manufacturer of model A decides to conduct an experiment. A number of *n*=10 drivers are given the task to drive a given course representing standard conditions at an average speed of 60 km/h until the battery runs empty. Let X_i be the driving range obtained by each driver i, $i = 1,2,\ldots,10$. The observed driving ranges in km are given below:

x_i: 123.0, 119.3, 119.4, 118.2, 119.5, 118.5, 117.7, 118.1, 119.6, 119.4

Let $E(X_i) = \mu_A$ and $Var(X_i) = \sigma^2$, $i = 1,2,\ldots,10$.

a) Write down unbiased estimators of μ_A and σ^2 given the sample. Compute the estimates when $\sum_{i=1}^{10} x_i = 1192.70$ and $\sum_{i=1}^{10} (x_i - \bar{x})^2 = 19.68$.

The manufacturer of model A claims that the expected driving range under standard conditions is 125 km. Assume that $X_i \sim N(\mu_A, \sigma^2)$ and independent, $i = 1,2,\ldots,10$.

b) Can you, based on these data, conclude that the manufacturer has reported a too long driving range? Formulate this as a hypothesis testing problem and carry out the test. Use $\alpha = 0.01$ as your level of significance.

In an advertisement, it is claimed that model A has a longer driving range than model B. The manufacturer of model B decides to carry out the same experiment as above for model B. Let Y_i be the obtained driving ranges for each driver i, $i = 1,2,\ldots,10$.

The observed driving ranges in km are given below:

y_i : 109.9, 110.4, 109.7, 111.5, 112.3, 111.8, 112.1, 108.2, 109.9, 116.1

$\sum_{i=1}^{10} y_i = 1111.9$ and $\sum_{i=1}^{10} (y_i - \bar{y})^2 = 41.75$

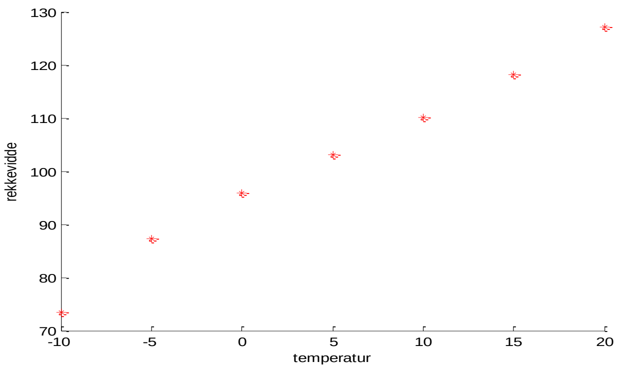
Assume that $Y_i \sim N(\mu_B, \sigma^2)$. The variances are thus equal in both samples but the expected values may be different. Assume that $X_1, \ldots, X_{10}, Y_1, \ldots, Y_{10}$ are all independent.

c) Write down an unbiased estimator of σ^2 based on the two samples. Construct a 95%-confidence interval for $\mu_A - \mu_B$ based on the observed data. Does the interval suggest that the claim made in the advertisment can be correct? Give an argument for you conclusion based on the interpretation of the confidence interval.

Several consumer associations complain that manufacturers of electric vehicles provide too little information about how dependent the driving range is on temperature. The manufacturer of model A therefore conduct an additional experiment in which the driving range is measured for 7 different temperatures between -10^0 C and 20^0 C. The temperatures and the driving ranges obtained are given in the Table below.

Temperature(^{0} C)	-10	-5	0	5	10	15	20
Driving range (km)	73.4	87.2	95.9	103.1	110.1	118.1	127.1

A plot of driving range versus temperature is given below.



Let $t_i, i = 1,2,\ldots,7$ be the 7 temperatures used in the experiment and let R_i, $i = 1,2,\ldots,7$ be the corresponding driving ranges. Based on the plot, the relationship appears to be approximately linear and it is decided to fit a model of the form $R_i = \beta_0 + \beta_1 t_i + \varepsilon_i, i = 1,2,\ldots,7$.

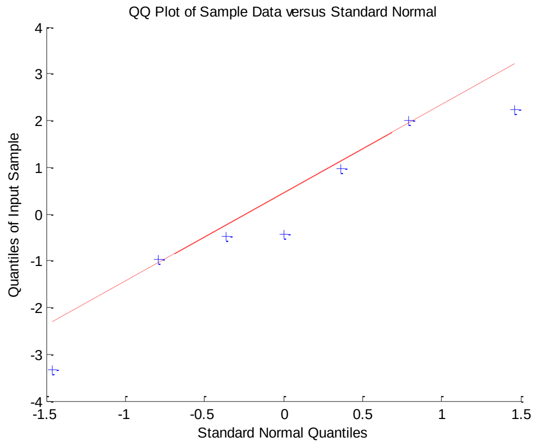
where the error terms are independent and satisfy $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$. The least squares estimator of

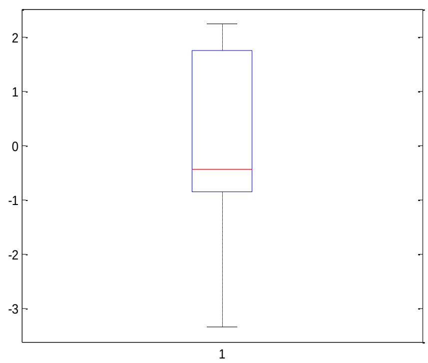
β_1 is then given by $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{7}(t_i - \bar{t})R_i}{\sum_{i=1}^{7}(t_i - \bar{t})^2}$.

d) Show that the variance of $\hat{\beta}_1$ is given by $\dfrac{\sigma_\varepsilon^2}{\sum_{i=1}^{7}(t_i - \bar{t})^2}$. The estimates of β_0, β_1 og σ_ε^2 equals 93.7, 1.7 and 4.5 respectively. Use this to construct a 95 % confidence interval for β_1.

The deviations between the observed driving ranges R_i and the corresponding values predicted by the regression model $93.7 + 1.7 t_i$ are called the residuals. The distribution of the observed residuals can be used to assess the assumption that $\varepsilon_i \sim N(0,\sigma_\varepsilon^2)$.

A QQ-plot and a box-plot of the residuals are given below.





e) Comment on the plots. In what follows, assume that $\sigma^2 = \sigma_\varepsilon^2 = 4$ and that all R_i, $i = 1,2,\ldots,7$ are independent of $X_1,\ldots,X_{10},Y_1,\ldots,Y_{10}$. It turns out that the 10 observations given previously for model A were made under temperatures equal to 15^0 C and the 10 observations for model B under temperatures equal to 10^0C. It is thus more reasonable to compare $\mu_A - 5\beta_1$ with μ_B. Use $\bar{X} - 5\hat{\beta}_1$ as an estimator of $\mu_A - 5\beta_1$ and assess if there is any evidence to conclude that the expected driving ranges for the two models A and B differ at temperatures equal to 10^0C. Answer the question by carrying out a hypothesis test. Choose your own level of significance.