NTNU
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4240/45 Statistics**

Academic contact during examination: Øyvind Bakke
Phone: 990 41 673

Examination date:

Examination time (from–to):

Permitted examination support material: Hjelpemiddelkode C:

– Tabeller og formler i statistikk, Akademika,
– A yellow sheet of paper (A5 with a stamp) with personal handwritten formulas and notes,
– A specific basic calculator

Other information:

All your answers should be justified.

The hand-in material should contain calculations leading to your answer.

There are 10 subtasks which have equal weights in the grading.

Language: English

Number of pages: 4

Number of pages enclosed: 0

Checked by:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
sort/hvit ☒	**farger** ☐
skal ha flervalgskjema ☐	

Date Signature

Problem 1 Probability

Let X be a random variable having probability mass function $p(x) = P(X = x)$ as shown in the table:

x	-1	0	1
$p(x)$	0.3	0.6	k

a) Find k such that $p(x)$ is a valid probability function.

Find the cumulative probability function of X, $F(x)$.

Find $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

Problem 2 Lifetime

Assume that the lifetime T, measured in hours (need not be an integer), of a randomly chosen electronic component is exponentially distributed with probability density

$$f(t;\beta) = \begin{cases} \frac{1}{\beta}e^{-t/\beta} & t \geq 0 \\ 0 & t < 0, \end{cases} \tag{1}$$

where $\beta > 0$ is a parameter.

a) Assume (only in this subtask) that $\beta = 30$.

Show that the cumulative distribution function of T is given as

$$P(T \leq t) = \begin{cases} 1 - e^{-t/30} & t \geq 0 \\ 0 & t < 0. \end{cases}$$

Find $P(T < 20)$ and $P(T < 20 \cup T > 40)$.

Find the median of T, that is, k such that $P(T \leq k) = 0.5$.

b) Show that the exponential distribution is memoryless, that is,

$$P(T \geq t + s \mid T \geq t) = P(T \geq s)$$

for $s \geq 0$.

Show that $\frac{2}{\beta}T$ has a chi-squared distribution with 2 degrees of freedom. (Hint: you can use that $\Gamma(1) = 1$ in the formula, found in the formula booklet, for the probability density of a chi-squared variable.)

Let T_1, T_2, $\ldots$, T_n be independent observations of the lifetime of n electronic components.

c) Argue briefly why $\frac{2}{\beta}\sum_{i=1}^{n} T_i$ has a chi-squared distribution with $2n$ degrees of freedom. (Hint: you can use the results from subtask **b)** without proof.)

Use this to derive a $(1-\alpha)100\%$ confidence interval for β.

What is the 95% confidence interval if $n = 20$ and $\sum_{i=1}^{20} t_i = 30$?

Problem 3 Sleeplessness

Assume that the time (in minutes), X, it takes for a patient to fall asleep after taking sleeping medicine against sleeplessness is normally distributed with mean (expected value) μ and standard deviation σ.

a) Assume, only in this subtask, that $\mu = 35$ and $\sigma = 10$.

Find the probability that the patient falls asleep within 20 minutes after having taken the sleeping medicine.

Given that the patient has not fallen asleep within the first 20 minutes, find the probability that the patient has not fallen asleep within the first 40 minutes.

Assume that two patients take the sleeping medicine independently of each other on a random day. Find the probability that the total time it takes before they fall asleep is over 90 minutes.

Two pharmaceutical companies competes to be the market leader within sleeping medicine, and both claim that their medicine is better than that of the competitor. Assume that both companies give their sleeping medicine to n unique patients and measure the time it takes before each patient falls asleep.

Assume that the times $X_1, X_2, \ldots, X_n$ it takes for the patients of company A to fall asleep is a random sample with unknown mean (expected value) μ and known standard deviation $\sigma = 10$. Further, assume that the corresponding times $Y_1, Y_2, \ldots, Y_n$ for company B are independent and normally distributed with unknown mean θ and known standard deviation $\tau = 12$. You can assume that the two samples are independent of each other.

Company B wants to perform the hypothesis test

$$\mathrm{H}_0\colon \mu = \theta \qquad \text{against} \qquad \mathrm{H}_1\colon \mu > \theta. \tag{2}$$

with a significance level $\alpha = 0.1$ based on the random samples $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$.

b) Assume that the true difference between the effects of the two companies is $\mu - \theta = 5$ minutes.

Derive an expression for the least number of patients n both companies must give their sleeping medicine to if they require that the test power shall be at least 95 % if the true difference is $\mu - \theta = 5$ minutes.

Problem 4 Table-tennis ball

A manufacturer produces table-tennis balls for professional use. One of the requirements for a table-tennis ball that is to be used at a competition is that the diameter is 40 millimetres. Assume that the diameter, in millimetres, of a table-tennis ball is normally distributed with unknown mean (expected value) μ and known standard deviation $\sigma = 0.5$.

In the latest competitions, several table-tennis players have claimed that the ball has had a wrong diameter. The manufacturer therefore wants to perform a hypothesis test to investigate whether the claim of the players is correct. The manufacturer performs the hypothesis test

$$\mathrm{H}_0\colon \mu = 40 \text{ millimetres} \qquad \text{against} \qquad \mathrm{H}_1\colon \mu \neq 40 \text{ millimetres} \tag{3}$$

at a significance level $\alpha = 0.1$ based on a random sample $X_1, X_2, \ldots, X_{10}$. It is given that $\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 40.2$.

a) Perform the above hypothesis test by using a p-value. Will the manufacturer reject the null hypothesis?

Derive an expression for a 95% prediction interval for the diameter X_0 of a new table-tennis ball, independent of $X_1, X_2, \ldots, X_{10}$.

Problem 5 Linear regression

Assume the following simple linear regression model,

$$Y = \beta_0 + \beta_1 x + \varepsilon, \tag{4}$$

where β_0 and β_1 are unknown parameters, x is a known explanatory (independent) variable and ε is assumed normally distributed with mean (expected value) 0 and known variance σ^2.

Assume that we from the model defined in (4) have a random sample (x_1, Y_1), (x_2, Y_2), $\ldots$, (x_n, Y_n).

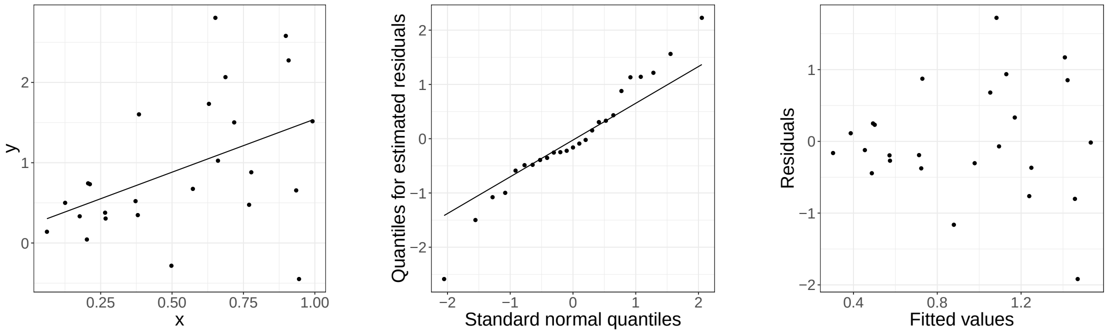
a) Assume (only in this subtask) that $\beta_0 = 0$.

Show that the likelihood maximum estimator of β_1 based on the random sample $(x_1, Y_1), (x_2, Y_2), \ldots, (x_n, Y_n)$ then is

$$\widetilde{\beta}_1 = \frac{\sum_{i=1}^n x_i Y_i}{\sum_{i=1}^n x_i^2}.$$

Find the expected value and variance of $\widetilde{\beta}_1$.

In the figure the regression model defined in (4) fitted to a random sample (x_1, Y_1), $(x_2, Y_2), \ldots, (x_{25}, Y_{25})$ is shown together with a normal probability plot (Q–Q plot) of the residuals, and the fitted observations plotted against the residuals. It is given that $\widehat{\beta}_0 = 0.22$ and $\widehat{\beta}_1 = 1.32$.



b) Find the estimated expected value of Y when it is given that $x = 0.25$.

Based on the figure, briefly discuss whether the fitted model is reasonable for the observations. In particular, give the assumptions that must be satisfied when using a simple linear regression model.

Problem 6

Let X_1, X_2 and X_3 be independent Poisson distributed random variables with expected values μ_1, μ_2 and μ_3, respectively.

Let

$$Z_i = \begin{cases} 1 & \text{if } X_i = X_3 = 0, \\ 0 & \text{otherwise,} \end{cases} \tag{5}$$

for $i = 1, 2$.

a) Show that Z_i is Bernoulli distributed, that is, binomially distributed with $n = 1$ trials, with success probability $p_i = e^{-(\mu_i + \mu_3)}$, and use this to find the expected value and variance of Z_i for $i = 1, 2$.

Use this to derive an expression for the covariance between Z_1 and Z_2, that is, $\text{Cov}(Z_1, Z_2)$.