Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **ST1201/ST6201 Statistical methods**

Academic contact during examination: Nikolai Ushakov
Phone: 45128897

Examination date: 20 December 2019

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

Permitted examination support material: C:
– Tabeller og formler i statistikk, Tapir forlag,
– K.Rottman. Matematisk formelsamling,
– Stamped yellow A4 sheet with your own handwritten notes,
– Calculator: HP30S, Citizen SR-270X, Citizen SR-270X College or Casio fx-82ES PLUS.

Other information:
Sensur:

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

Let $X \sim \mathrm{N}(\mu, 2\sigma^2)$ and $Y \sim \mathrm{N}(2\mu, \sigma^2)$ and assume X and Y to be independent. We want to estimate μ from observed values for X and Y.

a) The following two estimators have been proposed,

$$\hat{\mu}_1 = \frac{1}{3}X + \frac{1}{3}Y, \quad \hat{\mu}_2 = \frac{1}{2}X + \frac{1}{4}Y.$$

Which of the two estimators would you prefer? Give reason for the answer!

Problem 2

Let $X_1, X_2, ..., X_n$ be independent and identically distributed, $X_i \sim \mathrm{N}(0, \sigma^2)$, where σ^2 is unknown. We test the hypotheses

$$H_0 : \sigma^2 = 1 \quad \text{versus} \quad H_1 : \sigma^2 \neq 1.$$

a) Use

$$T = \frac{1}{n} \sum_{i=1}^{n} X_i^2$$

as the test statistic, work out a decision rule for when to reject H_0. Use significance level α.

Find the power function of the test.

b) Find the generalised likelihood ratio (GLR) λ, for the hypotheses H_0 and H_1 given above.

Explain why the test you worked out in **a**) is not a generalised likelihood ratio test (GLRT).

Problem 3

In this exercise we will consider a regression model that is somewhat modified relative to what is discussed in the textbook. Assume that we have pairs of variables

$$(x_1, Y_1), (x_2, Y_2), ...(x_n, Y_n)$$

where $x_1, x_2, ..., x_n$ are positive and nonstochastic while $Y_1, Y_2, ..., Y_n$ are assumed to be independent random variables with

$$Y_i \sim \mathrm{N}(\beta x_i, \sigma_0^2 x_i^2).$$

Thus, the variance of Y_i is assumed to be proportional to x_i^2. In this exercise we will assume the value of σ_0^2 to be known while the parameter β is estimated from the available data.

a) Work out the maximum likelihood estimator (MLE) $\hat{\beta}$ for β. Show that $\hat{\beta}$ is unbiased and find the variance of $\hat{\beta}$.

b) What is the probability distribution of $\hat{\beta}$? Give reason for the answer.

Work out a $100(1-\alpha)\%$-confidence interval for β.

Problem 4

Let X_1 and X_2 be independent and identically distributed, $X_i \sim \chi_n^2$ $(n > 2)$.

a) Show that

$$\mathrm{E}\left(\frac{1}{X_1 + X_2}\right) = \frac{1}{2(n-1)} \quad \text{and} \quad \mathrm{E}\left[\frac{1}{(X_1 + X_2)^2}\right] = \frac{1}{4(n-1)(n-2)}.$$

Problem 5

The following table is an ANOVA table in which some entries are lost (stars).

Source	df	SS	MS	F
Treatment	$\star$	18.1	3.62	$\star$
Error	80	$\star$	$\star$	
Total	$\star$	247.7		

a) Find lost values and fill in the ANOVA table. Show how you calculate values where there are $\star$ in the table. Do all the samples have the same size?

b) Test the hypothesis that

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$$

versus

$$H_1 : \text{ not all are equal.}$$

The significance level $\alpha = 0.05$.

Problem 6

One studies the difference between the effect of an old fertilizer (denoted as X_1) and a newly developed fertilizer (denoted as X_2) on the growth of plants. An experiment was conducted where plants were grown under identical conditions, and one randomly assigned fertilizer X_1 to $n_1 = 7$ plants and fertilizer X_2 to $n_2 = 8$ plants. After 3 weeks the height of the plants were measured in cm. Data from the experiment is presented below.

i	1	2	3	4	5	6	7	8
x_{1i}	31	33	29	35	39	42	38	
x_{2i}	34	36	30	32	37	43	39	40

Descriptive measures for this dataset are

$$\bar{x}_1 = \frac{1}{7}\sum_{i=1}^{7} x_{1i} = 35.29, \;\; s_{x1} = \sqrt{\frac{1}{6}\sum_{i=1}^{7}(x_{1i} - \bar{x}_1)^2} = 4.64,$$

$$\bar{x}_2 = \frac{1}{8}\sum_{i=1}^{8} x_{2i} = 36.38, \;\; s_{x2} = \sqrt{\frac{1}{7}\sum_{i=1}^{8}(x_{2i} - \bar{x}_2)^2} = 4.31.$$

a) We assume that X_{1i} and X_{2i} are normally distributed, $X_{1i} \sim N(\mu_1, \sigma^2)$, $i = 1, ..., 7$, and $X_{2i} \sim N(\mu_2, \sigma^2)$, $i = 1, ..., 8$.

Based on this experiment, can one conclude that the mean height of the plants given the two different types (X_1 and X_2) of fertilizers are different? Write down the null hypothesis and the alternative hypothesis. Choose a test statistic and perform a hypothesis test. Use significance level $\alpha = 0.05$. Specify the assumptions you make.

b) What is the difference between a non-parametric hypothesis test and the test used in a)?

Perform a Wilcoxon rank-sum test based on the data given above. You can assume a normal approximation to the test statistic.