NTNU
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4255 Applied statistics**

Academic contact during examination: Nikolai Ushakov

Phone: 45128897

Examination date: August 2017

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

Permitted examination support material: C:
– Tabeller og formler i statistikk, Tapir forlag,
– Stamped yellow A4 sheet with your own handwritten notes,
– Specified calculator.

Other information:
– In outputs from MINITAB comma is used as decimal separator.
– Significance level 5% should be used unless a different level is specified.
– all answers need to be justified.

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

Three independent samples $X_1, X_2, ..., X_n$; $Y_1, Y_2, ..., Y_n$; $Z_1, Z_2, ..., Z_n$ of independent, and identically distributed in each sample, random variables are drawn from normally distributed populations with means μ_X, μ_Y, μ_Z, respectively, and the same variance σ^2. A one-way analysis of variance model (ANOVA) was fitted to these data (X-s are in column C1, Y-s are in column C2, Z-s are in column C3,), and the MINITAB output (a part) is given below

```
One-way ANOVA: C1;C2;C3

Source   DF      SS      MS      F      P
Factor    2   47,23   23,61  22,87  0,000
Error    57   58,86    1,03
Total    59 106,08

Level    N    Mean    StDev
C1      20  -0,296    0,977
C2      20   0,341    1,025
C3      20   1,822    1,046

Pooled StDev = 1,016

Tukey 95% Simultaneous Confidence Intervals
All Pairwise Comparisons

C1 subtracted from:

      Lower  Center  Upper  -----+---------+---------+---------+----
C2   -0,135   0,637  1,410              (----*----)
C3    1,345   2,118  2,890                        (----*----)
                            -----+---------+---------+---------+----
                               -1,5       0,0       1,5       3,0

C2 subtracted from:

      Lower  Center  Upper  -----+---------+---------+---------+----
C3    0,708   1,481  2,253                    (----*----)
                            -----+---------+---------+---------+----
                               -1,5       0,0       1,5       3,0
```

a) Find

$$n, \sum_{i=1}^{n} Y_i, \sum_{i=1}^{n} Y_i^2.$$

How is [Pooled StDev = 1,016] related with the numbers in the column StDev?

If the hypothesis

$$H_0 : \mu_X = \mu_Y = \mu_Z$$

is tested versus the alternative

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

at the significance level 0.001, what is the conclusion?

b) The hypothesis $H_0 : \mu_X = 0$ is to be tested against the alternative $H_1 : \mu_X < 0$. Which test statistic should be used? How can the value of this test statistic be found from the numbers given in the output? Find this value. What is the conclusion if the significance level is 0.05?

c) How can the output be used for simultaneous testing of the three hypotheses:

$$H_0 : \mu_X = \mu_Y, \; H_1 : \mu_X \neq \mu_Y,$$

$$H_0 : \mu_X = \mu_Z, \; H_1 : \mu_X \neq \mu_Z,$$

$$H_0 : \mu_Y = \mu_Z, \; H_1 : \mu_Y \neq \mu_Z?$$

How are the three intervals in the Tukey part of the output obtained? What are the conclusions if the significance level is 0.05?

Problem 2

Consider an in-control process with mean $\mu_0 = 10$ and standard deviation $\sigma = \sqrt{5}$. Suppose that subgroups of size $n = 5$ are used with control limits $\mu_0 \pm 3\sigma/\sqrt{n}$, and centerline at μ_0. Suppose that a shift occurs in the mean, and the new mean is $\mu = 12$.

a) What is the average number of samples required (following the shift) to detect the out-of-control situation?

Problem 3

Let $X_1, ..., X_n$ be a random sample from a normal distribution with expectation μ_X and variance σ^2, and $Y_1, ..., Y_m$ be a random sample from a normal distribution with expectation μ_Y and variance $k\sigma^2$ Assume that X-s og Y-s all are independent. Parameters μ_X and μ_Y are unknown and we have to estimate them and to construct a confidence interval for the difference $\mu_X - \mu_Y$.

We use $\bar{X}$ and $\bar{Y}$ as estimators of μ_X and μ_Y respectively.

a) Argue that $\bar{X} - \bar{Y}$ has a normal distribution and show that

$$\mathrm{E}(\bar{X} - \bar{Y}) = \mu_X - \mu_Y, \quad \mathrm{Var}(\bar{X} - \bar{Y}) = \sigma^2 \left(\frac{1}{n} + \frac{k}{m} \right).$$

Assume in this point that both parameters σ^2 and k are known. Construct an $(1 - \alpha)$-confidence interval for the difference $\mu_X - \mu_Y$.

In the rest of the problem, we assume that the parameter σ^2 is unknown, while the parameter k is known.

Let S_x^2 and S_y^2 be empirical variances of X-s and Y-s, i.e.

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2, \quad S_y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y})^2.$$

It is known that

$$\frac{(n-1)S_x^2}{\sigma^2} \sim \chi^2_{n-1}, \quad \frac{(m-1)S_y^2}{k\sigma^2} \sim \chi^2_{m-1},$$

and that S_x^2 and $\bar{X}$ are independent, and S_y^2 and $\bar{Y}$ are independent.

b) Show that

$$S_p^2 = \frac{n-1}{n+m-2} S_x^2 + \frac{m-1}{k(n+m-2)} S_y^2$$

is an unbiased estimator of σ^2 and that

$$(n+m-2)S_p^2/\sigma^2 \sim \chi^2_{n+m-2}.$$

Show further that

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{S_p^2 (\frac{1}{n} + \frac{k}{m})}}$$

has the Student t-distribution with $n + m - 2$ degrees of freedom.

Specify which properties and connections between various types of distributions you use and explain why necessary conditions are satisfied.

c) Construct a $(1 - \alpha)$-confidence interval for the difference $\mu_X - \mu_Y$.

Problem 4

A simple linear regression model

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i,\ i = 1, 2, ..., n,$$

where ϵ_i are independent, $\epsilon_i \sim N(0, \sigma^2)$, is fitted to some data. The least squares estimate $\hat{\beta}_1$ of β_1 is $\hat{\beta}_1 = 1$, and the 95% confidence interval symmetric about $\hat{\beta}_1$ is $[0.5, 1.5]$.

a) Find a 99% confidence interval for β_1 if it is known that the statistic

$$\frac{(n-2)S^2}{\sigma^2},$$

where S^2 is the least squares estimator of σ^2, has the chi-square distribution with 18 degrees of freedom.