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: 04 June 2019

Examination time (from–to): 15:00 – 19:00

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

A research lab will investigate the vibration wich occurs when thin plates of polyethylen are exposed to wind.

The experiment uses plates of equal thickness and width, but with three different lengths. The plates are exposed to four different wind speeds, and the number of oscillations in a certain time period are measured. Two experiments are performed for each of the 12 combinations of length and wind speed.

The responses of the individual experiments are denoted Y_{ijk} and are given as one hundredth of the number of oscillations per second when plate no. k with length x_{1i} (factor level A_i) is exposed to wind of speed x_{2j} (factor level B_j); for $i = 1, 2, 3$; $j = 1, 2, 3, 4$ and $k = 1, 2$.

The result of the experiment is given in the table below. The unity for length is *inches*, while the unity for wind speed is *feet/second*.

	$B_1 = 62.5$	$B_2 = 54.6$	$B_3 = 44.3$	$B_4 = 31.3$
$A_1 = 1.50$	50.5	46.0	36.5	23.0
	50.0	45.1	37.0	24.5
$A_2 = 1.75$	47.0	41.5	33.1	22.0
	48.0	42.0	34.1	24.2
$A_3 = 2.00$	45.5	39.4	30.8	20.3
	45.1	38.8	31.0	21.6

The first analysis of the data is by two-factor analysis of variance, where the factors A and B have, respectively, 3 and 4 levels. You may in the problems below use the following output from MINITAB:

Two-way ANOVA: Y versus A; B

Source	DF	SS	MS	F
A	2	100,54	50,268	93,52
B	3	2145,40	715,134	1330,48
Interaction	6	8,99	1,498	2,79
Error	12	6,45	0,537	
Total	23	2261,38		

a) Write down the model for the observations $\{Y_{ijk}\}$ which takes into account a possible interaction between the length effect (A) and the wind speed effect (B).

How can you test whether there is an interaction between the two factors? Write down and motivate a test statistic for this, and then find the critical value when the significance level is set to 5%. What is the conclusion?

Are the main effects of length and wind speed significant at 5% level? Answer the question by referring to the values of the test statistics given in the MINITAB output.

b) Write down the expressions for SSE and $SS(AB)$ in the model from question (a).

It is known from the course, that SSE/σ^2 is chi-square distributed with 12 degrees of freedom. Write down an expression for the estimator S^2 for σ^2 based on SSE. Find the numerical value of S^2 in the MINITAB-output.

Assume now that there is no interaction between the factors A and B, i.e. that all the parameters $(\alpha\beta)_{ij}$ in the model in question (a) equal 0. It can then be shown that $SS(AB)/\sigma^2$ is chi-square distributed with 6 degrees of freedom, and that $SS(AB)$ is independent of SSE.

Show how a new estimator of σ^2 can be derived by using both SSE and $SS(AB)$. Find the estimator and compute the new estimate for σ^2 with the given data.

Problem 2

The situation is the same as in Problem 1. It is now assumed that

$$Y_{ijk} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2j} + \epsilon_{ijk} \tag{1}$$

for $i = 1, 2, 3$; $j = 1, 2, 3, 4$ and $k = 1, 2$.

The output from a regression analysis using MINITAB, based on the data in Problem 1, is given on the next page. You may use this when solving the Problem.

a) Explain briefly how this model can be treated as a multiple linear regression model with 24 observations and explanatory variables $x_1 =$ length and $x_2 =$ wind speed.

Which assumptions are made in the model (1)?

What portion of the variation in the data is explained by the model? Is a significant amount of variation explained?

One will test the hypotheses $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$ and $H_0 : \beta_2 = 0$ vs. $H_1 : \beta_2 \neq 0$ in this model.

Write down the relevant test statistics and perform the two tests with significance level 1%.

To what extent do the results of this testing support the conclusion from the analysis of variance in Problem 1?

Regression Analysis: Y versus x1; x2

```
The regression equation is
Y = 15,1 - 10,0 x1 + 0,808 x2

Predictor       Coef   SE Coef       T
Constant      15,140     1,848    8,19
x1          -10,0250    0,9466  -10,59
x2           0,80842   0,01653   48,89

Analysis of Variance

Source          DF      SS      MS        F       P
Regression       2  2242,6  1121,3  1251,40   0,000
Residual Error  21    18,8     0,9
Total           23  2261,4
```

Problem 3

Let $X_1, ..., X_{100}$ be a random sample from a normal distribution with unknown expectation μ and variance $\sigma^2 = 25$. The hypothesis $H_0 : \mu = 0$ is tested against $H_1 : \mu > 0$. H_0 is rejected for large values of $\bar{X}$, i.e. the critical region has the form $\{\bar{X} > c\}$. Suppose that for $\mu = 1$, the probability of rejecting the null hypothesis H_0 is 0.5.

a) What is the value of the significance level α?

b) What is the probability of rejecting H_0 if $\mu = 2$?

Problem 4

In order to investigate the connection between high blood pressure and smoking, one has collected the following information from 180 randomly selected persons:

	Non-smoker	Moderate smoker	Heavy smoker
High blood pressure	20	36	32
Normal blood pressure	48	26	18

a) Test the null hypothesis that there is independence between the occurrence of high/normal blood pressure and smoking habit. What is the conclusion? Use significance level 1%.

Problem 5

A part of some MINITAB output is given below:

One-Sample Z: C1

```
Test of mu = 0 vs not = 0
The assumed standard deviation = 1

Variable     N      Mean   StDev  SE Mean          95% CI         Z       P
C1          50     0,043   0,908    0,141    (-0,234; 0,320)  0,31   1,240
```

a) One value in the output is wrong. Which value? Why do you think it is wrong? Find the right value.

b) Find a 99.99% confidence interval for the expectation μ.