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 2018

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

A biologist will examine how the growth of mussels is affected by light when certain other factors are held constant. The growth (y) for 10 mussels are measured under different degrees of lighting (x). The observations (y_i, x_i) for $i = 1, 2, ..., 10$ are given in the table below:

i	1	2	3	4	5	6	7	8	9	10
x_i	1	1	2	2	3	3	4	4	5	5
y_i	16	18	17	20	25	21	23	20	17	19

The MINITAB output and plots given below show the results of fitting a multiple linear regression model where the expected growth y is a second order polynomial in lighting X. More precisely, the assumed model is:

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$$

for $i = 1, ..., 10$, where $\epsilon_1, ..., \epsilon_{10}$ are independent and $N(0, \sigma^2)$.

In the MINITAB output is the covariate x^2 denoted as $x * x$. Here is the MINITAB output:

```
Regression Analysis: Y versus x; x*x

The regression equation is
Y = 10,1 + 7,36 x - 1,14 x*x

Predictor      Coef   SE Coef      T      P
Constant     10,100     3,183   3,17  0,016
x             7,357     2,425   3,03  0,019
x*x         -1,1429    0,3966  -2,88  0,024

S = 2,09859   R-Sq = 57,4%   R-Sq(adj) = 45,3%

Analysis of Variance

Source          DF      SS      MS     F      P
Regression       2  41,571  20,786  4,72  0,050
Residual Error   7  30,829   4,404
Total            9  72,400
```

a) What is the meaning of the $F-$value 4.72 in the table "Analysis of Variance" in the output? Write down the null hypothesis that is tested with this test statistic. How can you conclude from the displayed $P-$value?

b) The hypothesis $H_0 : \beta_2 = 0$ is tested vs. $H_1 : \beta_2 < 0$. Write down a test statistic for this and find the critical value when the significance level is 0.05. What is the conclusion of the test? What is the $P-$value of this test?

c) Write down the estimate for σ that can be read out from the MINITAB output.

Find a 95% confidence interval for σ.

How can the confidence interval be used to test the null hypothesis $H_0 : \sigma = 1$? What is in that case the alternative hypothesis, and what is the significance level?

d) Show that the estimated regression equation has a maximum at $x_0 = 3.23$.

One wants a prediction interval for the response y_0 for this $x-$value. First, calculate the point prediction $\hat{y}_0$.It is given that the estimated standard deviation of $\hat{y}_0$ is 1.024. Use this to calculate a 95% prediction interval for y_0.

Problem 2

An agent goes to regular shooting practice. Experience tells him that the probability of hit is $p = 0.6$. During a practice session he has 20 trials. Assume that each shot is either a hit or a miss, and that the trials are independent. The boss decides that the agent should have a new gun. They hope that this new one results in a better hitting probability. They want to check if this may hold, and the agent does a usual practice session consisting of 20 trials with the new gun.

a) Formulate the problem as a hypothesis test.

Use the common normal approximation to perform the test at significance level $\alpha = 0.05$ when the observed number of hits is 18.

b) What is the P-value of the test when he hits on 18 shots?

Problem 3

Darwin (1876) studied the growth of pairs of corn plants, where one plant was produced by cross-fertilization and the other produced by self-fertilization. His goal was to demonstrate the greater fitness (e.g. survival and growth) of cross-fertilized plants compared to self-fertilized plants.

Fifteen pairs of plants were grown together under identical conditions in every pair (but possibly under different conditions in different pairs). The data recorded are the height (in inches) of the plants in each pair.

For pair i let X_{1i} denote the height of the plant from the seedling produced by cross-fertilization and X_{2i} denote the height of the plant from seedling produced by self-fertilization, $i = 1, ..., 15$. Further, let $D_i = X_{1i} - X_{2i}$. Data from the experiment are presented below.

i	1	2	3	4	5	6	7	8	9	10
x_{1i}	188	96	168	176	153	172	177	163	146	173
x_{2i}	130	163	160	160	147	149	149	122	132	144

i	11	12	13	14	15
x_{1i}	186	168	177	184	96
x_{2i}	130	144	102	124	144

Descritive measures for this dataset are

$$\bar{d} = \frac{1}{15}\sum_{i=1}^{15} d_i = 21.53,$$

$$s_d = \sqrt{\frac{1}{14}\sum_{i=1}^{15}(d_i - \bar{d})^2} = 38.29.$$

a) Assume that X_{1i} and X_{2i} are normally distributed, $X_{1i} \sim N(\mu_1 + \beta_i, \sigma^2)$, $X_{2i} \sim N(\mu_2 + \beta_i, \sigma^2)$, and $X_{1i} - X_{2i}$ are independent, $i = 1, 2, ..., 15$.

Based on this experiment, could Darwin conclude that cross-fertilazed plants are taller than self-fertilized plants? Write down the null hypothesis and the alternative hypothesis, choose a test statistics and perform a hypothesis test. Use significance level $\alpha = 0.05$.

b) Assume that X_{1i} og X_{2i} are not normally distributed (but have simmetrical around expectations distributions). Perform a sign test to test if cross-fertilized plants are taller than self-fertilized plants.