Exam question paper for:

Written examination: MAY 2025 TEST

Course name and number: **02402 Statistics (Polytechnical Foundation)**

Duration: 4 hours

The final answers should be handed in by filling out a separate "Answer Sheet".

This exam consists of 30 questions of the "multiple choice" type, which are divided between 13 exercises.
Only hand in the "Answer Sheet" and not the entire question paper.

Multiple choice questions: *Note that in each question, one and only one of the answer options is correct. Furthermore, not all the suggested answers are necessarily meaningful. Always remember to round off your own result to the number of decimals given in the answer options before you choose your answer.*

The use of Python code in this exam: *This exam includes Python code. Note that we use the following libraries and abbreviations:*

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import scipy.stats as stats
import statsmodels.api as sm
import statsmodels.formula.api as smf
import statsmodels.stats.power as smp
import statsmodels.stats.proportion as smprop
```

Continue on page 2

Exercise I

Let Y follow an exponential distribution with rate parameter 2, and let U follow a continuous uniform distribution on the interval $[3, 6]$. The two random variables are independent.

Question I.1 (1)

What is the probability that Y exceeds 2?

1 ☐ $2^{-2} = 0.250$

2 ☐ $2^{-1} = 0.500$

3 ☐ $1 - e^{-4} = 0.982$

4 ☐ $e^{-4} = 0.018$

5 ☐ $2e^{-4} = 0.037$

6 ☐ Don't know / No answer

Question I.2 (2)

What is the standard deviation of U?

1 ☐ $\frac{3}{2}$

2 ☐ $\frac{3}{4}$

3 ☐ $\frac{\sqrt{3}}{2}$

4 ☐ $\frac{\sqrt{3}}{4}$

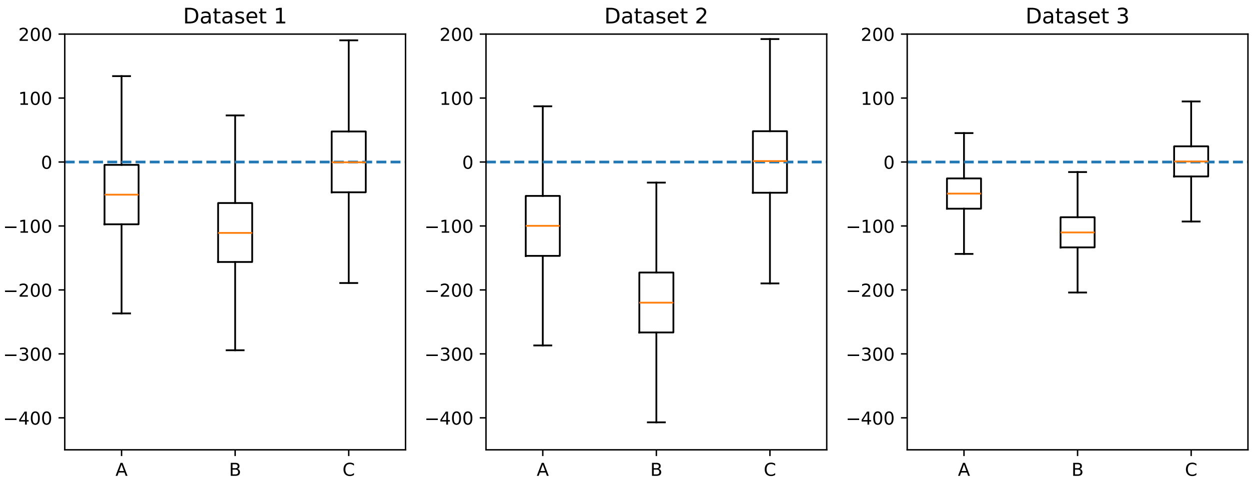
5 ☐ $\frac{9}{2}$

6 ☐ Don't know / No answer

Continue on page 3

Exercise II

Three experiments compare three medications (A, B, and C) designed to regulate appetite. For each individual in each experiment, the decrease in calorie intake was measured. The results are stored in Dataset 1, Dataset 2, and Dataset 3 and are visualized in the boxplots below:



Question II.1 (3)

Which of the following statements is false?

1 ☐ The average effects of the medications seem the same in Datasets 1 and 3.

2 ☐ The SS(TR) - *Treatment Sum of Squares* - is largest in Dataset 2.

3 ☐ The SSE - *Sum of Squared Errors* - is largest in Dataset 3.

4 ☐ Within each dataset, the variance in each group (A, B, and C) is approximately equal.

5 ☐ In all three datasets, medication C could be a placebo (i.e., have no average effect).

6 ☐ Don't know / No answer

Question II.2 (4)

The sample size of each group is the same in all three datasets (i.e., all nine groups are of the same size). For each dataset, an analysis of variance (ANOVA) is performed, and the observed F-statistic and p-value are computed for the usual null hypothesis. Which of the following statements is correct?

1 ☐ The observed F-statistic is the same for all three datasets.

2 ☐ The test performed is a Welch t-test.

3 ☐ The p-value is the same for all three datasets.

4 ☐ The assumption of equal variances across groups appears violated in Dataset 1.

5 ☐ The p-value for Dataset 1 is larger than those for Datasets 2 and 3.

6 ☐ Don't know / No answer

Continue on page 5

Exercise III

A fitness coach claims that the average weight loss for a new diet program is 4 kg. To test this claim, a random sample of 12 participants who completed the program was taken, and their weight losses (in kg) were recorded: 3.5, 4.2, 3.8, 4.4, 4.1, 3.9, 4.0, 4.3, 3.6, 3.7, 4.5, 4.1.

The researchers then perform a t-test to determine whether the average weight loss differs significantly from 4 kg at a significance level of 5%.

Question III.1 (5)

Let t_{obs} be the observed test statistic for the t-test. What is the rejection rule?

1 ☐ Reject H_0 if $t_{\text{obs}} \leq t_{1-\alpha}$, where $t_{1-\alpha}$ is the $(1-\alpha)$ quantile of a t-distribution with 11 degrees of freedom.

2 ☐ Reject H_0 if $t_{\text{obs}} \leq t_{1-\alpha}$, where $t_{1-\alpha}$ is the $(1-\alpha)$ quantile of a t-distribution with 12 degrees of freedom.

3 ☐ Reject H_0 if $t_{\text{obs}} \geq t_{1-\alpha}$, where $t_{1-\alpha}$ is the $(1-\alpha)$ quantile of a t-distribution with 11 degrees of freedom.

4 ☐ Reject H_0 if $t_{\text{obs}} \leq -t_{\alpha/2}$ or if $t_{\text{obs}} \geq t_{1-\alpha/2}$, where $t_{\alpha/2}$ and $t_{1-\alpha/2}$ are the $\alpha/2$ and $(1-\alpha/2)$ quantiles of a t-distribution with 11 degrees of freedom, respectively.

5 ☐ Reject H_0 if $t_{\text{obs}} \leq -t_{\alpha/2}$ or if $t_{\text{obs}} \geq t_{1-\alpha/2}$, where $t_{\alpha/2}$ and $t_{1-\alpha/2}$ are the $\alpha/2$ and $(1-\alpha/2)$ quantiles of a t-distribution with 10 degrees of freedom, respectively.

6 ☐ Don't know / No answer

Question III.2 (6)

Given the sample mean: 4.0083 and the sample standard deviation: 0.3175 What is the value of the observed test statistic (t_{obs})?

1 ☐ $t_{\text{obs}} = 2.83$

2 ☐ $t_{\text{obs}} = 1.09$

3 ☐ $t_{\text{obs}} = 1.03$

4 ☐ $t_{\text{obs}} = 0.32$

5 ☐ $t_{\text{obs}} = 0.09$

6 ☐ Don't know / No answer

Continue on page 6

Exercise IV

A computer engineer wants to test the efficiency of an algorithm. She records the execution time (`time`) for jobs of varying complexity (`complex`).

To analyze the data, she runs the code below, where `dat` contains the recorded data.

```
fit1 = smf.ols(formula = 'time ~ complex', data = dat).fit()
print(fit1.summary(slim=True))
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   time   R-squared:                       0.971
Model:                            OLS   Adj. R-squared:                  0.971
No. Observations:                  50   F-statistic:                     1631.
Covariance Type:            nonrobust   Prob (F-statistic):           1.03e-38
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     -0.1732      0.007    -23.620      0.000      -0.188      -0.158
complex        0.0006   1.47e-05     40.391      0.000       0.001       0.001
==============================================================================
```

Question IV.1 (7)

Consider the test statistic for the null hypothesis that the intercept is zero. Which distribution is the test statistic compared to in order to obtain the associated p-value?

1 ☐ A $t(50)$-distribution

2 ☐ A $t(48)$-distribution

3 ☐ An $F(1, 50)$-distribution

4 ☐ A $\mathcal{N}(0, 1^2)$-distribution

5 ☐ A $\mathcal{N}(0, 0.007^2)$-distribution

6 ☐ Don't know / No answer

Question IV.2 (8)

Based on the output above, which statement about the validity of the model (i.e., the model assumptions) is correct?

1 ☐ The validity of the model assumptions cannot be assessed based on the output.

2 ☐ The assumptions must be satisfied because the R^2-value is close to 1.

3 ☐ The model should be extended with additional terms because the p-values are small.

4 ☐ The assumptions must be satisfied because the p-values are close to zero.

5 ☐ The model should be extended with additional terms because the R^2-value is close to 1.

6 ☐ Don't know / No answer

Question IV.3 (9)

For the following questions, the numbers below may be useful:

```
print(np.mean(dat["complex"]))
484.58
print(np.var(dat["complex"], ddof=1))
12603.187346938777
```

Also, we provide the following quantiles from t-distributions (stated in the form "$t_{1-\alpha/2,\nu}$", where α is the significance level and ν is the degrees of freedom):

$$t_{0.975,\,50} = 2.0086, \qquad t_{0.975,\,49} = 2.0096, \qquad t_{0.975,\,48} = 2.0106.$$

Let $\hat{\sigma}$ denote the estimated residual standard deviation. What is the 95% confidence interval for the mean (execution) time of a job with complexity 300 according to the model?

1 ☐ $0.12 \pm 0.45 \cdot \hat{\sigma}$

2 ☐ $0.0068 \pm 0.15 \cdot \hat{\sigma}$

3 ☐ $0.12 \pm 0.201 \cdot \hat{\sigma}$

4 ☐ $0.0068 \pm 0.201 \cdot \hat{\sigma}$

5 ☐ $0.0068 \pm 0.55 \cdot \hat{\sigma}$

6 ☐ Don't know / No answer

Question IV.4 (10)

What is the estimate of the residual standard deviation ($\hat{\sigma}$)?

(To answer this question you may need some of the numbers and additional information provided under the previous question)

1 ☐ $2 \cdot 10^{-5}$

2 ☐ 0.01

3 ☐ 5.5

4 ☐ 0.97

5 ☐ 0.007

6 ☐ Don't know / No answer

Question IV.5 (11)

The engineer figures that the execution time is proportional to the third power of complexity and formulates the following model:

$$Y_i = \beta_0 x_i^{\beta_1} \varepsilon_i; \quad \varepsilon_i \sim LN(0, \sigma^2),$$

where Y_i and x_i are the time and complexity of job i, respectively. She hence tests the hypothesis $H_0 : \beta_1 = 3$ against a two-sided alternative by running the below code chunk:

```
ltime = np.log(dat['time'])
lcomplex = np.log(dat['complex'])
dat2 = pd.DataFrame({'ltime' : ltime, 'lcomplex' : lcomplex})
fit2 = smf.ols(formula = 'ltime ~ lcomplex',data=dat2).fit()
print(fit2.summary(slim=True))
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  ltime   R-squared:                       0.999
Model:                            OLS   Adj. R-squared:                  0.999
No. Observations:                  50   F-statistic:                 8.032e+04
Covariance Type:            nonrobust   Prob (F-statistic):           4.87e-79
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    -18.0015      0.055   -325.241      0.000     -18.113     -17.890
lcomplex       2.5459      0.009    283.406      0.000       2.528       2.564
==============================================================================
```

Based on the output above, what is the test statistic (t_{obs}) for the usual test of H_0, and what is the conclusion of the test at a significance level of 5%?

1 ☐ The hypothesis is rejected as $t_{\text{obs}} = -5.645$.

2 ☐ The hypothesis is accepted as $t_{\text{obs}} = 13.84$.

3 ☐ The hypothesis is rejected as $t_{\text{obs}} = -50.46$.

4 ☐ The hypothesis is rejected as $t_{\text{obs}} = 283.4$.

5 ☐ The hypothesis is accepted as $t_{\text{obs}} = -5.645$.

6 ☐ Don't know / No answer

Continue on page 10

Exercise V

The two questions in this exercise concern a simulation carried out in Python.

The number of customers shopping in a supermarket on a randomly selected day follows a Poisson distribution with a mean of 5000. The amount a randomly selected customer spends is exponentially distributed, with an average of 200 DKK. It is assumed that customers act independently of each other and independently of the number of customers.

The following code simulates k daily revenues, but you need to complete the missing parts of the code by specifying the parameters of the distributions:

```
# Set seed
np.random.seed(2025)

# Number of samples
k = 100000

# Number of customers (array of customer counts for the k days)
N = stats.poisson.rvs(mu=____,size=k)

# The revenues (array with revenues for the k days)
Y = np.array([np.sum(stats.expon.rvs(scale=_____, size=n)) for n in N])
```

Let Y denote the supermarket's revenue (or turnover) on a randomly selected day.

Question V.1 (12)

Identify and fill in the two missing parameters in the simulation code: the mean μ of the Poisson distribution and the scale parameter of the exponential distribution.

1 ☐ $\mu = 200$ and scale $= 5000$

2 ☐ $\mu = 5000$ and scale $= 1/200$

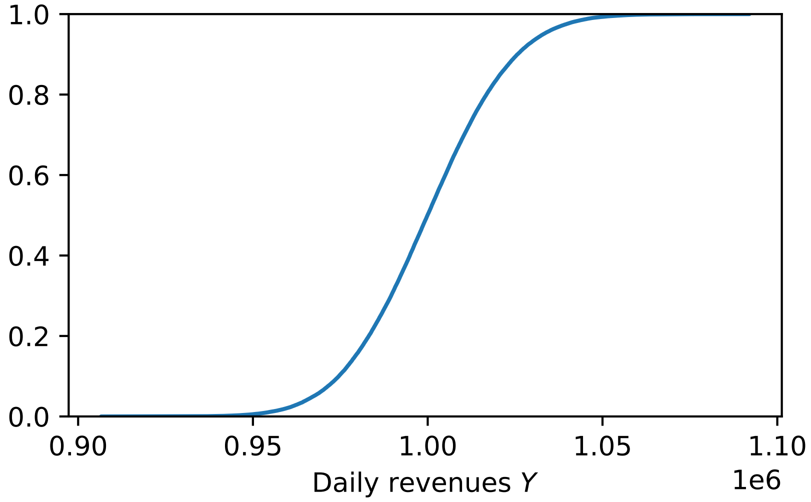
3 ☐ $\mu = 200$ and scale $= 1/5000$

4 ☐ $\mu = 5000$ and scale $= 200$

5 ☐ $\mu = 1/5000$ and scale $= 200$

6 ☐ Don't know / No answer

Question V.2 (13)



The figure above shows the empirical cumulative distribution function (ECDF) of the simulated daily revenues Y obtained from the simulation above.

What is the probability that the supermarket makes more than 1,500,000 DKK in revenue on a randomly selected day?

1 ☐ Almost 0%

2 ☐ Approximately 11%

3 ☐ Approximately 22%

4 ☐ Approximately 33%

5 ☐ Approximately 50%

6 ☐ Don't know / No answer

Continue on page 12

Exercise VI

A shipping company has hired a sales representative to recruit new clients. The sales representative visits n (independent) clients each month, and historical data suggests that the probability of recruiting a client after a visit is p. Let X denote the number of recruited clients in a randomly selected month.

Question VI.1 (14)

Which one of the following models is the most appropriate?

1 ☐ $X \sim U(0, n)$

2 ☐ $X \sim H(n, np, n)$

3 ☐ $X \sim \mathcal{N}(np, p^2)$

4 ☐ $X \sim \text{Pois}(np)$

5 ☐ $X \sim \text{Bin}(n, p)$

6 ☐ Don't know / No answer

The sales representative earns a salary of 5000 USD per month and receives an additional bonus of 100 USD for each client recruited during the month. The sales representative's total compensation in a randomly selected month (Y) is therefore given by

$$Y = 100X + 5000,$$

where X denotes the number of clients recruited during that month. You may assume that the standard deviation of X is 2.

Question VI.2 (15)

What is the variance of Y?

1 ☐ 5200

2 ☐ 20000

3 ☐ 25000

4 ☐ 40000

5 ☐ 45000

6 ☐ Don't know / No answer

Question VI.3 (16)

Which one of the following statements about the correlation between X and Y is true?

1 ☐ The correlation between X and Y is one: $\rho(X, Y) = 1$.

2 ☐ The correlation between X and Y is between one and zero: $0 < \rho(X, Y) < 1$.

3 ☐ The correlation between X and Y is zero: $\rho(X, Y) = 0$.

4 ☐ The correlation between X and Y is negative: $-1 \leq \rho(X, Y) < 0$.

5 ☐ The correlation cannot be determined without additional information.

6 ☐ Don't know / No answer

Continue on page 14

Exercise VII

A municipality tested a service that allows citizens to report damages to traffic infrastructure (e.g., roads and traffic lights) via an app. The reported issues were categorized into three groups: needs immediate repair, can be postponed, or can be ignored.

During a one-year trial, $n = 165$ independent issues were reported. The counts for each issue type were compiled by season and entered into the below table:

```
# Reading the data into Python
data = np.array([[44, 10, 32, 18], [20, 8, 4, 2], [11, 2, 10, 4]])
# Converting to a Pandas dataframe
data = pd.DataFrame(data, index=['Immediate', 'Postpone', 'Ignore'],
columns=['Spring', 'Summer', 'Fall', 'Winter'])
print(data)
           Spring  Summer  Fall  Winter
Immediate      44      10    32      18
Postpone       20       8     4       2
Ignore         11       2    10       4
```

The municipality plans to test if the distribution of fault types depends on the season. Specifically, it tests the null hypothesis

$$H_0 : p_{i1} = p_{i2} = p_{i3} = p_{i4} = p_i \text{ for } i = 1, 2, 3,$$

where p_{ij} represents the proportion of observations in column j that fall into row i i.e., the proportion of fault type i out of all faults in season j, and p_i represents the proportion of all observations that are in row i.

Question VII.1 (17)

Under the null hypothesis, what is the estimated proportion of faults requiring immediate repair?

1 ☐ 0.6303

2 ☐ 0.4231

3 ☐ 0.2061

4 ☐ 0.7939

5 ☐ None of the above answers are correct.

6 ☐ Don't know / No answer

Question VII.2 (18)

To assess whether the usual chi-square test is valid, the expected cell counts under the null hypothesis have been calculated using eq. 7–53 in chapter 7.5. The expected counts are shown in the table below:

	Spring	Summer	Fall	Winter
Immediate	47.27	12.61	28.99	15.13
Postpone	15.45	4.12	9.48	4.95
Ignore	12.27	3.27	7.53	3.93

At a significance level of $\alpha = 0.05$, what is the correct conclusion for the usual chi-square test of the null hypothesis? (Both argument and conclusion must be correct.)

1 ☐ The null hypothesis is rejected, indicating a significant difference in distribution between the seasons, as the p-value is 0.04.

2 ☐ The null hypothesis is rejected, indicating a significant difference in distribution between the seasons, as the p-value is 0.08.

3 ☐ The null hypothesis is accepted, indicating no significant difference in distribution between the seasons, as the p-value is 0.04.

4 ☐ The null hypothesis is accepted, indicating no significant difference in distribution between the seasons, as the p-value is 0.08.

5 ☐ No conclusion can be drawn, since some expected cell counts under the null hypothesis are too low, i.e. $e_{ij} < 5$ for some cells.

6 ☐ Don't know / No answer

Question VII.3 (19)

Another purpose was to test an AI system for automatically handling the reported faults, comparing its performance to manual evaluation, where a human categorized the faults. Based on expert knowledge, the number of discrepancies between the two methods is expected to be 21 (out of the 165 observed faults).

Let $X \sim \text{Bin}(n, p)$ be the count of how many faults were evaluated differently and let $p = \frac{X}{n}$ be the proportion. What is the estimate of the standard deviation of $\frac{X}{n}$?

1 ☐ $\hat{\sigma} = 0.3343$

2 ☐ $\hat{\sigma} = 0.1273$

3 ☐ $\hat{\sigma} = 0.1118$

4 ☐ $\hat{\sigma} = 0.0259$

5 ☐ $\hat{\sigma} = 0.0007$

6 ☐ Don't know / No answer

Continue on page 17

Exercise VIII

A researcher designs an experiment to estimate the mean time of a chemical reaction. The experiment should be able to detect a difference in mean of 0.5 seconds, assuming a standard deviation of 1.2 seconds for the reaction time. The researcher wants to use a 95% confidence level and a power of 80% for this experiment.

Question VIII.1 (20)

What is the minimum sample size required to achieve the desired specifications? (You may use the following quantiles from the standard normal distribution: $z_{0.975} \approx 1.96$ and $z_{0.80} \approx 0.84$)

1 ☐ 38

2 ☐ 45

3 ☐ 46

4 ☐ 48

5 ☐ 79

6 ☐ Don't know / No answer

Continue on page 18

Exercise IX

Consider the following model:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},$$

where the errors are assumed to be independent and normally distributed with $E[\varepsilon_{ij}] = 0$ and $V[\varepsilon_{ij}] = \sigma^2$. The parameters α_i are referred to as treatment effects, while the parameters β_j are referred to as block effects. The model is fitted to data from a study, where exactly one observation was made for each combination of treatment and block.

The model is associated with the following ANOVA table, in which some numbers have been replaced by letters:

Source	DF	SS	MS	F statistic	p-value
Treatment	A	C	E	G	0.078
Block	B	D	F	H	0.021
Residual	20	60	3		

Question IX.1 (21)

How many observations were in the study?

1 ☐ 19

2 ☐ 20

3 ☐ 29

4 ☐ 30

5 ☐ This cannot be determined from the given information.

6 ☐ Don't know / No answer

Question IX.2 (22)

What is the total sum of squares for the data?

1 ☐ C + D + 60

2 ☐ C + D - 60

3 ☐ C + D

4 ☐ E + F + 3

5 ☐ E + F - 3

6 ☐ Don't know / No answer

Consider the following quantiles of various F-distributions.

Quantile	1%	2.5%	5%	10%	90%	95%	97.5%	99%
F(A,20)-distribution	0.010	0.025	0.051	0.106	2.589	3.493	4.461	5.849
F(20,A)-distribution	0.171	0.224	0.286	0.386	9.441	19.446	39.448	99.449
F(B,20)-distribution	0.227	0.293	0.360	0.454	1.937	2.348	2.774	3.368
F(20,B)-distribution	0.297	0.361	0.426	0.516	2.201	2.774	3.419	4.405

Question IX.3 (23)

Which of the following pairs gives admissible values for G and H in the ANOVA table?

1 ☐ G = 0.082 and H = 0.278

2 ☐ G = 0.344 and H = 0.347

3 ☐ G = 2.906 and H = 2.884

4 ☐ G = 3.832 and H = 3.335

5 ☐ G = 12.264 and H = 3.594

6 ☐ Don't know / No answer

Continue on page 20

Exercise X

In an experiment, $n = 85$ individuals are tested for a binary response (positive or negative). Let the binomially distributed random variable

$$X \sim \text{Bin}(n, p)$$

denote the number of positive outcomes.

The following null hypothesis is considered

$$H_0 : p = 0.2.$$

Question X.1 (24)

Under the null hypothesis (i.e., assuming it is true), what is the probability of observing 3 or less positive outcomes?

1 ☐ $1.0 \cdot 10^{-5}$

2 ☐ $1.7 \cdot 10^{-6}$

3 ☐ $1.2 \cdot 10^{-4}$

4 ☐ $1.2 \cdot 10^{-3}$

5 ☐ $1.5 \cdot 10^{-2}$

6 ☐ Don't know / No answer

Question X.2 (25)

In the experiment 25 positive outcomes are observed. What is the observed test statistic in the usual test of the null hypothesis?

1 ☐ 8.000

2 ☐ 2.169

3 ☐ 1.904

4 ☐ 0.057

5 ☐ 0.030

6 ☐ Don't know / No answer

Continue on page 21

Exercise XI

A study is conducted to compare the average scores of two groups of students in a math test. Group A consists of 12 students with a mean score of 78 and a standard deviation of 10, while Group B consists of 15 students with a mean score of 82 and a standard deviation of 8. Researchers want to perform a two-sample t-test to determine if there is a significant difference between the means of the two groups at a 5% significance level i.e., they want to test the null hypothesis

$$H_0 : \mu_A - \mu_B = 0$$

against a two-sided alternative hypothesis.

Question XI.1 (26)

What is the observed test statistic in the appropriate test?

1 ☐ -3.422

2 ☐ -1.155

3 ☐ -1.127

4 ☐ -0.765

5 ☐ -0.317

Question XI.2 (27)

The test statistic follows a t-distribution with how many degrees of freedom?

1 ☐ 20.85

2 ☐ 22.38

3 ☐ 24.99

4 ☐ 25.00

5 ☐ 27.00

Exercise XII

Two students at DTU compute need to measure the force between two charged particles, given by:

$$F = k \cdot \frac{q_1 q_2}{r^2}$$

where $k = 8.98 \cdot 10^9 N \cdot m^2/C^2$.

Assume that q_1 and q_2 are given with perfect accuracy: $q_1 = 3.0 \cdot 10^{-6}$C and $q_2 = 2.5 \cdot 10^{-6}$C.

The students meassure the distance r many times and estimate the average distance $\overline{r} = 0.85$ m (85 cm) and its standard deviation $s_r = 0.03$ m (3 cm).

Question XII.1 (28)

The students wish to estimate the variance of F, namely V[F], what is its value?
(Recall: $\frac{\partial}{\partial x}(\frac{1}{x^2})) = \frac{-2}{x^3}$

1 ☐ $9.32 \cdot 10^{-3}$ N

2 ☐ $1.63 \cdot 10^{-5}$ N

3 ☐ $7.89 \cdot 10^{5}$ N

4 ☐ $8.88 \cdot 10^{-2}$ N

5 ☐ $9.32 \cdot 10^{3}$ m

6 ☐ Don't know / No answer

Question XII.2 (29)

After measuring the force between two charged particles, the students also record how many electrons are detected by their sensor per hour.
Past measurements show that, on average, the detector registers 7 electrons per hour, the electrons can be assumed to be detected at random times.
They want to estimate the median number of electrons detected per hour using a simulation of 10.000 samples.

Which code snippet correctly performs this simmulation?

1 ☐
```
y=stats.expon.cdf(7, 10000)
print(np.median(y))
```

2 ☐
```
y=stats.poisson.rvs(7, 10000)
print(np.median(y))
```

3 ☐
```
y=stats.poisson.cdf(7, 10000)
print(np.median(y))
```

4 ☐
```
y=stats.expon.rvs(7, 10000)
print(np.median(y))
```

5 ☐
```
y=2*(1-stats.poisson.cdf(7, 10000))
print(np.mean(y))
```

6 ☐ Don’t know / No answer

Continue on page 24

Exercise XIII

After seeing the skills which the two students at DTU compute had in measuring variance, two of their friends from DTU Chemistry wanted to show of their skills. After doing some research on enzyme kinematics, they came up with this model:

$$y_i = b_0 + b_1 x_i^2 + b_2 \frac{1}{x_i} + e_i \qquad \text{where } e \sim N(0, \sigma)$$

which they fitted to their data for y and x.

Question XIII.1 (30)

Which of the following, could be the top 5 rows in the corresponding Design Matrix?

1 ☐ $\begin{bmatrix} 1 & 0.25 & 2.000 \\ 1 & 1.00 & 1.000 \\ 1 & 2.25 & 0.667 \\ 1 & 4.00 & 0.500 \\ 1 & 9.00 & 0.333 \end{bmatrix}$

2 ☐ $\begin{bmatrix} 1 & 2.000 & 0.25 \\ 1 & 1.000 & 1.00 \\ 1 & 0.667 & 2.25 \\ 1 & 0.500 & 4.00 \\ 1 & 0.333 & 9.00 \end{bmatrix}$

3 ☐ $\begin{bmatrix} 0.25 & 2.000 \\ 1.00 & 1.000 \\ 2.25 & 0.667 \\ 4.00 & 0.500 \\ 9.00 & 0.333 \end{bmatrix}$

4 ☐ $\begin{bmatrix} 1 & 0.25 \\ 1 & 1.00 \\ 1 & 2.25 \\ 1 & 4.00 \\ 1 & 9.00 \end{bmatrix}$

5 ☐ $\begin{bmatrix} 0.25 & 2.000 \\ 1.00 & 1.000 \\ 2.25 & 0.667 \\ 4.00 & 0.500 \\ 9.00 & 0.333 \end{bmatrix}$

6 ☐ Don't know / No answer

The exam is finished.