This exam paper is available in both Danish and English. The English version appears after the Danish version.

Exam question paper for:

Written examination: 26.06.2026

Course name and number: **02323 Introduction to Statistics**

Duration: 4 hours

Aids allowed: All printed materials and a pocket calculator of type TI30XS or TI30XB

Weighting: The questions are weighted equally (see details below)

Final answers must be submitted by completing the separate "Answer Sheet".

This examination consists of 30 multiple-choice questions distributed across 12 exercises.
Only the "Answer Sheet"must be submitted; do not hand in the exam paper itself.

Multiple choice questions: *The exam is 30 multiple choice questions, with possible answers: 1, 2, 3, 4, 5, and 6 (where 6 always refers to the answer "Don't know/No answer"). A correct answer gives 5 points, a wrong answer gives −1 point, and "Don't know/No answer" (6) gives 0 points. It is not allowed to select more than one answer option for a single question. If this happens anyway, 0 points are given for the question.*

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

```
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

A student is taking a multiple choice exam with 30 questions. For each question there are five options where one and only one is correct. The student is wondering what the probability of answering at least 15 correctly if he answers every question completely at random.

Question I.1 (1)

Which of the following Python commands calculate the desired probability?

1 ☐ `1 - stats.hypergeom.cdf(14, 30, 6, 15)`

2 ☐ `1 - stats.hypergeom.cdf(15, 30, 5, 15)`

3* ☐ `1 - stats.binom.cdf(14, 30, 1/5)`

4 ☐ `1 - stats.hypergeom.pmf(15, 30, 6, 14)`

5 ☐ `1 - stats.binom.pmf(15, 30, 1/5)`

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

Answering at random imply that the number of correct answer follow a Binomial with success probability $p = 1/5$, and $n = 30$. Further the probability in question is

$$P(X \geq 15) = 1 - P(X \leq 14) \tag{1}$$

where $X \sim Binom(1/6, 30)$, which can be calculated by

```
1-stats.binom.cdf(14,30,1/5)


0.00023122561166777356
```

---------- FACIT-END ----------

In the exam students are awarded 5 points for a correct answer and -1 point for an incorrect answer.

Question I.2 (2)

Let the random variable Y be the total number of points achieved in the exam, still assuming that the student answer completely at random, what is the standard deviation of Y?

1 ☐ $\sqrt{30\left(\frac{4}{5}+5^2\cdot\frac{1}{5}\right)}=\sqrt{174}$

2* ☐ $\sqrt{30\left(1.2^2\cdot\frac{4}{5}+4.8^2\cdot\frac{1}{5}\right)}=\sqrt{172.8}$

3 ☐ $30\sqrt{\frac{4}{5}+5^2\cdot\frac{1}{5}}=30\sqrt{5.8}$

4 ☐ $30\sqrt{\frac{4}{5^2}}=30\sqrt{0.16}$

5 ☐ $\sqrt{30\cdot\frac{4}{5^2}}=\sqrt{4.8}$

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The X_i be random variable denoting the point for one question. The expected value X_i is

$$E[X_i]=-1\cdot\frac{4}{5}+5\cdot\frac{1}{5}=\frac{1}{5} \tag{2}$$

and the variance of X_i is

$$V[X_i]=\left(-1-\frac{1}{5}\right)^2\frac{4}{5}+\left(5-\frac{1}{5}\right)^2\frac{1}{5}=1.2^2\frac{4}{5}+4.8^2\frac{1}{5} \tag{3}$$

and hence the variance of the total number of points (Y) is

$$V[Y]=30\left(1.2^2\frac{4}{5}+4.8^2\frac{1}{5}\right) \tag{4}$$

taking the square root we get answer 2.

---------- FACIT-END ----------

Continue on page 4

Exercise II

A researcher is planning to study the effect of a new drug that reduces the number of recovery days needed after a certain procedure. In a previous pilot study, it was found that patients who were given the drug needed, on average, 19 days to recover from the procedure. This result seems promising, as patients who are not given the drug need, on average, 23 days to recover. The estimated standard deviation of the number of recovery days is 12 days, and it is assumed that this standard deviation applies to both patients who are given the drug and those who are not.

Since the previous pilot study was based on only a small number of patients, the researcher plans to conduct a new study with a larger sample. The researcher will measure the number of recovery days in a sample of n patients and estimate the average number of recovery days, and it is assumed that the data are normally distributed. The researcher will then test whether this average is significantly different from 23 days (the null hypothesis). A significance level of $\alpha = 0.05$ will be used, and the researcher wants the study to have a statistical power of 70% ($1 - \beta$=0.70) under an alternative hypothesis where the mean and variance equal the values observed in the pilot study.

To answer the following question, we provide the following quantiles from a standard normal distribution:

$z_{0.50}$	$z_{0.60}$	$z_{0.70}$	$z_{0.80}$	$z_{0.90}$	$z_{0.95}$	$z_{0.975}$	$z_{0.99}$	$z_{0.995}$
0.00	0.25	0.52	0.84	1.28	1.64	1.96	2.33	2.58

Question II.1 (3)

How large should the sample size n be?

1 ☐ 5

2 ☐ 42

3 ☐ 44

4* ☐ 56

5 ☐ 71

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

Method 3.65

$$n = \left(\frac{\sigma\left(z_{1-\beta} + z_{1-\alpha/2}\right)}{\mu_0 - \mu_1} \right)^2 = \left(\frac{12(0.52 + 1.96)}{4} \right)^2 = 55.4$$

We must round up, so the answer is 56.

---------- FACIT-END ----------

Continue on page 6

Question II.2 (4)

Which of the following statements is FALSE?

1 ☐ The risk of making a Type I error in the new study is 5%

2* ☐ The risk of making a Type II error in the new study is 70%

3 ☐ The significance level α is the same as the risk of making a Type I error

4 ☐ The statistical power is the same as *the probability of correctly rejecting the null hypothesis if the null hypothesis is false.*

5 ☐ If the true effect of the drug is smaller than hypothesized, then the statistical power of the study will also be smaller.

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The risk of making a Type II error in the new study is 30% ($\beta = 0.30$), hence option 2 is false. Lets just go through the other options:

1: This is the definition of the significance level and hence it is correct

3: Again the definition of significance level (just phrased slightly different)

4: This is the definition of power, hence it is also true

5: The larger the true effect the larger the power, and of course the other way around, hance this is also correct

and hence the correct answer is option 2.

---------- FACIT-END ----------

Continue on page 7

Exercise III

In a study, a university wants to investigate whether the level of Generative AI declaration in student reports is associated with the type of academic project.

After introducing a compulsory library course on responsible use of Generative AI, students are required to explicitly state in the methods chapter whether they used Generative AI tools (such as ChatGPT, Claude, or Copilot) as part of their project or for writing their reports.

The level of AI declaration in the report is classified as: (1) Explicit declaration (clearly documented use), (2) Partial declaration (mentioned "what" but not "how"), and (3) No declaration.

The project type is classified as: (1) Bachelor project, (2) Master thesis, and (3) Other project report.

In a random sample of 240 student reports, the results are summarized in the table below:

	AI Declaration Level:			
Project Type:	Explicit	Partial	None	Total
Bachelor project	40	20	20	80
Master thesis	50	30	20	100
Other report	10	20	30	60
Total	100	70	70	240

The researcher wants to test the hypothesis of independence between Generative AI declarations and project type. The desired significance level α has been stored in Python in a variable called `alpha`.

Question III.1 (5)

Which of the following Python code lines correctly calculates the critical value for the relevant hypothesis test?

1 ☐ `critical_value = stats.chi2.ppf(1 - alpha/2, df = 239)`

2* ☐ `critical_value = stats.chi2.ppf(1 - alpha, df=4)`

3 ☐ `critical_value = stats.f.ppf(1 - alpha, dfn=2, dfd=4)`

4 ☐ `critical_value = stats.f.cdf(1 - alpha, dfn=2, dfd=6)`

5 ☐ `critical_value = stats.t.cdf(1 - alpha/2, df=8)`

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

This problem involves two categorical variables with three categories each:

- Project type (Bachelor, Master, Other)
- AI declaration level (Explicit, Partial, None)

Thus the data forms a 3×3 contingency table.

We therefore apply a Chi-square test of independence.

Degrees of freedom:

$$df = (r-1)(c-1) = (3-1)(3-1) = 4$$

The hypothesis test compares the observed test statistic with the critical value:

$$\chi^2_{1-\alpha,df}$$

In Python, the critical value is obtained using the inverse cumulative distribution function (percent point function) of the chi-square distribution:

```
stats.chi2.ppf(1 - alpha, df=4)
```

---------- FACIT-END ----------

Question III.2 (6)

Continuing with the contingency table on Generative AI declaration in student reports (Bachelor projects, Master theses, and other reports), a researcher now wants to perform a test of independence between generative AI declaration and student report type, using Python.

The observed counts from the study are stored in Python as the following array:

```
observed = np.array([[40, 20, 20],
                     [50, 30, 20],
                     [10, 20, 30]])
```

Which of the following Python code lines correctly performs the test for this contingency table?

1 ☐ `stats.chi2.cdf(observed)`

2 ☐ `stats.chi2.ppf(observed)`

3* ☐ `stats.chi2_contingency(observed)`

4 ☐ `stats.ttest_ind(observed)`

5 ☐ `stats.chi2.pdf(observed)`

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The correct Python function for performing a χ^2-test of independence on a contingency table is:

`stats.chi2_contingency()`

This function takes the contingency table (observed frequencies) as input and returns:

- the Chi-square test statistic
- the p-value
- the degrees of freedom
- the expected frequencies

Example usage:

```
chi2, p, dof, expected = stats.chi2_contingency(observed)
```

This test evaluates whether the two categorical variables (here: project type and AI declaration level) are statistically independent.

---------- FACIT-END ----------

Continue on page 10

Exercise IV

A researcher studies daily online news consumption (measured in minutes) among 20 users. The ordered data are:

$$12, 18, 22, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 90, 100, 110, 125, 140$$

Question IV.1 (7)

What is the value of P_{90}, the 90th percentile?

1 ☐ 125

2* ☐ 117.5

3 ☐ 140

4 ☐ 110

5 ☐ 100

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

$np = 0.90 \cdot 20 = 18$

P_{90} is between the 18th and 19th values:

$(110 + 125)/2 = 117.5$

---------- FACIT-END ----------

Question IV.2 (8)

The researcher extends the dataset by including one additional observation: 165 minutes. The researcher defines outliers (extreme observations) as values that are either 1.5 × IQR above the third quartile (Q3) or 1.5 × IQR below the first quartile (Q1). According to the researcher's definition, what is the upper threshold for outliers, and is the new observation of 165 minutes an outlier?

1 ☐ The upper threshold for outliers is 157.5, and therefore the observation of 165 minutes is an outlier

2 ☐ The upper threshold for outliers is 162.5, and therefore the observation of 165 minutes is an outlier

3 ☐ The upper threshold for outliers is 162.5, and therefore the observation of 165 minutes is not an outlier

4* ☐ The upper threshold for outliers is 172.5, and therefore the observation of 165 minutes is not an outlier

5 ☐ The upper threshold for outliers is 157.5, and therefore the observation of 165 minutes is not an outlier

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

$np = 0.25 \cdot 21 = 5.25$ So, the number in 6th position, $Q_1 = 35$ $np = 0.75 \cdot 21 = 15.75$ So, the number in 16th position, $Q_3 = 90$

$IQR = Q_3 - Q_1 = 90 - 35 = 55$

Upper limit $= Q_3 + 1.5 \cdot IQR = 90 + 1.5 \cdot 55 = 172.5$

Since $165 < 172.5$, 165 is not an extreme observation.

---------- FACIT-END ----------

Continue on page 12

Exercise V

The following sample is available:

```
x = [-2.71, -1.85, -1.12, -0.64, 0.17, 0.43, 0.88, 1.05, 1.34]
```

Question V.1 (9)

What is the median of the sample?

1 ☐ -0.24

2 ☐ -0.64

3 ☐ 0.30

4* ☐ 0.17

5 ☐ 0.43

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The sample is already sorted in ascending order:

$$-2.71,\ -1.85,\ -1.12,\ -0.64,\ 0.17,\ 0.43,\ 0.88,\ 1.05,\ 1.34$$

The sample size is

$np = 9(0.5) = 4.5$

So, the median is the 5th value in the ordered sample:

In Python, this can be checked using:

```
import numpy as np
x = [-2.71, -1.85, -1.12, -0.64, 0.17, 0.43, 0.88, 1.05, 1.34]
np.sort(x)
np.median(x)
np.percentile(x, 50, method='averaged_inverted_cdf')
```

---------- FACIT-END ----------

Question V.2 (10)

Let the i'th observation of the sample be represented by X_i and assume $X_i \sim N(\mu, \sigma^2)$ where the observations are independent and identically distributed (i.i.d.)

What is the distribution of the sample mean $\bar{X}$, given the sample size $n = 10$?

1 ☐ The t-distribution with 8 degrees of freedom

2 ☐ The t-distribution with 9 degrees of freedom

3* ☐ The normal distribution $N(\mu, \sigma^2/n)$

4 ☐ The χ^2-distribution with 8 degrees of freedom

5 ☐ The χ^2-distribution with 9 degrees of freedom

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

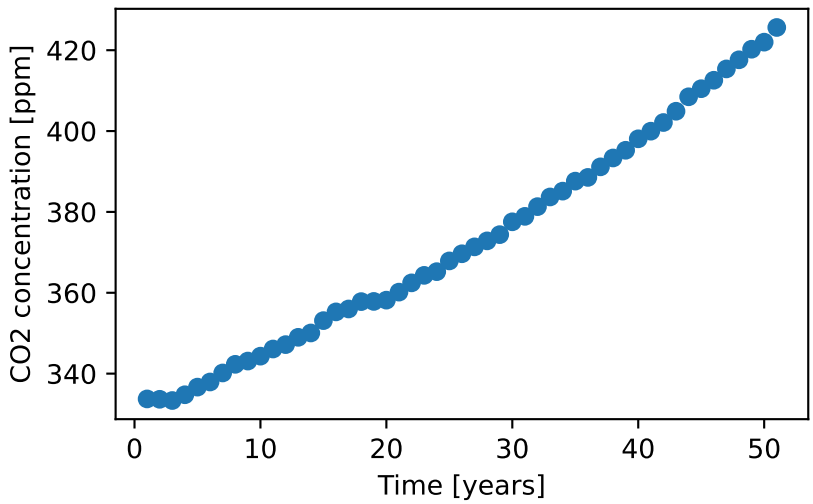
We have Theorem 3.3 The distribution of the mean of normal random variables.

---------- FACIT-END ----------

Continue on page 14

Exercise VI

The plot below shows average yearly CO_2-concentrations as a function of time since 1973 (i.e. the last observation is 2024) at a specific measurement station.



Question VI.1 (11)

Let r denote the empirical correlation coefficient between "Time" and CO_2-concentration. Which of the following statements about r could be correct?

1 ☐ $r \approx 0.8$

2 ☐ $r \approx 0$

3* ☐ $r \approx 0.99$

4 ☐ $r \approx -0.5$

5 ☐ $r \approx 0.5$

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The points fall almost on a straigth line with a positive slope. So imidiately answer 2 and 4 are out. Comparing to the figure in chapter 1.4 it is also clear that the correlation is stronger than 0.8 (actually it is clear that it is above 0.95), and hence the corrrect answer is 3.

---------- FACIT-END ----------

In order to explore the development in CO_2-concentration as a function of time, the following analysis have been performed in Python (`value` denotes the CO_2-concentration, and `co2` is a data-frame containing the data shown above). Some values have been replaced by letters.

Continue on page 16

```
fit1 = smf.ols(formula = "value ~ time", data = co2).fit()
print(fit1.summary(slim=True))
                            OLS Regression Results
==============================================================================
Dep. Variable:                  value   R-squared:                       R2
Model:                            OLS   Adj. R-squared:                  R2a
No. Observations:                  51   F-statistic:                     4128.
Covariance Type:            nonrobust   Prob (F-statistic):           5.68e-49
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    325.14        0.86           t1       p1         l1            u1
time           1.84        0.03           t2       p2         l2            u2
==============================================================================

print(np.sqrt(fit1.scale))
3.00
```

Question VI.2 (12)

What is the approximate values of `l2` and `u2`?

1 ☐ `l2`≈ 1.42, `u2`≈ 2.26

2* ☐ `l2`≈ 1.78, `u2`≈ 1.90

3 ☐ `l2`≈ 1.00, `u2`≈ 2.68

4 ☐ `l2`≈ 1.83, `u2`≈ 1.85

5 ☐ `l2`≈ 0.16, `u2`≈ 3.52

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The numbers `l2` and `u2` makes the 95% confidence interval for the parameter, i.e. it is defined as

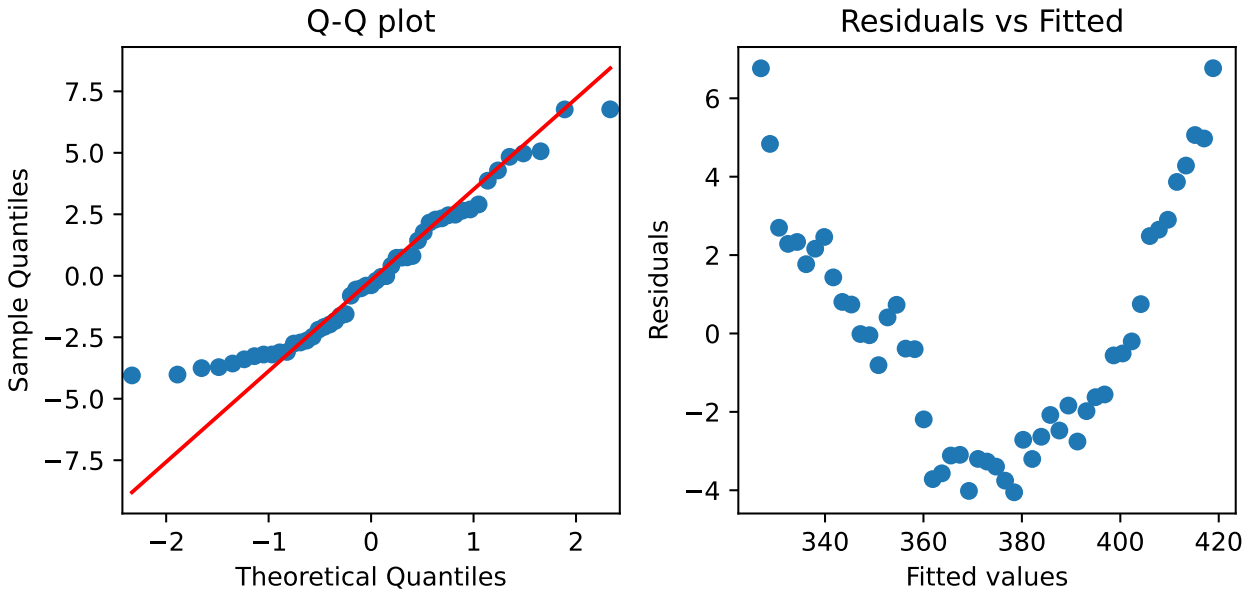
$$[\texttt{l2}, \texttt{u2}] = \hat{\beta}_1 \pm t_{0.975}(49) se_{\beta_1} \tag{5}$$

the 0.975 quantile in the t-distribution is around 2 (it is 1.96 for the normal and slightly higher for the specific t-distribution). Hence

$$[\texttt{l2}, \texttt{u2}] \approx 1.84 \pm 2 \cdot 0.03 = [1.78, 1.90]. \tag{6}$$

---------- FACIT-END ----------

In order to validate the model assumptions the diagnostic plots (based on the model presented above) below have been created.



Question VI.3 (13)

Based on the diagnostic plots, which of the following statements is correct (all parts of the answer must be correct)?

1 ☐ The independence assumption is clearly violated (Q-Q plot), and the variance of the residuals depends on the level (Residuals vs. Fitted).

2 ☐ A data transformation should be considered (Q-Q plot), and "Residual vs. Fitted" suggest a log-transformation.

3 ☐ The "Q-Q plot" suggests that the normality assumption is fulfilled (most dots are on the straight line), further it seems that the correlation between fitted values and residuals is close to zero, and hence all assumptions seems to be fulfilled.

4 ☐ The "Q-Q plot" suggest that systematic effects are missing in the model, and further "Residuals vs. fitted" suggest that the residual variance is a function of fitted values.

5* ☐ The normality assumption seems to be violated (Q-Q plot), and there are clearly systematic effects which are not included in the model (Residual vs. Fitted).

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The QQ-plot assess the normality assumption (which is violated in this case), and Residuals vs. Fitted show clear systematic effects (that should be modeled) in the residuals, this is answer 5. We berifly go through the other answwering options:

1) The independent assumption is not checked by the QQ-plot

2) Resiudal vs. fitted does not suggest a log-transformation.

3) The assumption are definitely not fulfilled

4) The QQ-plot does not suggest systematic effects.

---------- FACIT-END ----------

Continue on page 19

Exercise VII

In this exercise, we consider the quarterly m^2-prices for apartments in Copenhagen from the first quarter of 2010 to the fourth quarter of 2025. Some rows of the dataset are given below:

Observation	Quarter	Price (DKK)	t
1	Q1 2010	22,735	0
2	Q2 2010	23,654	1
⋮	⋮	⋮	⋮
63	Q3 2025	64,719	62
64	Q4 2025	69,192	63

The variable t indicates the number of quarters since Q1 2010 and will serve as the independent variable in the simple linear regression model:

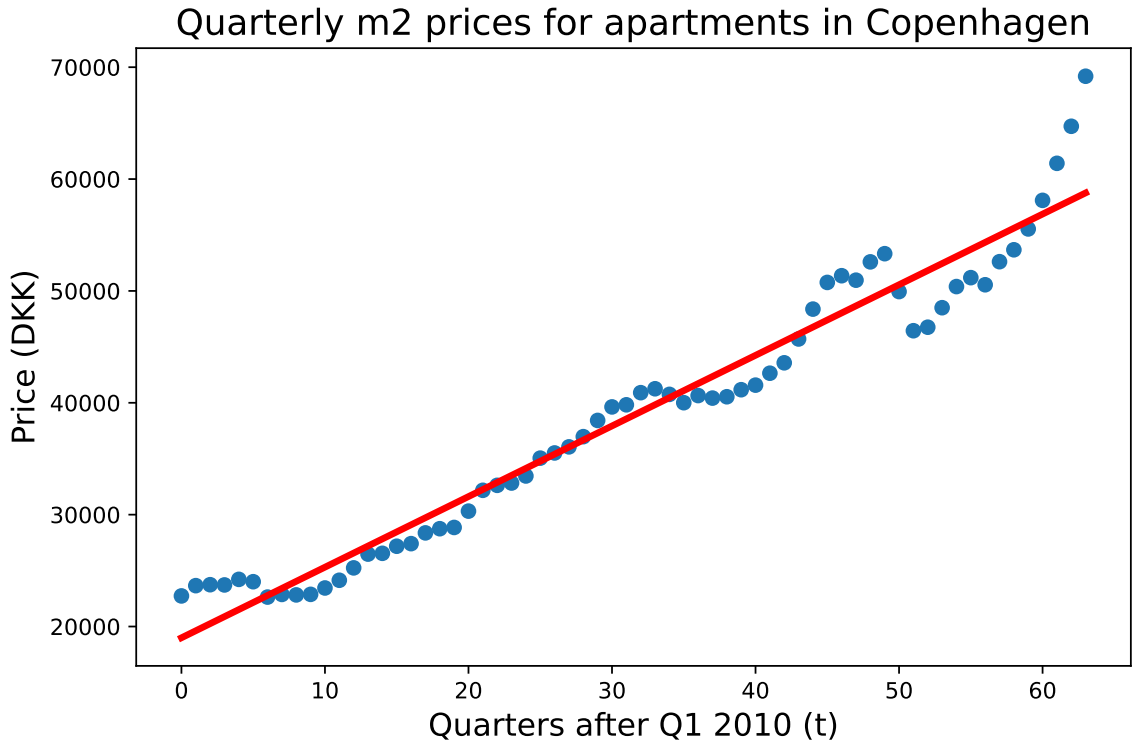
$$Y_i = \beta_0 + \beta_1 t_i + \varepsilon_i,$$

where Y_i is the price for observation i, and the errors $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ are independent and identically distributed.

The linear regression is fitted in Python using the least squares method, and the output from Python is shown below:

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                price_m2   R-squared:                       0.950
Model:                            OLS   Adj. R-squared:                  0.949
No. Observations:                  64   F-statistic:                     1169.
Covariance Type:            nonrobust   Prob (F-statistic):           5.98e-42
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   1.899e+04    674.150     28.170      0.000    1.76e+04    2.03e+04
t            631.3005     18.461     34.196      0.000     594.397     668.204
==============================================================================
```

The fitted regression line is depicted along with the data points in the below plot.



Question VII.1 (14)

What is the observed test statistic for the test of the null hypothesis $\mathcal{H}_0 : \beta_1 = 600$?

1 ☐ $t_{\text{obs}} = 0.950$

2* ☐ $t_{\text{obs}} = 1.695$

3 ☐ $t_{\text{obs}} = 4.017$

4 ☐ $t_{\text{obs}} = 28.170$

5 ☐ $t_{\text{obs}} = 34.196$

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

Theorem 5.12, equation (5-46) applies and yields

$$t_{\text{obs}} = \frac{\hat{\beta}_1 - \beta_{0,1}}{\hat{\sigma}_{\beta_1}} = \frac{631.3005 - 600}{18.461} = 1.695.$$

Here, $\hat{\beta}_1$ and $\hat{\sigma}_{\beta_1}$ are found in the two first columns of the row for t in the Python output, while $\beta_{0,1}$ is from the null hypothesis.

------------------------------------ FACIT-END ------------------------------------

Continue on page 22

Question VII.2 (15)

Which statement is inconsistent with the plot of the fitted regression model?

1* ☐ The observation in Q1 2026 will most likely result in a large negative residual.

2 ☐ The residual variance could vary over time (t).

3 ☐ The residuals could depend on time (t).

4 ☐ The sample mean of the residuals could be zero.

5 ☐ Some residuals could be dependent on each other.

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

According to the plot, the actual prices are rising much faster in 2025 than the model predicts. Therefore, unless the market crashes completely, the observed price in Q1 2026 will be considerably higher than the model prediction, leading to a large positive residual. Hence, option 1 seems highly inconsistent with the fitted model.

First, the sample mean of the residuals is zero by construction, so answer option 4 is consistent with the fitted model. A comparison of the actual prices with the fitted prices (i.e. the residuals) reveals some patterns. Before $t = 30$, the residuals are rather stable, indicating a low residual variance, but the model systematically either over or underestimates the prices. This suggests that the residuals could be dependent on each other or on the time (t). Therefore, answer options 3 and 5 are not inconsistent with the fitted model. After $t = 30$, the prices begin to fluctuate more and more rapid price changes are observed, leading to increasing residual variance. In conclusion, it does not seem inconsistent with the fitted model that the residual variance could vary over time (option 2).

---------- FACIT-END ----------

Question VII.3 (16)

Assuming the model assumptions are satisfied, which of the following statements is correct?

1 ☐ The model explains 95% of the variation in the quarterly prices, and the correlation between price and time (t) is 0.95.

2 ☐ The model explains 97.5% of the variation in the quarterly prices, and the correlation between price and time (t) is 0.95.

3* ☐ The model explains 95% of the variation in the quarterly prices, and the correlation between price and time (t) is 0.975.

4 ☐ The model explains 97.5% of the variation in the quarterly prices, and the correlation between price and time (t) is 0.975.

5 ☐ The model explains 95% of the variation in the quarterly prices, but the correlation between price and time (t) cannot be determined without further information.

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The Python output shows an R^2-value of 0.950. According to definition 5.25, this indicates that the model (the fitted regression line) explains 95% of the variation in the quarterly prices. Furthermore, the sample correlation coefficient between price and time, $\hat{\rho}$, satisfy that $\hat{\rho}^2 = R^2$, cf. the bottom of page 246. Thus, the sample correlation coefficient can be derived as

$$\hat{\rho} = \text{sign}(\hat{\beta}_1)\sqrt{R^2} = \sqrt{0.950} = 0.975.$$

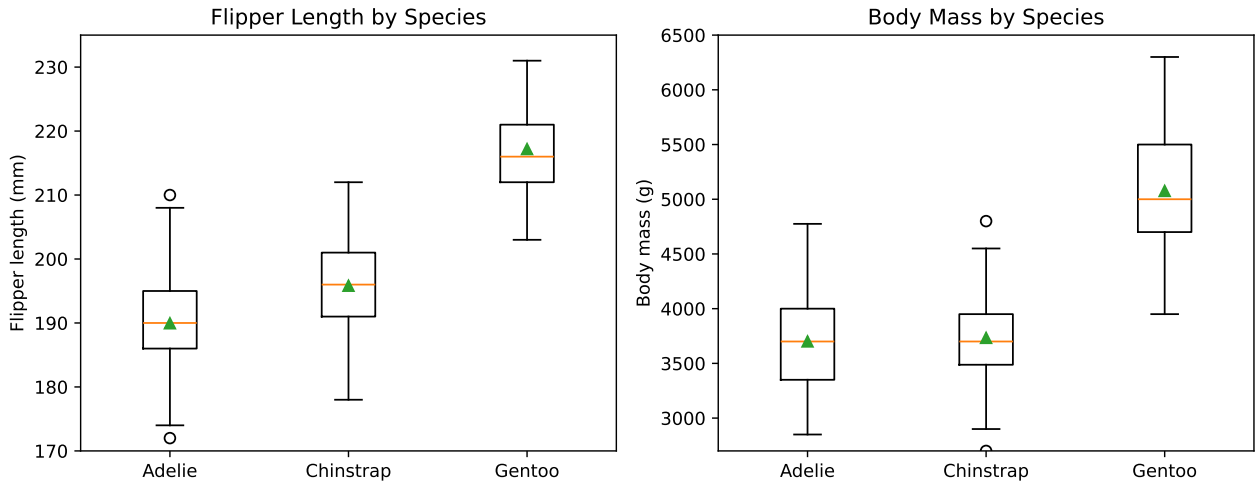
---------- FACIT-END ----------

Continue on page 24

Exercise VIII

We consider data on penguins from three different species: `Adelie`, `Chinstrap`, and `Gentoo`. The data are stored in a Pandas DataFrame named `penguins` and include the variables: `species`, `flipper_length_mm`, and `body_mass_g`.

The data are visualized in the plots below:



Summary statistics for each species are:

		Flipper length (mm)		Body mass (g)	
Species	n	Mean	Std. dev.	Mean	Std. dev.
Adelie	151	189.95	6.54	3700.7	458.6
Chinstrap	68	195.82	7.13	3733.1	384.3
Gentoo	123	217.19	6.48	5076.0	504.1

Assume that both flipper length and body mass within each species is approximately normally distributed.

To answer the following questions, you may need one or more of the following quantiles ($t_{1-\alpha,\nu}$ is the $1-\alpha$ quantile in a t-distribution with ν degrees of freedom):

Degrees of freedom	$t_{0.95,\nu}$	$t_{0.975,\nu}$	$t_{0.995,\nu}$
$\nu = 67$	1.668	1.996	2.651
$\nu = 120$	1.658	1.980	2.617
$\nu = 122$	1.657	1.980	2.617
$\nu = 150$	1.655	1.976	2.609
$\nu = 200$	1.653	1.972	2.601
$\nu = 250$	1.651	1.969	2.596
$\nu = 300$	1.650	1.968	2.592

Question VIII.1 (17)

What is the usual 95% confidence interval for the mean flipper length of Chinstrap penguins?

1 ☐ [195.7; 195.9]

2 ☐ [176.9; 214.7]

3 ☐ [183.9; 207.7]

4* ☐ [194.1; 197.5]

5 ☐ [181.5; 210.1]

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

Sample statistics for Chinstrap penguins:

$$\bar{x} = 195.82, \quad s = 7.13, \quad n = 68$$

We use the t-quantile $t_{0.975,\nu=67} = 1.996$.

The upper and lower limits of the 95% confidence interval are calculated as:

$195.82 \pm 1.996 \cdot 7.13/\sqrt{68}$

---------- FACIT-END ----------

An older source reports that the mean flipper length of Chinstrap penguins is $\mu_0 = 190$ mm. Use the sample data from the current study to test

$$H_0 : \mu = 190 \quad \text{vs} \quad H_1 : \mu \neq 190.$$

Question VIII.2 (18)

At significance level $\alpha = 0.05$, what are the correct test statistic and conclusion of the test?

1 ☐ The test statistic becomes $t \approx 0.816$, and the conclusion is that the observed mean flipper length in the Chinstrap penguins is NOT significantly different from 190 mm (since 0.816 is smaller than 1.996).

2 ☐ The test statistic becomes $t \approx 0.816$, and the conclusion is that the observed mean flipper length in the Chinstrap penguins IS significantly different from 190 mm (since 0.816 is larger than 0.05).

3* ☐ The test statistic becomes $t \approx 6.73$, and the conclusion is that the observed mean flipper length in the Chinstrap penguins IS significantly different from 190 mm (since 6.73 is larger than 1.996).

4 ☐ The test statistic becomes $t \approx 6.73$, and the conclusion is that the observed mean flipper length in the Chinstrap penguins is NOT significantly different from 190 mm (since 6.73% is larger than 5%).

5 ☐ The test statistic becomes $t \approx 5.45$, and the conclusion is that the observed mean flipper length in the Chinstrap penguins is NOT significantly different from 190 mm (since 5.45% is larger than 5%).

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

Sample statistics for Chinstrap penguins:

$$\bar{x} = 195.82, \quad s = 7.13, \quad n = 68$$

Test statistic:

$$t = (195.82 - 190)/(7.13/\sqrt{68}) = 6.73$$

The test statistic should be compared to the critical value 1.996.

---------- FACIT-END ----------

We wish to investigate whether there is a significant difference in mean body mass between **Adelie** and **Chinstrap** penguins, which corresponds to testing

$$H_0 : \mu_A = \mu_C \qquad \text{vs} \qquad H_1 : \mu_A \neq \mu_C.$$

We will use a significance level of $\alpha = 0.05$ and apply a two-sample t-test that assumes equal variances in the two groups. For this purpose, you may use the pooled two-sample estimate of the variance (estimated from the Adelie and Chinstrap samples):

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_C - 1)s_C^2}{n_A + n_C - 2} = 190977 = 437^2.$$

Question VIII.3 (19)

Which of the following statements is TRUE?

1 ☐ The estimated difference $\hat{\mu}_A - \hat{\mu}_C$ is -32.4 grams, which is a significant difference, since it is smaller than 0.05% of the average body mass of both Adelie and Chinstrap penguins.

2 ☐ The relevant test statistic is approximately $t = -0.07$, and the final conclusion is that there is a significant difference in body mass between the Adelie and the Chinstrap species, since $|-0.07| = 0.07$ is greater than 0.05.

3* ☐ The relevant test statistic is approximately $t = -0.5$, and the final conclusion is that we do NOT find a statistically significant difference in mean body mass between the Adelie and the Chinstrap species.

4 ☐ The relevant test statistic is approximately $t = -0.5$, but given the information in the exercise, we are not able to determine if the difference is significant or not.

5 ☐ The relevant test statistic is approximately $t = -0.07$, and the final conclusion is that we do NOT find a significant difference in body mass between the Adelie and the Chinstrap species, since $|-0.07| = 0.07$ is greater than 0.05.

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

For **Adelie**:

$$n_A = 151, \quad s_A = 458.6, \quad \bar{x}_A = 3700.7$$

For **Chinstrap**:

$$n_C = 68, \quad s_C = 384.3, \quad \bar{x}_C = 3733.1$$

And we already have:

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_C - 1)s_C^2}{n_A + n_C - 2} = 190977 = 437^2$$

$$s_p = 437$$

Therefore:

$$t_{pooled} = \frac{\bar{x}_A - \bar{x}_C}{\sqrt{s_p^2 \left(\frac{1}{n_A} + \frac{1}{n_C}\right)}} = \frac{3700.7 - 3733.1}{\sqrt{437^2 \left(\frac{1}{68} + \frac{1}{151}\right)}} \approx -0.5$$

Also, the degrees of freedom are:

$$n_A + n_C - 2 = 217$$

From the quantiles given in the assignment, we do not see the exact quantile we need (which is $t_{0.975,\nu=217}$), but we can see that the critical value will be between $1.980 (t_{0.975,\nu=120})$ and 1.996 $(t_{0.975,\nu=67})$. Therefore, the value $|-0.5| = 0.5$ is so small that we will not be able to reject the null hypothesis.

---------- FACIT-END ----------

Continue on page 29

Question VIII.4 (20)

Which statement is TRUE regarding the pooled two-sample t-test?

1* ☐ The method assumes that the two populations have equal variances.

2 ☐ The method requires equal sample sizes in the two groups.

3 ☐ The method is inappropriate because the sample sizes are larger than 30.

4 ☐ The method allows the two populations to have different variances and adjusts the degrees of freedom accordingly.

5 ☐ The method tests whether the two sample variances are equal.

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The pooled method assumes the samples come from population with equal variances.

---------- FACIT-END ----------

Continue on page 30

Exercise IX

A researcher models daily news consumption (in minutes) using a normal distribution with mean 90 and standard deviation 10, based on some data.

Question IX.1 (21)

Which of the following one-line Python commands correctly generates random samples from this distribution?

1 ☐ `scipy.stats.norm.cdf(90, loc=90, scale=10)`

2 ☐ `scipy.stats.norm.pdf(90, loc=90, scale=10)`

3 ☐ `scipy.stats.norm.ppf(0.90, loc=90, scale=10)`

4 ☐ `scipy.stats.norm.pmf(90, loc=90, scale=10)`

5* ☐ `scipy.stats.norm.rvs(loc=90, scale=10, size=10)`

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The correct function is `rvs` (random variates), which generates random numbers from a specified probability distribution.

Example:

```
import scipy.stats as stats
stats.norm.rvs(loc=90, scale=10, size=10)
```

Here:

- `loc = 90` specifies the mean daily news consumption.
- `scale = 10` specifies the standard deviation.
- `size = 10` generates 10 simulated users.

Other functions:

- `pdf` – density (not random generation)

- `cdf` – cumulative probability
- `ppf` – quantiles (percentiles)

---------- FACIT-END ----------

Continue on page 32

Exercise X

A transportation researcher is studying whether the adoption rate of electric cars (EVs) increased after a government introduced incentives such as free municipal parking and expanded charging station infrastructure.

In a small pilot survey conducted before the incentives, it was found that about 15% of households owned an electric car. Since the pilot study was relatively small ($n = 20$ households), the estimate of 15% has considerable uncertainty.

To plan a larger follow-up study, the researcher wants to estimate the true proportion of households owning an electric car after the incentives. The researcher wishes to construct a 95% confidence interval with a *Margin of Error* (ME) of only 6%.

Question X.1 (22)

What minimum sample size n is required for the follow-up study, assuming the true proportion is equal the proportion found in the pilot study?

1 ☐ $n = 120$

2* ☐ $n = 137$

3 ☐ $n = 210$

4 ☐ $n = 260$

5 ☐ $n = 310$

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The margin of error is

$$ME = z_{1-\alpha/2}\sqrt{\frac{p(1-p)}{n}}$$

Solving for n:

$$n = p(1-p)\left(\frac{z_{1-\alpha/2}}{ME}\right)^2$$

Insert the known values:

$$n = 0.15(1 - 0.15)\left(\frac{1.96}{0.06}\right)^2 = 0.1275(32.67)^2 = 0.1275 \times 1067.11 = 136.1$$

Thus, the researcher should plan for a sample size of 137 households.

---------- FACIT-END ----------

The researcher later conducts a survey of 300 households. With the result

- Before the incentives: 36 out of 150 households owned an electric car.
- After the incentives: 51 out of 150 households owned an electric car.

The researcher wants to test whether the proportion of EV ownership changed after the incentives using a significance level $\alpha = 0.05$.

Continue on page 34

Question X.2 (23)

Which of the following conclusions is correct?

1* ☐ $z_{obs} = 1.91 < 1.96$, so we fail to reject H_0 and CANNOT conclude that EV adoption increased significantly.

2 ☐ $z_{obs} = 1.45 < 1.96$, so we fail to reject H_0 and conclude that EV adoption did not increase significantly.

3 ☐ $z_{obs} = 0.87 < 1.96$, so we fail to reject H_0 and conclude that EV adoption did not increase significantly.

4 ☐ $z_{obs} = 1.91 < 1.96$, so we fail to reject H_0 and conclude that EV adoption increased significantly.

5 ☐ $z_{obs} = 3.02 > 1.96$, so we reject H_0 and conclude that EV adoption increased significantly.

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

We estimate the proportions:

$$\hat{p}_1 = \frac{36}{150} = 0.24$$

$$\hat{p}_2 = \frac{51}{150} = 0.34$$

Under $H_0 : p_1 = p_2 = p$ the pooled estimate is

$$\hat{p} = \frac{36 + 51}{300} = \frac{87}{300} = 0.29$$

The test statistic is

$$z_{obs} = \frac{\hat{p}_2 - \hat{p}_1}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{150} + \frac{1}{150}\right)}} = \frac{0.34 - 0.24}{\sqrt{0.29(0.71)(0.0133)}} \frac{0.10}{0.0463} = \frac{0.10}{0.0524} = 1.91$$

Since

$$|z_{obs}| = 1.91 < 1.96$$

we **fail to reject the null hypothesis** at the significance level $\alpha = 0.05$.

Therefore, there is not sufficient statistical evidence to conclude that the proportion of households owning an electric car changed after the incentives.

---------- FACIT-END ----------

Continue on page 36

Exercise XI

A researcher assesses the toxicity of various drugs against a cardiovascular disease using a one-way ANOVA model, which yields the following entries in the ANOVA table:

Source	DF	SS
Drug	3	33.66
Residual	36	220.20
Total	39	253.86

Question XI.1 (24)

Given the model assumptions are satisfied, what is the observed F-statistic for the null hypothesis that the mean toxicity (mean effect) is the same for all the drugs against the usual alternative hypothesis?

1 ☐ $F_{\text{obs}} = 11.22$

2 ☐ $F_{\text{obs}} = 6.12$

3* ☐ $F_{\text{obs}} = 1.83$

4 ☐ $F_{\text{obs}} = 0.55$

5 ☐ $F_{\text{obs}} = 0.15$

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

Theorem 8.6 states that the F-statistic is

$$F_{\text{obs}} = \frac{SS(Tr)/(k-1)}{SSE/(n-k)} = \frac{33.66/3}{220.20/36} = 1.83.$$

The values in the denominators are the degrees of freedom for the treatments and the residuals, and these are found in the ANOVA table under DF.

---------- FACIT-END ----------

Question XI.2 (25)

The observed F-statistic should be compared to a quantile from which distribution when testing the null hypothesis that the mean toxicity (mean effect) is the same for all the drugs against the usual alternative hypothesis?

1* ☐ An F-distribution with $\nu_1 = 3$ and $\nu_2 = 36$ degrees of freedom.

2 ☐ An F-distribution with $\nu_1 = 4$ and $\nu_2 = 39$ degrees of freedom.

3 ☐ An F-distribution with $\nu_1 = 36$ and $\nu_2 = 39$ degrees of freedom.

4 ☐ A t-distribution with $\nu = 36$ degrees of freedom.

5 ☐ A t-distribution with $\nu = 39$ degrees of freedom.

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

Theorem 8.6 states that the F-statistic follows an F-distribution with $\nu_1 = k - 1 = 3$ and $\nu_2 = n - k = 36$. These values are the degrees of freedom for the treatments and the residuals, and they are found in the ANOVA table under DF.

---------- FACIT-END ----------

Question XI.3 (26)

What is the estimated standard deviation of the residuals in the one-way ANOVA model?

1 ☐ 220.20

2 ☐ 14.84

3 ☐ 6.12

4 ☐ 3.35

5* ☐ 2.47

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

In ANOVA models, the residual variance is estimated by MSE, cf. page 314. Hence,

$$\hat{\sigma}^2 = MSE = \frac{SSE}{n-k} = \frac{220.20}{36}.$$

Consequently, the estimated residual standard deviation is given as

$$\hat{\sigma} = \sqrt{\frac{220.20}{36}} = 2.47.$$

---------- FACIT-END ----------

Question XI.4 (27)

The Bonferroni correction is used to correct the significance level (α) when multiple tests are conducted. Which of the following statements about the effect on the risk of Type I and Type II errors is correct for the Bonferroni correction?

1 ☐ The Bonferroni correction increases the Type I risk but decreases the Type II risk of the individual hypothesis tests.

2* ☐ The Bonferroni correction decreases the Type I risk but increases the Type II risk of the individual hypothesis tests.

3 ☐ The Bonferroni correction increases the Type I risk and decreases the power of the individual hypothesis tests.

4 ☐ The Bonferroni correction decreases the Type I risk and increases the power of the individual hypothesis tests.

5 ☐ The Bonferroni correction increases the Type I risk of the individual hypothesis tests but decreases the family-wise Type I risk.

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The idea of applying a Bonferroni correction is to reduce the family-wise Type I error rate by reducing the Type I risk in each of the individual tests. However, for a fixed experimental design, decreasing the Type I risk implies increasing the Type II risk i.e., decreasing the power.

---------- FACIT-END ----------

Continue on page 39

Exercise XII

Let X be a discrete random variable taking values in the non-negative integers. We denote the probability density function (pdf) by f and the cumulative distribution function (cdf) by F. The below table shows some values of F:

x	0	1	2	3	4	5	6	7	8	9	10
$F(x)$	0.05	0.12	0.14	0.24	0.25	0.48	0.68	0.80	0.89	0.97	1.00

Question XII.1 (28)

What is the most probable outcome of X?

1 ☐ The most probable outcome of X is 3.

2 ☐ The most probable outcome of X is 4.

3* ☐ The most probable outcome of X is 5.

4 ☐ The most probable outcome of X is 6.

5 ☐ The most probable outcome of X is 10.

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

The distribution function of a discrete random variable is defined in definition 2.9. Following this definition, we have that for x in the non-negative integers

$$f(x) = F(x) - F(x-1).$$

Therefore, the density function takes the following values: Since f is largest for $x = 5$, the

x	0	1	2	3	4	5	6	7	8	9	10
$f(x)$	0.05	0.07	0.02	0.10	0.01	0.23	0.20	0.12	0.09	0.08	0.03

most probable value is therefore 5.

---------- FACIT-END ----------

Question XII.2 (29)

What is the probability $\mathbb{P}(3 \leq X \leq 8)$?

1 ☐ 0.83

2* ☐ 0.75

3 ☐ 0.73

4 ☐ 0.66

5 ☐ 0.65

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

Using equation (2-13):

$$\mathbb{P}(3 \leq X \leq 8) = \mathbb{P}(2 < X \leq 8) = F(8) - F(2) = 0.89 - 0.14 = 0.75.$$

---------- FACIT-END ----------

Continue on page 41

Question XII.3 (30)

We will define the q-quantile as the smallest value x that satisfies $\mathbb{P}(X \leq x) \geq q$. What is the 15%-quantile in the distribution of X?

1 ☐ The 15%-quantile in the distribution of X is 2.0.

2 ☐ The 15%-quantile in the distribution of X is 2.1.

3 ☐ The 15%-quantile in the distribution of X is 2.5.

4* ☐ The 15%-quantile in the distribution of X is 3.0.

5 ☐ The 15%-quantile in the distribution of X is 8.0.

6 ☐ Don't know / No answer

---------- FACIT-BEGIN ----------

According to the definition in the problem, we are looking for the smallest value x such that $F(x) \geq 0.15$. We first notice that $F(x) \geq 0.15$ for all $x \geq 3$. Since X is a discrete random variable, the distribution function is flat for values not in the image of X. Hence, the smallest value x for which $F(x) \geq 0.15$ is exactly 3.

---------- FACIT-END ----------

The exam is finished.