# i Forside mai 2021

**Department of Mathematical Sciences**

**Examination paper for TMA4245 Statistics**

**Examination date:** Friday May 14, 2021

**Examination time (from-to):** 09.00 – 13.00

**Permitted examination support material:** A / All support material is allowed

**Academic contact during examination:** Geir-Arne Fuglstad and Jarle Tufto
**Phone:** 45270806, 99705519

**Technical support during examination:** Orakel support services
**Phone:** 73 59 16 00

If you experience technical problems during the exam, contact Orakel support services as soon as possible before the examination time expires. If you don't get through immediately, hold the line until your call is answered.

**OTHER INFORMATION**

**Make your own assumptions:** If a question is unclear/vague, make your own assumptions and specify them in your answer. Only contact academic contact in case of errors or insufficiencies in the question set.

**Cheating/Plagiarism:** The exam is an individual, independent work. Examination aids are permitted, but make sure you follow any instructions regarding citations. During the exam it is not permitted to communicate with others about the exam questions, or distribute drafts for solutions. Such communication is regarded as cheating. All submitted answers will be subject to plagiarism control. *Read more about cheating and plagiarism here.*

**Notifications:** If there is a need to send a message to the candidates during the exam (e.g. if there is an error in the question set), this will be done by sending a notification in Inspera. A dialogue box will appear. You can re-read the notification by clicking the bell icon in the top right-hand corner of the screen. All candidates will also receive an SMS to ensure that nobody misses out on important information. Please keep your phone available during the exam.

**Weighting:** There are in total 100 points. Problems 1--6 are automatically marked and count 5 points each. Problems 7a, 7b, 7c, 7d, 8a, 8b og 9a are marked manually and count 10 points each.

**General instructions:** In Problems 1--6 you should only give the correct answer and no justification. Use more significant digits during calculations than in the final answers to avoid errors due to rounding. **Follow the instructions regarding decimal places in the final answers. If not, the answer may be evaluated as wrong.** For Problems 7a, 7b, 7c, 7d, 8a, 8b og 9a all answers must be justified and all natural intermediate calculations must be included. It must be completely clear how you arrived at the final answers.

## ABOUT SUBMISSION

**How to answer questions:** All question types other than Upload assignment must be answered directly in Inspera. In Inspera, your answers are saved automatically every 15 seconds. **NB!** We advise against pasting content from other programs, as this may cause loss of formatting and/or entire elements (e.g. images, tables).

**File upload**: All files must be uploaded before the examination time expires. All uploaded files must be in "pdf"-format. 30 minutes are added to the examination time to manage the sketches/calculations/files. The additional time is included in the remaining examination time shown in the top left-hand corner.

NB! You are responsible to ensure that the file(s) are correct and not corrupt/damaged. Check the file(s) you have uploaded by clicking “Download” when viewing the question. All files can be removed or replaced as long as the test is open.

*How to digitize your sketches/calculations*
*How to create PDF documents*
*Remove personal information from the file(s) you want to upload*

**Automatic submission:** Your answer will be submitted automatically when the examination time expires and the test closes, if you have answered at least one question. This will happen even if you do not click “Submit and return to dashboard” on the last page of the question set. You can reopen and edit your answer as long as the test is open. If no questions are answered by the time the examination time expires, your answer will not be submitted. This is considered as “did not attend the exam”.

**Withdrawing from the exam:** If you become ill, or wish to submit a blank test/withdraw from the exam for another reason, go to the menu in the top right-hand corner and click “Submit blank”. This cannot be undone, even if the test is still open.

**Accessing your answer post-submission:** You will find your answer in Archive when the examination time has expired.

1 # 1A

**Problem 1**

A dataset consists of seven measurements: 4, 6, 5, 3, 5, 1 and 3. Calculate the following quantities. **Use three decimal places in intermediate calculations and round the final answer to one decimal place.**

- Average: ☐.
- Median: ☐.
- Empirical variance: ☐.

Maximum marks: 5

2 # 1B

**Problem 1**

A dataset consists of seven measurements: 4, 5, 3, 4, 3, 2 and 1. Calculate the following quantities. **Use three decimal places in intermediate calculations and round the final answer to one decimal place.**

- Average: ☐.
- Median: ☐.
- Empirical variance: ☐.

Maximum marks: 5

## 3 1C

**Problem 1**

A dataset consists of seven measurements: 2, 6, 4, 6, 5, 3 og 3. Calculate the following quantities. **Use three decimal places in intermediate calculations and round the final answer to one decimal place.**

- Average: [ ].
- Median: [ ].
- Empirical variance: [ ].

Maximum marks: 5

## 4 1D

**Problem 1**

A dataset consists of seven measurements: 2, 2, 1, 5, 4, 2 og 3. Calculate the following quantities. **Use three decimal places in intermediate calculations and round the final answer to one decimal place.**

- Average: [ ].
- Median: [ ].
- Empirical variance: [ ].

Maximum marks: 5

# 5 2A

**Problem 2**

Suppose that $X$ is a standard normally distributed random variable with probability density function $f_X(x)$. Consider the events $A = \{X \leq 3\}$ and $B = \{X \geq 3\}$. Determine if the following statements are true or false.

1. $\mathrm{P}(A \cap B) = f_X(3)$.

**Select one alternative:**

- True
- False

2. $\mathrm{P}(A \cap B) = 0$.

**Select one alternative:**

- True
- False

3. $A$ and $B$ are disjoint events.

**Select one alternative:**

- True
- False

4. $A$ and $B$ are dependent events.

**Select one alternative:**

- True
- False

---

Maximum marks: 5

# 6 2B

**Problem 2**

Suppose that $X$ is a Gamma distributed random variable with probability density function $f_X(x)$. Consider the events $A = \{X \leq 2\}$ and $B = \{X \geq 2\}$. Determine if the following statements are true or false.

1. $\mathrm{P}(A \cap B) = f_X(2)$.

**Select one alternative:**

- True
- False

2. $\mathrm{P}(A \cap B) = 0$.

**Select one alternative:**

- True
- False

3. $A$ and $B$ are disjoint events.

**Select one alternative:**

- True
- False

4. $A$ and $B$ are dependent events.

**Select one alternative:**

- True
- False

---

Maximum marks: 5

# 7 2C

**Problem 2**

Suppose that $X$ is a t-distributed (with 10 degrees of freedom) random variable with probability density function $f_X(x)$. Consider the two events $A = \{X \leq -1\}$ and $B = \{X \geq -1\}$. Determine if the following statements are true or false.

1. $\mathrm{P}(A \cap B) = f_X(-1)$.

**Select one alternative:**

- True
- False

2. $\mathrm{P}(A \cap B) = 0$.

**Select one alternative:**

- True
- False

3. $A$ are $B$ are disjoint events.

**Select one alternative:**

- True
- False

4. $A$ and $B$ are dependent events.

**Select one alternative:**

- True
- False

---

Maximum marks: 5

# 8 2D

**Oppgave 2**

Suppose that $X$ is chi-square distributed with 10 degrees of freedom and probability density function $f_X(x)$. Consider the two events $A = \{X \leq 10\}$ and $B = \{X \geq 10\}$. Determine if the following statements are true or false.

1. $\mathrm{P}(A \cap B) = f_X(10)$.

**Select one alternative:**

- True
- False

2. $\mathrm{P}(A \cap B) = 0$.

**Select one alternative:**

- True
- False

3. $A$ and $B$ are disjoint events.

**Select one alternative:**

- True
- False

4. $A$ and $B$ are dependent events.

**Select one alternative:**

- True
- False

---

Maximum marks: 5

## 9 3A

**Problem 3**

We draw 5 cards at random from a deck of cards (52 cards in total, 13 cards of each suit). Give your answers to each question below as integers.

- What is the number of possible outcomes (without regard to ordering of the cards)?
- In how many ways can we draw 5 cards that are all hearts?

Maximum marks: 5

## 10 3B

**Problem 3**

We draw 5 cards at random from a deck of cards (52 cards in total, 13 cards of each suit). Give your answers to each question below as integers.

- What is the number of possible outcomes (without regard to ordering of the cards)?
- In how many ways can we draw a straight flush (all cards of the same suit and in sequential rank, for example, 8, 9, 10, Jack and Queen all in clubs)? **Note that Ace can be be both the first and last card in a straight flush.**

Maximum marks: 5

11 **3C**

**Problem 3**

We draw 5 cards at random from a deck of cards (52 cards in total, 13 cards of each suit). Give your answers to each question below as integers.

- What is the number of possible outcomes (without regard to ordering of the cards)? [ ]
- In how many ways can we draw 5 cards that are all honour cards (knight, queen, king or ace)? [ ]

Maximum marks: 5

12 **3D**

**Problem 3**

We draw 5 cards at random from a deck of cards (52 cards in total, 13 cards of each suit). Give your answers to each question below as integers.

- What is the number of possible outcomes (without regard to ordering of the cards)? [ ]
- In how many ways can we draw 5 cards out of which four are aces? [ ]

Maximum marks: 5

# 13 4A

**Problem 4**

Let $X$ denote the total number of eyes when we throw a pair of fair dice and let $Z = 10X + 5$. Calculate the following probabilities and give the answer rounded to three decimal places.

- $P(X < 4) =$
- $P(Z \leq 35) =$
- $P(Z = 25) =$
- $P(Z = 25|X < 4) =$

Maximum marks: 5

# 14 4B

**Problem 4**

Let $X$ denote the total number of eyes when we throw a pair of fair dice and let $Z = 5X + 10$. Calculate the following probabilities and give the answer rounded to three decimal places.

- $P(X > 10) =$
- $P(Z \geq 65) =$
- $P(Z = 70) =$
- $P(Z = 70|X > 10) =$

Maximum marks: 5

## 15 4C

**Problem 4**

Let $X$ denote the total number of eyes when we throw a pair of fair dice and let $Z = 2X + 10$ . Calculate the following probabilities and give the answer rounded to three decimal places.

- $P(X \leq 4) =$
- $P(Z < 20) =$
- $P(Z = 14) =$
- $P(Z = 14|X \leq 4) =$

---

Maximum marks: 5

## 16 4D

**Problem 4**

Let $X$ denote the total number of eyes when we throw a pair of fair dice and let $Z = 10X$ . Calculate the following probabilities and give the answer rounded to three decimal places.

- $P(X \leq 4) =$
- $P(Z < 50) =$
- $P(Z = 20) =$
- $P(Z = 20|X \leq 4) =$

---

Maximum marks: 5

## 17 5A

**Problem 5**

Let $X$ be a normally distributed stochastic variable with expected value 4.5 and standard deviation 3, and let $Y$ be a normally distributed stochastic variable with expected value -1.5 and standard deviation 4. Assume that $X$ og $Y$ are independent stochastic variables. Calculate the following probabilities and round the answers to three decimal places.

- $\mathrm{P}(X > 3) =$ ____.
- $\mathrm{P}(X + Y > 3) =$ ____.
- $\mathrm{P}(X + Y > 3 | Y = 3) =$ ____.

---

Maximum marks: 5

## 18 5B

**Problem 5**

Let $X$ be a normally distributed stochastic variable with expected value 2.5 and standard deviation 3, and let $Y$ be a normally distributed stochastic variable with expected value -3 and standard deviation 4. Assume that $X$ og $Y$ are independent stochastic variables. Calculate the following probabilities and round the answers to three decimal places.

- $\mathrm{P}(X > 3) =$ ____.
- $\mathrm{P}(X + Y > 3) =$ ____.
- $\mathrm{P}(X + Y > 3 | Y = 2) =$ ____.

---

Maximum marks: 5

## 19 5C

**Problem 5**

Let $X$ be a normally distributed stochastic variable with expected value 2 and standard deviation 4, and let $Y$ be a normally distributed stochastic variable with expected value 0 and standard deviation 3. Assume that $X$ og $Y$ are independent stochastic variables. Calculate the following probabilities and round the answers to three decimal places.

- $\mathrm{P}(X > 3) =$ ______.
- $\mathrm{P}(X + Y > 3) =$ ______.
- $\mathrm{P}(X + Y > 3 | Y = 1) =$ ______.

---

Maximum marks: 5

## 20 5D

**Problem 5**

Let $X$ be a normally distributed stochastic variable with expected value -1 and standard deviation 4, and let $Y$ be a normally distributed stochastic variable with expected value -1 and standard deviation 3. Assume that $X$ og $Y$ are independent stochastic variables. Calculate the following probabilities and round the answers to three decimal places.

- $\mathrm{P}(X > 3) =$ ______.
- $\mathrm{P}(X + Y > 3) =$ ______.
- $\mathrm{P}(X + Y > 3 | Y = 3) =$ ______.

---

Maximum marks: 5

## 21 6A

**Problem 6**

We consider two stochastic variables $X$ and $Y$. Assume the expected values and variances are known: $\mathrm{E}[X] = 2$, $\mathrm{Var}[X] = 2$, $\mathrm{E}[Y] = -1$ og $\mathrm{Var}[Y] = 1$. Additionally, we know the covariance $\mathrm{Cov}[X, Y] = 0.5$. Calculate the quantities below and round the answers to one decimal place.

- $\mathrm{E}[X + Y] =$ ☐.
- $\mathrm{Var}[X + Y] =$ ☐.

Maximum marks: 5

## 22 6B

**Problem 6**

We consider two stochastic variables $X$ and $Y$. Assume the expected values and variances are known: $\mathrm{E}[X] = -1$, $\mathrm{Var}[X] = 1$, $\mathrm{E}[Y] = 1$ og $\mathrm{Var}[Y] = 1.5$. Additionally, we know the covariance $\mathrm{Cov}[X, Y] = 0.25$. Calculate the quantities below and round the answers to one decimal place.

- $\mathrm{E}[X + Y] =$ ☐.
- $\mathrm{Var}[X + Y] =$ ☐.

Maximum marks: 5

## 23 6C

**Problem 6**

We consider two stochastic variables $X$ and $Y$. Assume the expected values and variances are known: $\mathrm{E}[X] = 2$, $\mathrm{Var}[X] = 2$, $\mathrm{E}[Y] = 1$ and $\mathrm{Var}[Y] = 2$. Additionally, we know the covariance $\mathrm{Cov}[X, Y] = 0.5$. Calculate the quantities below and round the answers to one decimal place.

- $\mathrm{E}[X + Y] =$ ☐.
- $\mathrm{Var}[X + Y] =$ ☐.

---

Maximum marks: 5

## 24 6D

**Problem 6**

We consider two stochastic variables $X$ and $Y$. Assume the expected values and variances are known: $\mathrm{E}[X] = 3$, $\mathrm{Var}[X] = 1.5$, $\mathrm{E}[Y] = -4$ and $\mathrm{Var}[Y] = 1.5$. Additionally, we know the covariance $\mathrm{Cov}[X, Y] = 0.25$. Calculate the quantities below and round the answers to one decimal place.

- $\mathrm{E}[X + Y] =$ ☐.
- $\mathrm{Var}[X + Y] =$ ☐.

---

Maximum marks: 5

# 25 7A

**Problem 7**

Suppose that the income $Y$ of individuals in a population is a continuously distrubuted random variabel with probability density function

$$f_Y(y) = \begin{cases} \frac{\lambda y_0^\lambda}{y^{\lambda+1}}, & \text{for} \quad y \geq y_0, \\ 0, & \text{for} \quad y < y_0, \end{cases}$$

where $\lambda > 0$ is an unknown parameter and $y_0 = 300\,000$ kr is a known minimum income set by government regulations.

**a)**

Assume that $\lambda = 1.0$ in this point.

- Find the cumulative distribution function of income $Y$.
- Use the cumulative distribution function to derive the derive the probability that the lowest income in a random sample of $n = 10$ individuals is smaller than 310 000 kroner.

Suppose we observe a random sample of $n = 10$ incomes, $y_1, y_2, \ldots, y_n$, given by 456000, 578000, 325000, 324000, 382000, 1498000, 594000, 405000, 510000 og 326000 kr. Suppose that the statistic $\sum_{i=1}^{n} \ln(y_i/y_0) = 4.68$.

**b)**

- Derive the maximum likelihood estimator of $\lambda$ for a random sample of size $n$.
- Compute the maximum likelihood estimate of $\lambda$ for the data given above.

**c)**

- Let $X = 2\lambda \ln(Y/y_0)$. Show that $X$ has a chi-square distribution with 2 degrees of freedom. What is the distribution of $\sum_{i=1}^{n} 2\lambda \ln(Y_i/y_0)$? Explain why.
- Derive a 95%-confidence interval for $\lambda$ based on a random sample of size n. Compute the interval for the data given above.

**d)**

The median of $Y$ is given by $\mu^* = 2^{1/\lambda} y_0$ for $\lambda > 0$. We wish to test if the data above provides evidence for the claim that the median income in the population, $\mu^*$, is greater than 600 000 kr.

- Show that this is equivalent to testing the hypothesis $H_0 : \lambda = 1.0$ versus $H_1 : \lambda < 1.0$.
- Construct and carry out the test for the above data using a level of significance of 0.05. What is the conclusion of the test?

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 40

26 **7B**

**Problem 7**

Suppose that the income $Y$ of individuals in a population is a continuously distrubuted random variabel with probability density function

$$f_Y(y) = \begin{cases} \frac{\lambda y_0^\lambda}{y^{\lambda+1}}, & \text{for} \quad y \geq y_0, \\ 0, & \text{for} \quad y < y_0, \end{cases}$$

where $\lambda > 0$ is an unknown parameter and $y_0 = 300\,000$ kr is a known minimum income set by government regulations.

**a)**

Assume that $\lambda = 1.36$ in this point.

- Find the cumulative distribution function of income $Y$.
- Use the cumulative distribution function to derive the derive the probability that the lowest income in a random sample of $n = 8$ individuals is smaller than 310 000 kroner.

Suppose we observe a random sample of $n = 8$ incomes, $y_1, y_2, \ldots, y_n$, given by 417000, 501000, 320000, 319000, 363000, 1056000, 512000, 379000 kr. Suppose that the statistic $\sum_{i=1}^{n} \ln(y_i/y_0) = 3.19$.

**b)**

- Derive the maximum likelihood estimator of $\lambda$ for a random sample of size $n$.
- Compute the maximum likelihood estimate of $\lambda$ for the data given above.

**c)**

- Let $X = 2\lambda \ln(Y/y_0)$. Show that $X$ has a chi-square distribution with 2 degrees of freedom. What is the distribution of $\sum_{i=1}^{n} 2\lambda \ln(Y_i/y_0)$? Explain why.
- Derive a 95%-confidence interval for $\lambda$ based on a random sample of size n. Compute the interval for the data given above.

**d)**

The median of $Y$ is given by $\mu^* = 2^{1/\lambda} y_0$ for $\lambda > 0$. We wish to test if the data above provides evidence for the claim that the median income in the population, $\mu^*$, is greater than 500 000 kr.

- Show that this is equivalent to testing the hypothesis $H_0 : \lambda = 1.36$ versus $H_1 : \lambda < 1.36$.
- Construct and carry out the test for the above data using a level of significance of 0.05. What is the conclusion of the test?

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 40

27 **7C**

**Problem 7**

Suppose that the income $Y$ of individuals in a population is a continuously distributed random variabel with probability density function

$$f_Y(y) = \begin{cases} \frac{\lambda y_0^\lambda}{y^{\lambda+1}}, & \text{for} \quad y \geq y_0, \\ 0, & \text{for} \quad y < y_0, \end{cases}$$

where $\lambda > 0$ is an unknown parameter and $y_0 = 400\,000$ kr is a known minimum income set by government regulations.

**a)**

Assume that $\lambda = 1.24$ in this point.

- Find the cumulative distribution function of income $Y$.
- Use the cumulative distribution function to derive the derive the probability that the lowest income in a random sample of $n = 6$ individuals is smaller than 410 000 kroner.

Suppose we observe a random sample of $n = 6$ incomes, $y_1, y_2, \ldots, y_n$, given by 514000, 593000, 420000, 419000, 463000, 1050000 kr. Suppose that the statistic $\sum_{i=1}^{n} \ln(y_i/y_0) = 1.85$.

**b)**

- Derive the maximum likelihood estimator of $\lambda$ for a random sample of size $n$.
- Compute the maximum likelihood estimate of $\lambda$ for the data given above.

**c)**

- Let $X = 2\lambda \ln(Y/y_0)$. Show that $X$ has a chi-square distribution with 2 degrees of freedom. What is the distribution of $\sum_{i=1}^{n} 2\lambda \ln(Y_i/y_0)$? Explain why.
- Derive a 95%-confidence interval for $\lambda$ based on a random sample of size n. Compute the interval for the data given above.

**d)**

The median of $Y$ is given by $\mu^* = 2^{1/\lambda} y_0$ for $\lambda > 0$. We wish to test if the data above provides evidence for the claim that the median income in the population, $\mu^*$, is greater than 700 000 kr.

- Show that this is equivalent to testing the hypothesis $H_0 : \lambda = 1.24$ versus $H_1 : \lambda < 1.24$.
- Construct and carry out the test for the above data using a level of significance of 0.05. What is the conclusion of the test?

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 40

28 **7D**

**Problem 7**

Suppose that the income $Y$ of individuals in a population is a continuously distrubuted random variabel with probability density function

$$f_Y(y) = \begin{cases} \frac{\lambda y_0^\lambda}{y^{\lambda+1}}, & \text{for} \quad y \geq y_0, \\ 0, & \text{for} \quad y < y_0, \end{cases}$$

where $\lambda > 0$ is an unknown parameter and $y_0 = 400\,000$ kr is a known minimum income set by government regulations.

**a)**

Assume that $\lambda = 3.11$ in this point.

- Find the cumulative distribution function of income $Y$.
- Use the cumulative distribution function to derive the derive the probability that the lowest income in a random sample of $n = 12$ individuals is smaller than 410 000 kroner.

Suppose we observe a random sample of $n = 12$ incomes, $y_1, y_2, \ldots, y_n$, given by 584000, 722000, 430000, 429000, 497000, 1701000, 740000, 524000, 645000, 431000, 802000, 586000 kr. Suppose that the statistic $\sum_{i=1}^n \ln(y_i/y_0) = 5.29$.

**b)**

- Derive the maximum likelihood estimator of $\lambda$ for a random sample of size $n$.
- Compute the maximum likelihood estimate of $\lambda$ for the data given above.

**c)**

- Let $X = 2\lambda \ln(Y/y_0)$. Show that $X$ has a chi-square distribution with 2 degrees of freedom. What is the distribution of $\sum_{i=1}^n 2\lambda \ln(Y_i/y_0)$? Explain why.
- Derive a 95%-confidence interval for $\lambda$ based on a random sample of size n. Compute the interval for the data given above.

**d)**

The median of $Y$ is given by $\mu^* = 2^{1/\lambda} y_0$ for $\lambda > 0$. We wish to test if the data above provides evidence for the claim that the median income in the population, $\mu^*$, is greater than 500 000 kr.

- Show that this is equivalent to testing the hypothesis $H_0 : \lambda = 3.11$ versus $H_1 : \lambda < 3.11$.
- Construct and carry out the test for the above data using a level of significance of 0.05. What is the conclusion of the test?

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 40

# 29 8A

**Problem 8**

We consider a normal population with unknown expected value $\mu$ and known variance $\sigma^2 = 1$. The goal is to test the null hypothesis $H_0 : \mu = 5$ against the alternative hypothesis $H_1 : \mu < 5$. We use the significance level $\alpha = 0.05$ and construct the test based on a random sample $X_1, X_2, \ldots, X_n$ from the normal population.

**a)**

Assume that we choose to reject $H_0$ if $\bar{x} \leq 5 - \frac{z_\alpha}{\sqrt{n}}$, where $\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$ and $z_\alpha$ denotes the value so that $\mathrm{P}(Z > z_\alpha) = \alpha$ for a standard normally distributed stochastic variable $Z$.

- Calculate the probability of making a type I error.
- Calculate the power of the hypothesis test when $n = 10$ and the true expected value is $\mu = 4.5$.

**b)**

A researcher performs two (2) independent random samples each of size $n = 10$. Assume that instead of the decision rule given in a), the researcher chooses to reject $H_0$ if at least one of the two random samples satisfy the decision rule $\bar{x} \leq 5 - \frac{z_\alpha}{\sqrt{n}}$.

- Calculate the probability of making a type I error using the researcher's approach. In other words, calculate the probability that at least one of the two random samples satisfy the decision rule in a) when the null hypothesis is true.
- Explain why the probability of making a type I error is larger with the researcher's approach than the answer in a).
- We are now going to change the decision rule and reject $H_0$ if at least one of the two random samples satisfy the decision rule $\bar{x} \leq k$. Determine the value of $k$ that results in a significance level of $0.05$.

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 20

30 **8B**

**Problem 8**

We consider a normal population with unknown expected value $\mu$ and known variance $\sigma^2 = 1$. The goal is to test the null hypothesis $H_0 : \mu = 5$ against the alternative hypothesis $H_1 : \mu < 5$. We use the significance level $\alpha = 0.10$ and construct the test based on a random sample $X_1, X_2, \ldots, X_n$ from the normal population.

**a)**

Assume that we choose to reject $H_0$ if $\bar{x} \leq 5 - \frac{z_\alpha}{\sqrt{n}}$, where $\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$ and $z_\alpha$ denotes the value so that $\mathrm{P}(Z > z_\alpha) = \alpha$ for a standard normally distributed stochastic variable $Z$.

- Calculate the probability of making a type I error.
- Calculate the power of the hypothesis test when $n = 10$ and the true expected value is $\mu = 4.5$.

**b)**

A researcher performs two (2) independent random samples each of size $n = 10$. Assume that instead of the decision rule given in a), the researcher chooses to reject $H_0$ if at least one of the two random samples satisfy the decision rule $\bar{x} \leq 5 - \frac{z_\alpha}{\sqrt{n}}$.

- Calculate the probability of making a type I error using the researcher's approach. In other words, calculate the probability that at least one of the two random samples satisfy the decision rule in a) when the null hypothesis is true.
- Explain why the probability of making a type I error is larger with the researcher's approach than the answer in a).
- We are now going to change the decision rule and reject $H_0$ if at least one of the two random samples satisfy the decision rule $\bar{x} \leq k$. Determine the value of $k$ that results in a significance level of $0.10$.

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 20

# 31 8C

**Problem 8**

We consider a normal population with unknown expected value $\mu$ and known variance $\sigma^2 = 1$. The goal is to test the null hypothesis $H_0 : \mu = 5$ against the alternative hypothesis $H_1 : \mu < 5$. We use the significance level $\alpha = 0.05$ and construct the test based on a random sample $X_1, X_2, \ldots, X_n$ from the normal population.

**a)**

Assume that we choose to reject $H_0$ if $\bar{x} \leq 5 - \frac{z_\alpha}{\sqrt{n}}$, where $\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$ and $z_\alpha$ denotes the value so that $\mathrm{P}(Z > z_\alpha) = \alpha$ for a standard normally distributed stochastic variable $Z$.

- Calculate the probability of making a type I error.
- Calculate the power of the hypothesis test when $n = 20$ and the true expected value is $\mu = 4.5$.

**b)**

A researcher performs two (2) independent random samples each of size $n = 20$. Assume that instead of the decision rule given in a), the researcher chooses to reject $H_0$ if at least one of the two random samples satisfy the decision rule $\bar{x} \leq 5 - \frac{z_\alpha}{\sqrt{n}}$.

- Calculate the probability of making a type I error using the researcher's approach. In other words, calculate the probability that at least one of the two random samples satisfy the decision rule in a) when the null hypothesis is true.
- Explain why the probability of making a type I error is larger with the researcher's approach than the answer in a).
- We are now going to change the decision rule and reject $H_0$ if at least one of the two random samples satisfy the decision rule $\bar{x} \leq k$. Determine the value of $k$ that results in a significance level of $0.05$.

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 20

32 **8D**

**Problem 8**

We consider a normal population with unknown expected value $\mu$ and known variance $\sigma^2 = 1$. The goal is to test the null hypothesis $H_0 : \mu = 5$ against the alternative hypothesis $H_1 : \mu < 5$. We use the significance level $\alpha = 0.10$ and construct the test based on a random sample $X_1, X_2, \ldots, X_n$ from the normal population.

**a)**
Assume that we choose to reject $H_0$ if $\bar{x} \leq 5 - \frac{z_\alpha}{\sqrt{n}}$, where $\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$ and $z_\alpha$ denotes the value so that $\mathrm{P}(Z > z_\alpha) = \alpha$ for a standard normally distributed stochastic variable $Z$.

- Calculate the probability of making a type I error.
- Calculate the power of the hypothesis test when $n = 20$ and the true expected value is $\mu = 4.5$.

**b)**
A researcher performs two (2) independent random samples each of size $n = 20$. Assume that instead of the decision rule given in a), the researcher chooses to reject $H_0$ if at least one of the two random samples satisfy the decision rule $\bar{x} \leq 5 - \frac{z_\alpha}{\sqrt{n}}$.

- Calculate the probability of making a type I error using the researcher's approach. In other words, calculate the probability that at least one of the two random samples satisfy the decision rule in a) when the null hypothesis is true.
- Explain why the probability of making a type I error is larger with the researcher's approach than the answer in a).
- We are now going to change the decision rule and reject $H_0$ if at least one of the two random samples satisfy the decision rule $\bar{x} \leq k$. Determine the value of $k$ that results in a significance level of $0.10$.

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 20

# 33 9A

**Problem 9**

We use the linear regression model

$$Y_i = \ln(x_i)\beta + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\ln$ denotes the natural logarithm, $x_i > 0$ is a covariate, $\beta$ is the coefficient of the covariate, and $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent and normally distributed stochastic variables with the same expected value 0 and **known** variance $\sigma^2 = 2$. The estimator $\hat{\beta} = \frac{\sum_{i=1}^{n} \ln(x_i) Y_i}{\sum_{i=1}^{n} (\ln(x_i))^2}$ is given for $\beta$.

- Show that $\hat{\beta}$ is unbiased and determine the variance of $\hat{\beta}$.
- Write an expression for a 95% prediction interval for a new value $Y$ with covariate value $x$. You need to justify your choice of expression, but you do not need to derive the expression.
- Calculate numerical values for the prediction interval when $x = 2$, $n = 10$, $\sum_{i=1}^{10} \ln(x_i) y_i = 4$ and $\sum_{i=1}^{10} (\ln(x_i))^2 = 9.3$.

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 10

# 34 9B

**Problem 9**

We use the linear regression model

$$Y_i = x_i^3\beta + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

where $x_i$ is a covariate, $\beta$ is the coefficient of the covariate, and $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent and normally distributed stochastic variables with the same expected value 0 and **known** variance $\sigma^2 = 2$. The estimator $\hat{\beta} = \frac{\sum_{i=1}^n x_i^3 Y_i}{\sum_{i=1}^n x_i^6}$ is given for $\beta$.

- Show that $\hat{\beta}$ is unbiased and determine the variance of $\hat{\beta}$.
- Write an expression for a 95% prediction interval for a new value $Y$ with covariate value $x$. You need to justify your choice of expression, but you do not need to derive the expression.
- Calculate numerical values for the prediction interval when $x = 2$, $n = 10$, $\sum_{i=1}^{10} x_i^3 y_i = 3$ and $\sum_{i=1}^{10} x_i^6 = 10$.

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 10

35 **9C**

**Problem 9**

We use the linear regression model

$$Y_i = \mathrm{e}^{x_i}\beta + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

where $x_i$ is a covariate, $\beta$ is the coefficient of the covariate, and $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent and normally distributed stochastic variables with the same expected value 0 and **known** variance $\sigma^2 = 2$. The estimator $\hat{\beta} = \frac{\sum_{i=1}^{n} \mathrm{e}^{x_i} Y_i}{\sum_{i=1}^{n} \mathrm{e}^{2x_i}}$ is given for $\beta$.

- Show that $\hat{\beta}$ is unbiased and determine the variance of $\hat{\beta}$.
- Write an expression for a 95% prediction interval for a new value $Y$ with covariate value $x$. You need to justify your choice of expression, but you do not need to derive the expression.
- Calculate numerical values for the prediction interval when $x = 2$, $n = 10$, $\sum_{i=1}^{10} \mathrm{e}^{x_i} y_i = 5$ and $\sum_{i=1}^{10} \mathrm{e}^{2x_i} = 15$.

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 10

36 **9D**

**Problem 9**

We use the linear regression model

$$Y_i = \sqrt{x_i}\beta + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

where $x_i > 0$ is a covariate, $\beta$ is the coefficient of the covariate, and $\epsilon_1, \epsilon_2, \ldots, \epsilon_n$ are independent and normally distributed stochastic variables with the same expected value 0 and **known** variance $\sigma^2 = 2$. The estimator $\hat{\beta} = \frac{\sum_{i=1}^n \sqrt{x_i} Y_i}{\sum_{i=1}^n x_i}$ is given for $\beta$.

- Show that $\hat{\beta}$ is unbiased and determine the variance of $\hat{\beta}$.
- Write an expression for a 95% prediction interval for a new value $Y$ with covariate value $x$. You need to justify your choice of expression, but you do not need to derive the expression.
- Calculate numerical values for the prediction interval when $x = 2$, $n = 10$, $\sum_{i=1}^{10} \sqrt{x_i} y_i = 100$ and $\sum_{i=1}^{10} x_i = 50$.

**Upload your file here. Maximum one file.**

The following file types are allowed: **.pdf** Maximum file size is **50 GB**

Select file to upload

Maximum marks: 10