Department of Mathematical Sciences, NTNU

Examination paper for **TMA4240 Statistics**

Academic contact during examination: Sara Martino and Håkon Tjelmeland
Phone: 44903330, 48221896

Examination date: November 30th 2019
Examination time (from-to): 09:00 - 13:00
Permitted examination support material: Support material code C.

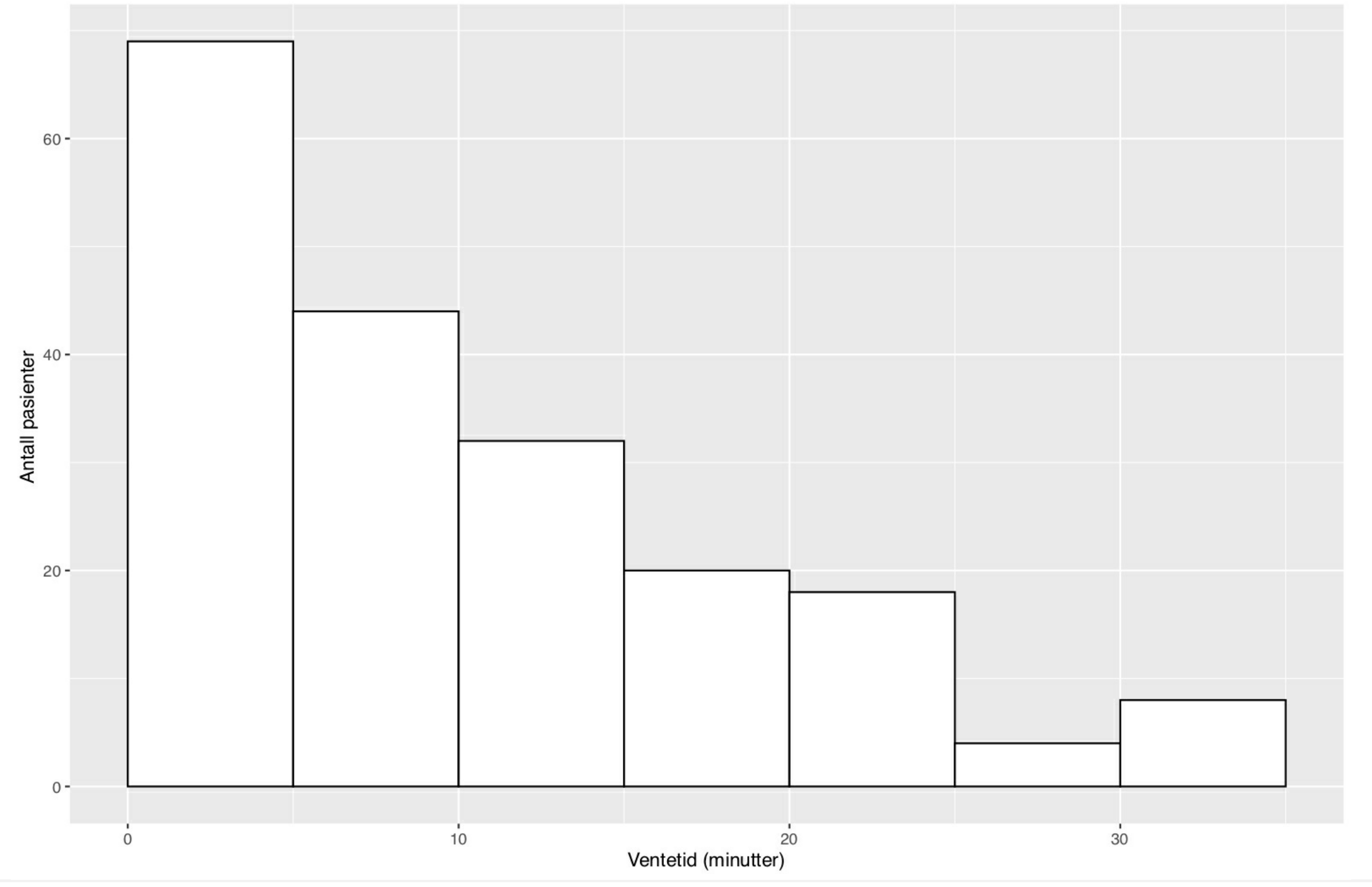
- *Tabeller og formler i statistikk*, (Akademika)
- A yellow sheet of paper (A5 with a stamp) with personal handwritten formulas and notes
- A specific basic calculator

Other information:

- For the exercises that are not multiple choice, all your answers should be justified and the hand-in material should contain sufficient calculations to make your reasoning clear.
- The weight in the grading of each subtask is indicated at the bottom of the exercises.

Language: English

1(a)



Multiple choice exercises

Introduction: The histogram above displays the frequencies of waiting times, in minutes, for 200 patients in a dentist's office.

Exercise: What is (approximately) the median of the waiting times?

Select one alternative:

- 2.50
- 7.25
- 12.25
- 15.00
- 17.50

Weight: 4 %.

Maximum marks: 4

1(b) **Introduction**: You are conducting a one-sided test of the null hypothesis that the population mean is 532 versus the alternative that the population mean is less than 532. Assume that the sample mean is 529 and the p-value is 0.01.

Exercise: Which of the following statements is then true?

Select one or more alternatives:

- There is a 0.01 probability that the population mean is smaller than 529.
- The probability of observing a sample mean smaller than 529 when the population mean is 532 is 0.01.
- There is a 0.01 probability that the population mean is smaller than 532.
- If the significance level is 0.05, you will not reject he null hypothesis H_0.
- None of the above.

Weight: 4 %.

Maximum marks: 4

1(c) **Introduction**: A manufacturer of balloons claims that p, the proportion of its balloons that burst when inflated to a diameter of up to 12 inches, is no more than 0.05. Some customers have complained that the balloons are bursting more frequently.

Exercise: If the customers want to conduct an experiment to test the manufacturer's claim, which of the following hypotheses would be appropriate?

Select one or more alternatives:

- $H_0 : p \neq 0.05$ against $H_1 : p = 0.05$.
- $H_0 : p = 0.05$ against $H_1 : p \neq 0.05$.
- $H_0 : p = 0.05$ against $H_1 : p < 0.05$.
- $H_0 : p = 0.05$ against $H_1 : p > 0.05$.
- $H_0 : p < 0.05$ against $H_1 : p = 0.05$.

Weight: 4 %.

Maximum marks: 4

1(d) **Introduction**: Suppose a 95% confidence interval for the proportion of Norwegians who exercise regularly is 0.29 to 0.37.

Exercise: Which one of the following statements is then FALSE?

Select one or more alternatives:

- It is reasonable to say that more than 25% of Norwegians exercise regularly.
- It is reasonable to say that more than 40% of Norwegians exercise regularly.
- The hypothesis that 33% of Norwegians exercise regularly cannot be rejected.
- It is reasonable to say that fewer than 40% of Norwegians exercise regularly.

Weight: 4 %.

Maximum marks: 4

1(e) **Introduction**: Let X and Y be two independent random variables following a geometric probability distribution with parameter p, that is

$X \sim g(x;p) = p(1-p)^{x-1}; x = 1, 2, \ldots$ and
$Y \sim g(y;p) = p(1-p)^{y-1}; y = 1, 2, \ldots$

Let $Z = \min(X, Y)$. In this exercise you may need the formula for a geometric series,

$$\sum_{k=0}^{n} a^k = \frac{1-a^{n+1}}{1-a}$$

Exercise: What is the probability mass function $f_Z(z)$ for Z?

Select one alternative:

- $f_Z(z) = (1-p)^{2(z-1)} p(2-p); z = 1, 2, \ldots$
- $f_Z(z) = 1 - (1-p)^{2z}; z = 1, 2, \ldots$
- $f_Z(z) = (1 - (1-p)^z)^2; z = 1, 2, \ldots$
- $f_Z(z) = p^2(1-p)^{2(z-1)}; z = 1, 2, \ldots$
- $f_Z(z) = -2(1-p)^{2z} \ln(1-p); z = 1, 2, \ldots$

Weight: 4 %.

Maximum marks: 4

2 Normal distribution

Introduction: Let X and Y be two independent and normally distributed stochastic variables. Assume that X has mean 0 and standard deviation 2, while Y has mean 1 and standard deviation 1.

Exercise:

- Draw the probability mass function of X and Y in a common plot.
- Find the following probabilities,

$$P(X \leq 1), \quad P(Y \geq -1) \quad \text{and} \quad P(X - Y \leq 0)$$

Fill in your answer here

Format

Words: 0

Weight: 10 %.

Maximum marks: 10

3(a) Sickle cell anemia

Introduction: Sickle cell anemia is a serious disease that causes anemia, i.e. low red blood cell counts. Sickle cell anemia is hereditary, and a child gets the disease if it inherits a particular recessive gene (a) from both the mother and the father. The child does not get the disease if it inherits the dominant gene (A) from at least one of its parents.

Thus, a person has either genotype AA, Aa or aa. People with genotype aa have sickle cell anemia, while people with AA or Aa do not have the disease. A child inherits one gene from each of its parents. If a parent has genotype Aa, the child will inherit either a or A from that parent with probability 0.5 for each of the two possibilities. Genes from the mother and the father are inherited independently from each other.

People with genotype Aa are called carriers of the disease, they do not have sickle cell anemia but can get children who have it.

Consider a couple where neither the man nor the woman have sickle cell anemia, but we do not know if they are carriers of the disease. Assume that the man and the woman each have probability 8 % of being carriers of the disease (this is the proportion of carriers in the African American population) and that the man and the woman are carriers or not independently of each another.

In this exercise, we will calculate the probability that this couple will have children who have sickle cell anemia or are carriers of the disease. Define the events

- M: The man is a carrier of sickle cell anemia,
- K: The woman is a carrier of sickle cell anemia,
- D: The couple's first born has sickle cell anemia, and
- B: The couple's first born is a carrier of sickle cell anemia.

Exercise:

- Draw the events M, K, D og B in a Venn diagram.
- Compute the probability that the couple's first born has sikle cell anemi.
- Compute the probability that the couple's first born is a carrier of sickle cell anemia.

Fill in your answer here

Format B I U x₂ x² Ω Σ

Words: 0

Weight: 10 %.

Maximum marks: 10

3(b) **Introduction**: We assume here that the couple has already got a child that does not have sickle cell anemia, and that they are planning to get another child.

Exercise:

- What is the probability that the second child will have sickle cell anemia?
- What is the probability that the second child will be a carrier of sickle cell anemia?

Fill in your answer here

Format B I U x_2 x^2 I_x Ω Σ

Words: 0

Weight: 10 %.

Maximum marks: 10

4(a) Binomially distributed variables

Introduction: Assume that X and Y are two independent stochastic variables, where $X \sim b(x; n, p)$ and $Y \sim b(y; n, 2p)$. That is, X represents the number of successes in n independent trials where each trial has probability p of success, while Y represents the number of successes in another set of n independent trials where each trial has probability of success $2p$.

Exercise:

- When $n = 12$ and $p = 0.2$, find the following probabilities,

$$P(X \leq 3),\ \ P(Y \geq 4 | X \leq 3) \text{ and } P(X + Y \leq 1).$$

Fill in your answer here

Format B I U Ω Σ

Words: 0

Weight: 10 %.

Maximum marks: 10

4(b) **Introduction**: In the rest of this exercise we assume that the value of the parameter $p \in [0, 0.5]$ is unknown and has to be estimated based on X and Y. The following three estimators are proposed:

$$\hat{p} = \frac{X+Y}{2n}, \ \tilde{p} = \frac{X+Y}{3n} \ \text{ and } \ p^\star = \frac{X}{2n} + \frac{Y}{4n} \tag{1}$$

Exercise:

- Which of the three estimators would you prefer? Justify your answer.

Fill in your answer here

Format B I U x₂ x² Tx

Words: 0

Weight: 10 %.

Maximum marks: 10

4(c) **Introduction**: Assume that we also want to estimate p, again based on X and Y, using the maximum likelihood principle.

Exercise:

- Find the expression for the log-likelihood function for p, $l(p)$.
- Find an estimate for p based on the maximum likelihood principle when $n = 25$, and we observe $x = 3$ and $y = 8$.

Fill in your answer here

Format

Words: 0

Weight: 10 %.

Maximum marks: 10

4(d) **Introduction**: Assume that we want to use X and Y to test

$$H_0 : p = 0.2 \text{ against } H_1 : p > 0.2.$$

In the rest of this exercise we assume that n is so large that one can assume with good approximation that X and Y are normally distributed.

Exercise:

- Consider the estimator in (1) that you found to best, choose a test statistic and define an (approximate) rejection criteria when the significance level is set to $\alpha = 0.05$. If you did not conclude which of the three estimators in (1) was the best, you may choose a test statistic based on any them.

Fill in your answer here

Format B I U x₂ x² Ix Ω Σ

Words: 0

Weight: 10 %.

Maximum marks: 10

4(e) **Exercise**:

- If the significance level for the hypothesis test is $\alpha = 0.05$, how large should n be in order to have at least probability 0.9 to find out that H_0 is wrong when $p = 0.25$?

Fill in your answer here

Words: 0

Weight: 10 %.

Maximum marks: 10