NTNU
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **TMA4245 Statistics**

Academic contact during examination: Gunnar Taraldsen[a], Torstein Fjeldstad[b]

Phone: [a]464 32 506, [b]962 09 710

Examination date: May 23th 2018

Examination time (from–to): 09:00 – 13:00

Permitted examination support material: C
- Tabeller og formler i statistikk, Akademika,
- Yellow stamped A5-sheet with personal hand written notes.
- A specific basic calculator

Other information:

All your answers should be justified.
The hand-in material should contain calculations leading to your answer.
The 10 tasks have equal weights in the grading.

Language: English

Number of pages: 4

Number of pages enclosed: 0

Checked by:

Informasjon om trykking av eksamensoppgave	
Originalen er:	
1-sidig ☐	**2-sidig** ☒
sort/hvit ☒	**farger** ☐
skal ha flervalgskjema ☐	

Date Signature

Problem 1

Let

$$f(x) = \begin{cases} \frac{1}{2\sqrt{x}} & x \in (0,1), \\ 0 & \text{otherwise.} \end{cases}$$

be the probability density of the random variable X.

a) Determine the cumulative distribution function of X and sketch its graph.

Calculate $P(X \geq 0.5)$ and $P(X \leq 0.7 | X \geq 0.5)$.

b) Calculate the probability density function of $Y = -\ln(X)$.

The variable Y has a well-known distribution. What well-known distribution? Find an expression for $\mu = \mathrm{E}(Y)$.

Briefly discuss the interpretation of expected value.

Problem 2 Illegal download of movies

The company KTK is monitoring illegal employee download of movies on their computer network.

a) Assume that 20% of the male employees have illegally downloaded at least one movie. The corresponding female percentage is 17%. The proportion of female employees at KTK is 67%.

Define relevant events.

Calculate the probability for the event that a randomly chosen employee has downloaded at least one movie.

Given that a randomly chosen employee has downloaded an illegal movie what is then the probability of this being a female employee?

Assume that the computer department at KTK sends out an e-mail warning for each violation of their computer policy guidelines. Assume that the number X of warnings for each year has a Poisson distribution with expected value μ. Assume that the number of e-mails sent in different years are independent. Assume that the true expected value is $\mu = 18$ in the years 2009-2016.

The number of warnings sent out in the year 2017 where 13 and the computer department wants to test if this is a statistically significant improvement by the hypothesis test:

$$H_0 : \mu = 18 \qquad \text{against} \qquad H_1 : \mu < 18.$$

b) Determine a critical value c such that $P(X \leq c | H_0 : \mu = 18) \leq 0.10$.

What is the critical region for the test with a significance level of 10%.

Will you keep or reject H_0 based on the given observation?

c) Derive an expression for the type-II error probability given that the expected number of warnings in 2017 is $\mu = \mu_1$ with $\mu_1 < 18$.

What is the type-II error probability when $\mu_1 = 14$?

Problem 3 Growth of soya bean plants

The growth of soya bean plants under given conditions is investigated by the planting of a large amount of soya bean seeds. The height of a randomly selected plant is measured one week after the first detected sprout. This measurement is repeated at the same time in the following weeks. The following table gives the resulting heights $h_1, h_2 \ldots, h_{11}$ in cm for weeks $t_1, t_2, \ldots, t_{11}$:

t (week)	1	2	3	4	5	6	7	8	9	10	11
h (cm)	2	4	11	16	19	25	27	32	34	39	42

Assume that each $H_1, H_2, \ldots, H_{11}$ is independently drawn from a normal distribution with mean $\mu_i = a + b(t_i - 6)$ for $i = 1, 2, \ldots, 11$ and standard deviation σ. The numbers a, b, and σ are unknown model parameters.

a) Plot graphically the observed heights as a function of the week.

Briefly discuss if a linear regression model is reasonable.

The data table gives

$$\sum_{i=1}^{11} h_i = 251, \quad \sum_{i=1}^{11} h_i^2 = 7577, \quad \sum_{i=1}^{11}(t_i - 6) = 0,$$
$$\sum_{i=1}^{11}(t_i - 6)^2 = 110 \quad \text{and} \quad \sum_{i=1}^{11}(t_i - 6)h_i = 449.$$

b) Show that the least squares estimators for a and b are given by

$$\hat{a} = \frac{1}{11}\sum_{i=1}^{11} H_i \qquad , \qquad \hat{b} = \frac{\sum_{i=1}^{11}(t_i - 6)\,H_i}{\sum_{i=1}^{11}(t_i - 6)^2}.$$

Calculate the estimates for the given data.

Sketch the resulting fitted line in the figure in question **a)**. Briefly discuss the result.

c) Write down the likelihood function for the regression model described above.

Show that the maximum likelihood estimators for a and b coincide with the least squares estimators.

In the following you can use, without proof, that

$$S^2 = \frac{1}{11-2}\sum_{i=1}^{11}\Big(H_i - \hat{a} - \hat{b}(t_i - 6)\Big)^2$$

is an unbiased estimator for σ^2, and that

$$V = \frac{(11-2)\cdot S^2}{\sigma^2}$$

is chi-squared distributed with $11 - 2 = 9$ degrees of freedom. Furthermore, $(11-2)\cdot S^2/\sigma^2$ is independent of $\hat{a}$ and $\hat{b}$. Given the observed values, the estimated variance is $s^2 = 1.88$.

d) Derive expressions for the mean and the variance of $\hat{b}$.

Derive a 95% confidence interval for the growth per week. Compute the confidence interval numerically.

What is the interpretation of a confidence interval?

Problem 4 Strike?

In a local worker union with 100 members, 20 members are selected at random to participate in a sample poll. The poll is about whether the union should go to strike or not. In the sample poll 3 are voting for a strike and 17 are voting against. Let X be the number of members voting for a strike.

What is the probability mass function of the corresponding random variable X? You must explain your reasoning.

Can you conclude that the majority of the members of the union are against strike? Your answer must be validated by an appropriate hypothesis test and the calculation of a corresponding p-value.

You can use that

$$\sum_{x=0}^{3} \frac{\binom{50}{x}\binom{50}{20-x}}{\binom{100}{20}} \approx 0.0004.$$