**Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag**

English

Contact during exam:
Jan Terje Kvaløy 73 59 35 20
Stian Lydersen 73 59 35 20

EXAM IN COURSE SIF5062/SIF5506 STATISTICS

Tuesday May 30th 2000
Time: 09:00–14:00

Permitted aids:
Approved calculator with empty memory, Statistiske tabeller og formler (Tapir forlag), Rottman: Matematisk Formelsamling

Results are ready: June 27th 2000.

Problem 1 pH-measurements

A method for measuring the pH-value of a solution gives measurement results which are assumed to be independent and normally distributed, with expectation μ equal to the true pH-value and variance $\sigma^2 = 0.060^2$. Let $X_1, \ldots, X_n$ be independent measurements of the pH of a certain solution.

a) Assume (only in this point) that the true pH-value of a solution is 6.8.

What is the probability that a single measurement gives a result which is below 6.74?

What is the probability that a single measurement gives a result which is between 6.74 and 6.86?

What is the probability that a single measurement, X, gives a result which deviates more than 0.06 from μ, in other words find $P(|X - \mu| > 0.06)$?

You are going to estimate the pH of a solution, and use the average of 5 independent measurements as your estimate. Let Y be the average of 5 independent measurements.

b) What is the probability that Y deviates more than 0.06 from μ?

Determine a 95% confidence interval for μ. Compute the numerical value of the interval when the average of five independent measurements is 6.76.

Problem 2 Cooling water pumps

In a factory two electrically driven cooling water pumps are installed. To have sufficient cooling capacity both pumps must function. Let

$$E_i = \text{Pump number } i \text{ is not functioning}, \quad i = 1, 2.$$

The following probabilities are given:

$$P(E_1) = P(E_2) = 0.01, \quad P(E_2|E_1) = 0.1.$$

Are the events E_1 and E_2 independent? Are they mutually exclusive? State the reason for your answers.

Find the probability for not having sufficient capacity, in other words find $P(E_1 \cup E_2)$.

The management is not satisfied with this, and wants to install a turbine driven pump in addition. The cooling capacity is then sufficient as long as at least 2 of the 3 pumps are functioning. Let

$$T = \text{The turbine driven pump is not functioning}$$

It is given that the turbine driven pump is functioning independently of the electrically driven pumps, and that $P(T) = 0.04$.

Find the probability of not having sufficient cooling capacity, in other words find the probability that at least 2 of the 3 pumps are not functioning.

Problem 3 Bacteria in food

In many kinds of food, there is often a certain amount of unwanted bacteria, and too high concentrations of such bacteria can be harmful. We are going to study the occurrence of a specific kind of bacteria in a specific type of food. Earlier studies have shown that the number of bacteria, Y, in a sample of t milligram (mg) of this kind of food, can be assumed to be Poisson distributed with expectation μt.

A critical limit for the number of bacteria in this kind of food is assumed to be 12 bacteria per mg. If the number of bacteria per mg is larger than 12, this might be harmful for the consumer.

Assume for now that it is known that the expected number of bacteria per mg of the food is 5, in other words $\mu = 5$.

a) Find the probability of finding more than 12 bacteria in a food sample of 1 mg.

Find the probability of finding more than 6 bacteria in a food sample of 0.5 mg.

If you take 10 independent samples of 1 mg, what is the probability that at least one of the 10 samples contains more than 12 bacteria?

In practice μ, the expected number of bacteria per mg food, is unknown. Based on independent measurements $Y_1, \ldots, Y_n$ of the number of bacteria in n samples of respective size $t_1, \ldots, t_n$ mg we want to estimate μ.

b) Show that the maximum likelihood estimator (MLE) for μ is

$$\hat{\mu} = \frac{Y_1 + \cdots + Y_n}{t_1 + \cdots + t_n}$$

Calculate the expectation and variance of the estimator.

To ensure that the limit of 12 bacteria per mg food is not exceeded too often, the authorities recommend that the expected number of bacteria per mg food should not exceed 6, in other words $\mu \leq 6$.

c) For a specific lot of food, we wishes to examine if there is any reason to claim that the authorities recommendation is exceeded.

Formulate this as a hypothesis test.

In the rest of this point you can use without proof that a Poisson distributed variable with expectation larger than 15 is approximately normally distributed. Explain why the statistic

$$\frac{\hat{\mu} - 6}{\sqrt{6/\sum_{i=1}^{n} t_i}}$$

is approximately normally distributed under the null hypothesis if $\sum_{i=1}^{n} t_i > 2.5$.

Perform the hypothesis test using an (approximate) 5% level of significance when $n = 10$ independent measurements gave the results reported in the table below.

measurement i	1	2	3	4	5	6	7	8	9	10
y_i	6	4	9	7	5	8	9	3	7	7
t_i	0.8	0.7	1.3	1.1	1.2	0.9	1.0	0.8	0.8	1.4

It is given that $\sum_{i=1}^{10} y_i = 65$ and $\sum_{i=1}^{10} t_i = 10$.

In order to make the routines more efficient, it is of interest to use a new and faster method for measuring the number of bacteria in a sample of food. The drawback with this new measurement method is that not all of the bacteria in a sample are discovered. Assume that each single bacterium is discovered with the same probability p, and that each bacterium is either discovered or not discovered independently of all the other bacteria.

Let, as before, Y be the total number of bacteria in a sample of t mg, and let X be the number of these discovered by the new measurement method.

d) Explain why X given $Y = y$ will be binomially distributed with parameter y and p.

Find the joint probability distribution of X and Y, and use this joint distribution to show that the (marginal) distribution of X is a Poisson distribution with expectation μtp.

The new measurement method has been used to perform n independent measurements $X_1, \ldots, X_n$ on samples of respectively $t_1, \ldots, t_n$ mg. Based on these measurements we want to estimate μ. Assume that p is known, and suggest how the estimator of μ from point b) can be modified to take into account the fact that we now know $X_1, \ldots, X_n$ instead of $Y_1, \ldots, Y_n$.

Problem 4 The runner

A 45 year old man started running 9 years ago, and has each year since participated in an exercise run. The time he spent to complete the race each year is reported in the table below.

year i	1	2	3	4	5	6	7	8	9
age x_i	37	38	39	40	41	42	43	44	45
time y_i	45.54	41.38	42.50	38.80	41.26	37.20	38.19	38.05	37.45

It is given that $\sum_{i=1}^9 x_i = 369$, $\sum_{i=1}^9 y_i = 360.37$, $\sum_{i=1}^9 (x_i - \bar{x})^2 = 60$, $\sum_{i=1}^9 (y_i - \bar{y})^2 = 63.28$ and $\sum_{i=1}^9 (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^9 (x_i - \bar{x})y_i = -52.57$.

We shall assume that the observations can be considered as realizations of independent normally distributed variables $Y_1, \ldots, Y_9$, where $\mathrm{E}(Y_i) = \alpha + \beta x_i$ and $\mathrm{Var}(Y_i) = \sigma^2$.

a) Write down the usual unbiased estimators $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$ for α, β and σ^2. Calculate the estimates of α and β using the reported data. Plot the data set and the estimated regression line.

It is given that the estimate of σ^2 is 1.568^2.

b) Calculate an expression for the variance of the estimator $\hat{\beta}$.

Perform the test $H_0: \beta = 0$ against $H_1: \beta \neq 0$, using an 1% level of significance.

What is the practical interpretation of the test above?

The runner wishes to predict the time he will spend to complete the race next year (at an age $x_0 = 46$ years).

c) Calculate the predicted time.

It is given that $\mathrm{Var}(\hat{\alpha} + \hat{\beta}x_0) = \sigma^2(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2})$. Determine a 95% prediction interval for Y at $x_0 = 46$ years. Calculate the interval numerically using the reported data.

If the runner asks you to predict the time he will spend to complete the race in 15 years (age 60 years), what will your answer be?