Norges teknisk–
naturvitenskapelige universitet
Institutt for matematiske fag

English

Contact during exam:
Øyvind Bakke, phone 73 59 81 26
Odd Kolbjørnsen, phone 73 59 35 27

SIF5062/SIF5506 Statistics

Saturday 25 May 2002
Time: 09.00–14.00

Permitted aids: *Tabeller og formler i statistikk,* Tapir Akademisk Forlag
K. Rottmann: *Matematisk formelsamling*
Specific simple calculator

Final grading: 29 June 2002

All points have equal impact on the grading. Your answers must include the reasoning.

Problem 1

A rare species of grass is planted in a particular field at a botanical research station. At a given time the length X measured in cm of a randomly selected blade of grass is exponentially distributed; that is, X has probability density f given by $f(x) = \frac{1}{\beta}e^{-x/\beta}$ for $x \geq 0$ and cumulative distribution F given by $F(x) = 1 - e^{-x/\beta}$ for $x \geq 0$.

a) Assume (only for this point) that $\beta = 10$. Calculate

$$P(X \leq 4), \quad P(X > 7) \quad \text{and} \quad P(X > 7 \mid X > 4).$$

We assume the lengths of the blades of grass to be independent. The plan is to make a random selection of blades of grass and measure the lengths of the blades in order to estimate β.

b) Based on measured lengths in a a random selection of k blades of grass, work out the maximum likelihood estimator (MLE) of β.

By a mistake the gardener mows the field of grass early the day the measurements are to be taken; then he plows the field. The scientist now plans to measure the lengths Y of a random selection of the blades of grass from the mower's collector; he plans to estimate β based on these lengths.

All blades of grass were cut at the same height c, and only those higher than c were cut. Given that a blade of grass had length $X > c$, it is now in the collector of the mower, and has length $Y = X - c$.

c) Show that $P(X > c) = e^{-c/\beta}$.

Find $P(Y > y)$, where $y > 0$.

Which familiar distribution does Y have?

Unfortunately, the blades of grass in the collector have dried and shrunk, making measurements of the blades impossible. It is however possible to count the number of blades in the collector. The collector was empty when the gardener started to mow the field. The scientist also knows the total number of blades of grass in the field prior to the mowing (the field was continuously monitored and dead blades of grass were removed on a regular basis).

d) Let Z be the number of blades in the collector. Assume the total number of blades of grass alive in the field that day to be n. Under which conditions will Z be binomially distributed?

Assume Z to be binomially distributed. Express the probability of success in terms of β and c.

Propose an estimator of β based on Z.

Problem 2

An analyzing laboratory has implemented quality control on a regular basis. A control solution with known concentration, 0.10 mg/l is analyzed together with every series of analysis. Ordinarily the outcome when analyzing the control solution can be regarded as normally distributed with expected value (mean) μ and variance σ^2, with σ^2 being the variance of the measurement error in the method used for analyzing and μ under normal circumstances equals 0.10 mg/l.

An *event of alarm* A occurs when the measured value, X, of the control solution deviates more than two standard deviations from the concentration 0.10 mg/l, that is, $|X - 0.1| > 2\sigma$ (i.e. $X < 0.1 - 2\sigma$ or $X > 0.1 + 2\sigma$).

An *event of action* B occurs when the measured value, X, of the control solution deviates more than three standard deviations from the concentration 0.10 mg/l, that is, $|X - 0.1| > 3\sigma$.

a) Assume (only for this point) that $\sigma = 0.01$ mg/l.

Compute $P(B)$ and $P(B \mid A)$ when $\mu = 0.10$ mg/l.

Assume the control solution is contaminated so that $\mu = 0.11$ mg/l, and compute $P(B)$ in this situation.

Assume for the remaining part of this problem that $\mu = 0.10$ mg/l. In order to estimate σ^2 the results of several independent analyses of control solutions, $X_1, X_2, \ldots, X_n$, are used. Two estimators of σ^2 are proposed.

$$\widehat{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu)^2,$$

where $\mu = 0.10$ mg/l is the known concentration of the control solution, and

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2,$$

where $\overline{X}$ is the sample mean.

In the points to follow you can use without proof that $\sum_{i=1}^{n}(X_i - \mu)^2/\sigma^2$ and $\sum_{i=1}^{n}(X_i - \overline{X})^2/\sigma^2$ have the χ^2distribution (chi-squared distribution) with n and $n-1$ degrees of freedom, respectively.

b) What two properties does a good estimator have?

Which of the two estimators $\widehat{\sigma^2}$ and S^2 would you recommend?

The results measured in units of mg/l coming from analyzing 20 control solutions are:

0.0939	0.1069	0.1002	0.1106	0.0866	0.1048	0.0837	0.0856	0.1029	0.0986
0.0887	0.0971	0.0942	0.0910	0.1025	0.0851	0.1031	0.0797	0.1053	0.1034

You are informed that $\sum_{i=1}^{20} x_i = 1.9240$ and $\sum_{i=1}^{20} x_i^2 = 0.1866$.

c) Find a 90 % confidence interval for σ^2 by using the estimator you recommended in point (b).

d) A 90 % confidence interval with length at most 50 % of the point estimate is wanted. How many observations are needed for this? State your answer rounded up to the nearest number divisible by 10 (10, 20, 30, ...). (Hint: Try out numbers from the table in *Tabeller og formler i statistikk.*)

Problem 3

The municipal sewage purification plant has a permit to let out water containing no more than 0.20 mg/l phosphorus. An environmental organization claims that the plant is outdated and that the content of phosphorus in the water from the plant exceeds the permit. To argue their case the organization gets the content of phosphorus analyzed in ten samples of the water from the purification plant, X_1, X_2, ... , X_{10}. It is reasonable to assume the samples to be independent and normally distributed with expected value (mean) equal to the true phosphorus content and *known* standard deviation of 0.02 mg/l.

The results of the samples measured in mg/l is:

0.2188	0.1782	0.1960	0.1884	0.2300	0.2446	0.2242	0.1950	0.2146	0.2055

You are informed that $\sum_{i=1}^{10} x_i = 2.0954$.

The environmental organization wants to prove that the phosphorus content in the water from the purification plant exceeds 0.20 mg/l.

a) State the problem as a hypothesis test. Compute the critical region. Use significance level 0.05. What is the conclusion when the data is as given above?

The environmental organization claims the conclusion in the test is due to a bad experimental design, i.e., ten samples are not sufficient to uncover the defect of the plant.

b) What is the probability of detecting the defect of the plant when using the test above with ten new samples, if the phosphorus content of the water is 0.21 mg/l?

How many observations are needed to have 90 % probability of detecting a phosphorus content in the water of 0.21 mg/l?