NTNU – Trondheim
Norwegian University of
Science and Technology

Department of Mathematical Sciences

Examination paper for **ST0103 Statistics with Applications**

Academic contact during examination: Øyvind Bakke

Phone: 73 59 81 26, 990 41 673

Examination date: 19. December 2015

Examination time (from–to): 9:00–13:00

Permitted examination support material: Yellow A4 sheet with your own handwritten notes, specific calculator (Casio fx-82ES Plus, Citizen SR-270X, Citizen SR-270X College or HP 30s), *Tabeller og formler i statistikk* (Tapir forlag), *Matematisk formelsamling* (K. Rottmann)

Other information:
In the grading, each of the ten points counts equally.

You should demonstrate how you arrive at your answers (e.g. by including intermediate answers or by referring to theory or examples from the reading list).

Language: English

Number of pages: 2

Number of pages enclosed: 0

Checked by:

Date Signature

Problem 1

In connection with an investigation of a mineral deposit, n holes are drilled. For each hole there is a probability p that the mineral is found. Assume that the number X of holes in which the mineral is found has the binomial distribution with parameters n and p.

a) Which assumptions must be satisfied for X to have a binomial distribution? Find $P(3 \leq X \leq 4)$ when $n = 20$ and $p = 0.3$.

b) Find an approximate value of $P(50 \leq X \leq 75)$ when $n = 200$ and $p = 0.3$.

Now assume that p is unknown. A new sequence of m holes are drilled. The number Y of holes in which the mineral occurs in the new sequence is assumed to be independent of X and to have the binomial distribution with success probability p. Two estimators of p are suggested, $\hat{p} = (X+Y)/(n+m)$ and $\tilde{p} = \frac{1}{2}(X/n+Y/m)$.

c) Which of the two estimator would you prefer? It is sufficient that you investigate the properties of $\hat{p}$ and $\tilde{p}$ for $n = 20$ and $m = 10$.

Problem 2

A certain pattern of nucleotides occurs on a DNA sequence according to a Poisson process with intensity (rate) $\lambda = 0.03$, so that the number of times the pattern occurs on a DNA sequence of length l (measured in number of thousand base pairs) has the Poisson distribution with parameter λl.

a) Find the probability that the pattern occurs 2 or more times on a DNA sequence of length 100. Find the conditional probability that the pattern occurs exactly twice given that it occurs 2 or more times.

b) Briefly argue why the distance from a pattern to the next occurrence of the pattern has the exponential distribution with parameter λ (i.e., having expected value $1/\lambda$). What is the probability that this distance is greater than 50?

We measure the distance from the first occurrence of the pattern to the second, the distance from the second occurrence to the third, and the distance from the third occurrence of the pattern to the fourth. By properties of a Poisson process, these three distances are independent (you are not required to show this fact).

c) Find the probability that the least of the three distances is greater than $1/\lambda$.

Problem 3

The concentration (measured in mM) of a substance that is formed in a reaction has the normal distribution with expected value (mean) 9.0 and standard deviation 2.0. The reaction is performed twice. The concentrations are X the first time and Y the second, and X and Y are independent.

a) Find $P(X > 10.0)$ and $P\Big(\frac{1}{2}(X + Y) > 10.0\Big)$.

An analysis method is dependent on the concentration following the first reaction being greater than 10.0 and the concentration following the second reaction being greater than the concentration following the first reaction.

b) Find the probability of this – that is, $P(10.0 < X < Y)$. (Hint: It may be helpful to sketch the event in the xy-plane.)

Problem 4

In 1894, John James Strutt (Lord Rayleigh) published *On an anomaly encountered in determinations of the density of nitrogen gas* (*Proc. R. Soc. Lond.* 1894 **55**, pp. 340–344). In the article he describes how he in a series of 8 experiments from November 1893 to January 1894 produced nitrogen gas from nitrogenous compounds and weighed the gas contained in a flask of known volume. The measurements of the 8 masses were 2.30143, 2.29890, 2.29816, 2.30182, 2.29869, 2.29940, 2.29849 and 2.29889 g.

Assume that the measurements are independent and come from the same normal distribution. It is given that the sample mean of the 8 measurements is 2.29947 g and that the sample standard deviance is 0.00138 g.

a) Find a 95% confidence interval for the standard deviation of the normal distribution.

In another sequence of 7 experiments from December 1893 to February 1894 Strutt weighed nitrogen gas derived from air (as opposed to the aforementioned 8 experiments, in which the gas was produced from nitrogenous compounds). The measurements of the masses were 2.31017, 2.30986, 2.31010, 2.31001, 2.31024, 2.31010 and 2.31028 g.

Assume that also these measurements are independent (also of the previous experimental series) and come from a normal distribution. It is given that the sample mean of the 7 measurements is 2.31011 g and that the sample standard deviation is 0.00014 g.

b) Test the null hypothesis stating that the expected value (mean) of the measured mass of the two methods are equal, against the alternative that the expected values are different. Use significance level 0.001. Which assumptions must be satisfied for the test method to be correct? Briefly discuss whether the assumptions seem to be satisfied.