TMA4245 Statistics
Exam May 2017

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Oppgave 1

The quality division in a company producing watches wants to have a look at the defect watches that sometimes come out of the production. They decide to use $k = 3$ defect watches in their inspection. From a production line there is a continuous stream of watches and every watch produced has a probability p of being defect, independent of each other.

Let X denote the smallest number of watches that need to be inspected from this production line to identify exactly $k = 3$ defect watches. We then know that the random variable X has a negative binomial distribution given as follows:

$$b^*(x; k, p) = \binom{x-1}{k-1} p^k (1-p)^{x-k} \qquad ; \qquad x = k, k+1, \ldots$$

a) Assume in this point that the probability of being defect $p = 0.1$, and calculate the probabilities

$$P(X > 3),$$
$$P(X < 6),$$
$$P(X \geq 6 | X > 3).$$

The probability p of being defect is now assumed to be unknown and needs to be estimated. The quality division repeats the trial of identifying exactly $k = 3$ defect watches n times, and obtains a random sample: $X_1, \ldots, X_n$. Based on this random sample, one wants to estimate p.

b) Derive the maximum likelihood estimator $\hat{p}$ for p, based on the random sample.

Assume that the watch factory has two separate production lines, denoted by A and B, and with unequal probabilities that watches are defect denoted by p_A og p_B, respectively.

The statistician in the quality division receives an observation $X = x$ of the number of watches that has to be inspected before $k = 3$ defect watches are identified from one of the production lines. He does not know whether the observation comes from production line A or B, so he assumes that the probability is 0.5 for each of the possibilities.

c) Use Bayes' rule to derive an expression for the probability that the observation comes from production line A given that $X = x$.

Let $p_A = 0.1$ and $p_B = 0.2$, and also that $x = 5$, and calculate the numerical value of the probability that the observation comes from production line A.

Oppgave 2

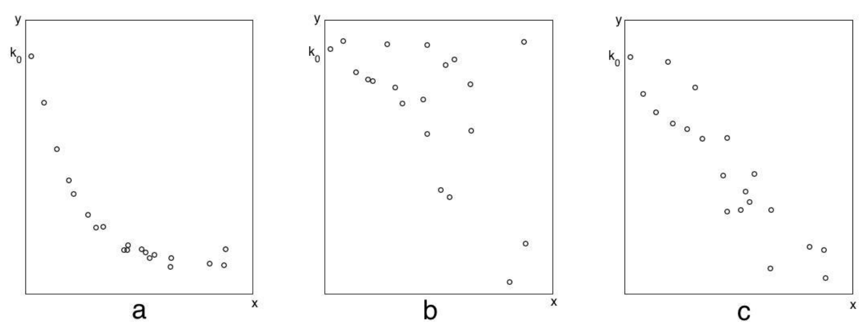
A car manufacturer wants to evaluate the wear on the brake pads of the cars that are produced. A simple linear regression model is defined,

$$Y = k_0 - \beta x + \epsilon,$$

where the response variable Y is the thickness of the brake pad, the explanation variable x is the number of kilometers driven, k_0 is the pad thickness of a new car and β is the rate of wear. The error term ϵ is assumed to be normally distributed, $n(\epsilon; 0, \sigma)$. We assume that k_0 is known, while the rate β and the variance σ^2 are unknown model parameters that need be estimated.

The car manufacturer designs a test to estimate β and σ^2. A group of n testdrivers drive different cars over a varying number of kilometers and the pad thickness is then measured. This defines a random sample from the model, $(x_1, Y_1), \ldots, (x_n, Y_n)$.

In Figure 1 are presented three plots of possible outcomes $(x_1, y_1), \ldots, (x_n, y_n)$ for $n = 20$.



Figur 1: Three possible outcomes of the test in Problem 2.

a) For which of the three plots in Figure 1 does the simple linear regression model appear to be a good model? Justify your answer.

Why is the model not good for the other two plots?

The simple linear regression model defined above is by necessity an approximation that is valid only for an interval of values of the explanation variable x. Explain briefly why this is so.

b) Use either a least squares method or the maximum likelihood method to derive a estimator $\hat{\beta}$ for β based on the random sample. Show that the estimator becomes

$$\hat{\beta} = \frac{k_0 \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i Y_i}{\sum_{j=1}^{n} x_j^2}.$$

Derive an expression for the mean value and variance of $\hat{\beta}$.

As an estimator of σ^2 based on the random sample it is reasonable to use

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\Big[Y_i - (k_0 - \hat{\beta}x_i)\Big]^2 .$$

It is given that $\hat{\beta}$ is normally distributed, that

$$V = \frac{(n-1)\hat{\sigma}^2}{\sigma^2}$$

is chi-square distributed with $(n-1)$ degrees of freedom, and that $\hat{\beta}$ and V are independent random variables.

c) Derive a $100(1-\alpha)\%$ confidence interval for β.

Describe briefly how the confidence interval may be used to test if the rate of wear is exactly equal to β_0.

Oppgave 3

A farmer from Sogn grows apples. He sells his apples in what is designated as '3 kg bags'. It is obviously an entire number of apples in every bag, so the weight of the bags will necessarily vary. A randomly chosen bag weighs X kg, where X is normally distributed with expectation μ and variance σ^2. Consider μ as unknown and let $\sigma^2 = 0.4^2$. It is obviously desirable that the expectation μ is 3 kg.

The store house of Rema 1000 at Sandmoen receives a big truck load of '3 kg bags' of apples from the farmer. The purchasing office of Rema 1000 wants to verify that the bags are heavy enough. They take a random sample of $n = 3$ bags from the truck load. They weigh the bags and register the following weights: X_1, X_2, X_3. A reasonable estimator for the expected weight μ is

$$\hat{\mu} = \frac{1}{3}\sum_{i=1}^{3} X_i = \bar{X}.$$

a) The estimator $\hat{\mu}$ is normally distributed. Explain briefly in your own words why.

Derive expressions for the expected value and variance of this normal distribution.

Is the estimator $\hat{\mu}$ unbiased? Justify the answer.

Desribe briefly in your own words what it means for an estimator to be unbiased.

The purchasing office wants to make sure that the expected weight of the bags, μ, is at least 3 kg. The statistician at the office formulates the weight control as a hypothesis testing problem,

$$H_0 : \mu = 3 \text{ against } H_1 : \mu < 3$$

and uses significance level $\alpha = 0.05$ in a test with the estimator $\hat{\mu}$ as test statistic.

b) Derive the rejection region for $\hat{\mu}$ with respect to the hypotheses defined above.

c) Derive an expression for the power of the test defined in **b)**.

Make a simple graphical sketch of the power of the test.

If the expected weight μ is only 2.9 kg, then the statistician wants to find out that the bags weigh too little with a probability of at least 0.9. Find out how many bags n there must then be in the random sample from the truck load.

The statistician continues to play a little with this problem after working hours. He considers the ordered version of the random sample, $X_{(1)}, X_{(2)}, X_{(3)}$ in increasing order. Then he defines an alternative estimator for μ,

$$\tilde{\mu} = X_{(2)}$$

d) Derive an expression for the probability distribution of the estimator $\tilde{\mu}$.

Show that $\tilde{\mu}$ is an unbiased estimator for μ.

Fasit

1. **a)** 0.999, 0.00856, 0.9924 **b)** $\widehat{p} = \frac{nk}{\sum_{i=1}^{n} X_i}$ **c)** 0.1366

2. **b)** $\mathrm{E}\left[\widehat{p}\right] = \beta$, $\mathrm{Var}\left[\widehat{\beta}\right] = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2}$

3. **a)** $\mathrm{E}\left[\widehat{\mu}\right] = \mu$, $\mathrm{Var}\left[\widehat{\mu}\right] = \frac{\sigma^2}{n}$, $\widehat{\mu}$ is unbiased **b)** Reject H_0 if $\widehat{\mu} < 2.62$ **c)** 138 **d)** $f_{X_{(2)}}(x) = 6F_X(x)f_X(x)\left[1 - F_X(x)\right]$