UPPSALA UNIVERSITY
Department of Mathematics
Rolf Larsson

Exam in Mathematical Statistics
Inference 1 1MS035
2021–08-16

Time: 14.00-19.00. Limits for the credits 3, 4, 5 are 18, 25 and 32 points, respectively. The solutions should be well motivated.

Permitted aids: Pocket calculator. Dictionary. Formelsamling för inferensteori 1.

1. We have a random sample 2,1,1,3 from a random variable X with probability function

$$P(X=k)=\begin{cases}1-6\theta, & k=0,\\ \theta k, & k=1,2,3,\\ 0, & \text{otherwise},\end{cases}$$

where $0 \le \theta \le 1/6$. Estimate θ using

   (a) the method of moments, (1p)

   (b) the least squares method, (2p)

   (c) maximum likelihood. (2p)

2. We have a random sample x_1, x_2, x_3, x_4, x_5 from a random variable X with expectation μ and variance 10, and another random sample y_1, y_2 from a random variable Y with expectation 2μ and variance 1. We may assume that X and Y are independent. The sample means are denoted by $\bar{x}$ and $\bar{y}$.

   The following estimates of μ are proposed:

$$\mu_1^* = \frac{\bar{x}+\bar{y}}{3}, \quad \mu_2^* = \frac{\bar{x}+2\bar{y}}{5}.$$

   (a) Show that μ_1^* and μ_2^* are both unbiased. (2p)

   (b) Which one of μ_1^* and μ_2^* is most efficient? (3p)

Please turn the page!

3. Burt sells umbrellas at a summer party. The number of umbrellas that he sells until it starts raining, X, is supposed to be a discrete random variable with distribution function

$$F(x) = P(X \leq x) = 1 - (1 - \theta)^x,$$

where $x = 1, 2, ...$ and $0 < \theta < 1$. For this distribution, it holds that $E(X) = 1/\theta$, so the smaller the θ, the bigger we expect X to be.

Burt's friend, the meterologist Börthie, claims that θ is less than 0.1.

At this particular occation, Burt sells 32 umbrellas before it starts to rain.

(a) Is Börthie right? Investigate this by testing a suitable hypothesis. Test at the significance level 5%. (2p)

(b) Calculate the critical region for the test. (1p)

(c) In case $\theta = 0.01$, what is the power of your test? (2p)

4. On a calm summer day, Lennart and Helga are out fishing. They weigh the fish that they catch and note the weight (in kilograms) before they let them back into the water. They go on until each of them has caught five fish. The results are given in the table below.

The heavier the fish, the better.

In terms of this, are Lennart and Helga equally good at fishing? Try to find this out by testing a suitable hypothesis. Be careful to specify all assumptions that you make. (5p)

Lennart's fish	Helga's fish
1.7	2.3
1.6	1.2
0.7	0.8
3.2	0.5
0.2	2.7

5. Amy picks 200 strawberries on a field, without taking notice of if they are rotten or not. Let the probability that a strawberry is rotten be p.

After picking, Amy gives all the strawberries to Sheldon. Sheldon checks them, and finds out that 15 are rotten.

(a) Calculate a 95% confidence interval for p. (4p)

(b) The owner of the field claims that $p = 0.02$. In the light of your result in (a), what do you think about this statement? (1p)

Please turn the page!

6. We have a random sample of 100 observations from $X \sim N(\mu, \sigma^2)$, where it is known that $\sigma^2 = 4$, while μ is unknown. The mean of the sample is $\bar{x} = 50$.

   (a) Calculate a 99% confidence interval for μ. (2p)

   (b) Calculate a 99% confidence interval for $P(X \leq 49.5)$. (3p)

7. A light bulb of a certain brand is supposed to have a lifelength that is exponentially distributed with expectation μ. A random sample of 200 such light bulbs was taken. Their mean lifelength was 3.2 years.

   The manifacturer claims that the average lifelength of a light bulb of this brand is at least 4 years. Perform a hypothesis test to check if the manifacturer is wrong. (5p)

8. Data is available for selling of 20 houses in Milwaukee, Wisconsin. The price in thousands of dollars is given by y, and the living space (i hundreds of square feet) is given by x. The data is plotted in the figure below.

   The object is to estimate the regression model

   $$Y_i = \alpha + \beta x_i + \varepsilon_i,$$

   for $i = 1, 2, ..., 20$, where the ε_i are independent $N(0, \sigma^2)$.

   (a) By looking at the plot, try to judge which pair of estimated β (β^*) and coefficient of determination R^2 that we have for our data, out of the following alternatives. Motivate your answer. (3p)

      i. $\beta^* = 2.68$, $R^2 = 2\%$.
      ii. $\beta^* = 1.03$, $R^2 = 75\%$.
      iii. $\beta^* = 2.75$, $R^2 = 83\%$.
      iv. $\beta^* = 1.12$, $R^2 = 3\%$.

   (b) The observed means are $\bar{x} = 16.22$ and $\bar{y} = 76.55$.
   What is the estimated α? (2p)

GOOD LUCK!

Appendix: figures

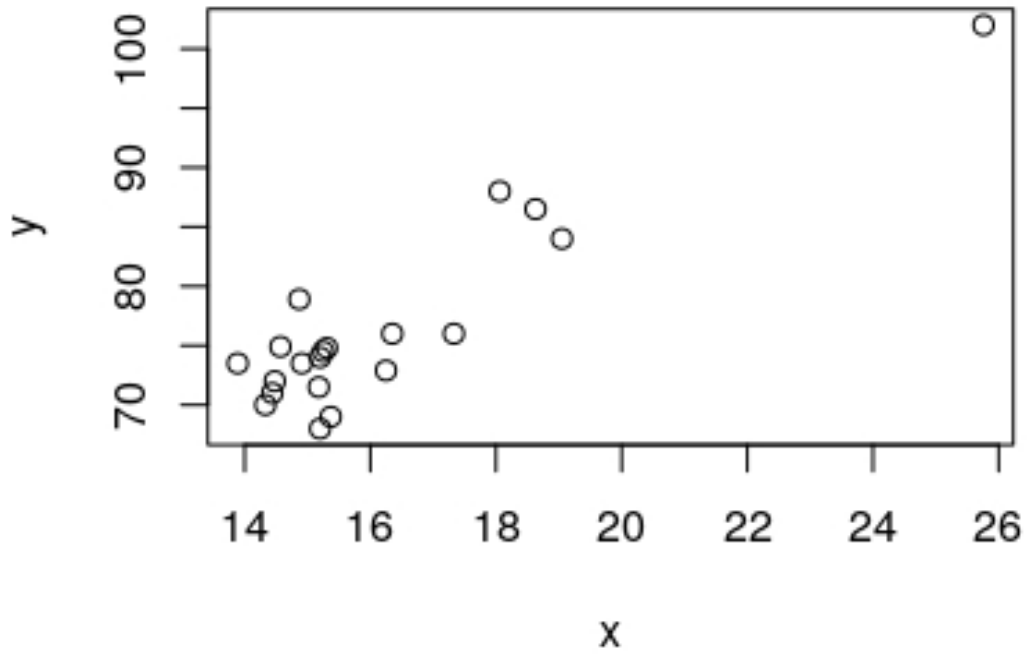

Figure 1: Plot for problem 8.