UPPSALA UNIVERSITY
Department of Mathematics
Rolf Larsson

Exam in Mathematical Statistics
Inference 1 1MS035
2021–06–12

Time: 8.00-13.00. 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 2.3, 1.2, 0.2, 6.7, 3.1 from a continuous random variable X with density function

$$f_X(x) = \frac{3x^2}{\theta^3} \exp\left(-\frac{x^3}{\theta^3}\right),$$

where $x > 0$, and 0 otherwise. Assume that $\theta > 0$.
Without proof, you may use that $E(X) = \theta\Gamma(4/3) \approx 0.893\theta$.

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 from a random variable X with expectation $\mu - m$ and variance 4, and another random sample $y_1, y_2, ..., y_5$ from a random variable Y with expectation m and variance 1. Moreover, we have one observation z from a random variable Z with expectation m and variance 1. We may assume that X, Y and Z are independent. The sample means are denoted $\bar{x}$ and $\bar{y}$.

The following estimates of μ are proposed:

$$\mu_1^* = \bar{x} + \bar{y}, \quad \mu_2^* = \bar{x} + 5\bar{y} - 4z.$$

(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)

3. The time in days (and fractions of days) after the first of June until the water temperature in the lake 'Blåvattnet' exceeds 20 degrees (so that it is suitable to go swimming there) is exponentially distributed with expectation μ.

   Eskil believes that $\mu = 20$.

   (a) One year, Eskil has to wait 35 days from the first of June until the water temperature in the lake exceeds 20 degrees. Perform a suitable hypothesis test to find out if this in accord with $\mu = 20$. (2p)

   (b) For $\mu = 40.0$, calculate the power of the test in (a). (3p)

4. A group of five sprinters tests two types of running shoes, Neki and Pamu. One day, they run 100 meters with the Neki shoes, and the next day they run with Pamu. The weather conditions and the shapes of the sprinters are the same for both days. Their times are given in the table below.

Sprinter:	Ed	Beth	Sue	Bob	Usain
Neki	10.27	11.52	11.22	10.75	9.82
Pamu	10.12	11.12	11.27	10.50	9.72

   In terms of achieving the shortest time, are the types of shoes equally good? Try to answer this question by performing a suitable statistical test. Make sure to specify all your assumptions. (5p)

5. During autumn, a researcher randomly selects 20 moose bulls (Swedish: älgtjurar) in the forest "Storskogen", and measures their weights. She then does the same with 15 randomly selected moose bulls in the forest "Lillskogen". The weights of the selected moose bulls in "Storskogen" have mean 495 kg, and the standard deviation is 10. For "Lillskogen", the corresponding mean is 450 kg, with standard deviation 8.

   Let μ_1 be the expected weight for a moose bull in "Storskogen", and let μ_2 be the corresponding for "Lillskogen".

   (a) Calculate a 99% confidence interval for $\mu_1 - \mu_2$. Make sure to specify all your assumptions. (4p)

   (b) On the significance level 1%, can you conclude that the expected weights μ_1 and μ_2 are different for the two forests? Motivate your answer. (1p)

6. In the November 2020 opinion poll by Statistics Sweden, the liberal party (L) got 3.8% of the sympathies. The number of respondents was 4692 people.

   In case of an election, L will have to leave the Swedish parliament if they get less than 4% of the votes.

   In case of an election in November 2020, perform a statistical test to judge if L would have had to leave the Swedish parliament. Motivate your answer. (5p)

7. The number of thunderstorms (Swedish: åskskurar) in Upptuna during summer is assumed to be Poisson distributed with parameter λ.

   One summer, there were 20 thunderstorms in Upptuna.

   (a) Calculate a 95% confidence interval for λ. (2p)

   (b) Calculate a 95% confidence interval for the probability of 25 thunderstorms or more during one summer in Upptuna. (3p)

8. Consider the regression model

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

   where $i = 1, 2, ..., n$ and all ε_i are independent $N(0, \sigma^2)$.

   Data were simulated from this model four times, each time with different values on the parameters α, β and σ^2. The number of observations was $n = 200$. Afterwards, the data were plotted, the parameters were estimated and the coefficients of determination (R^2) were calculated. The plots are shown in figures 1-4 below.

   Match figures 1-4 with the estimated models and corresponding R^2 values. Motivate your solution. (5p)

   Models:

$$\begin{aligned} y_i^* &= 10.4 + 2.0x_i, \quad R^2 = 96\% \quad (a) \\ y_i^* &= 10.3 - 2.0x_i, \quad R^2 = 97\% \quad (b) \\ y_i^* &= 11.4 + 2.0x_i, \quad R^2 = 60\% \quad (c) \\ y_i^* &= 11.4 - 2.0x_i, \quad R^2 = 60\% \quad (d) \end{aligned}$$

GOOD LUCK!

Appendix: figures

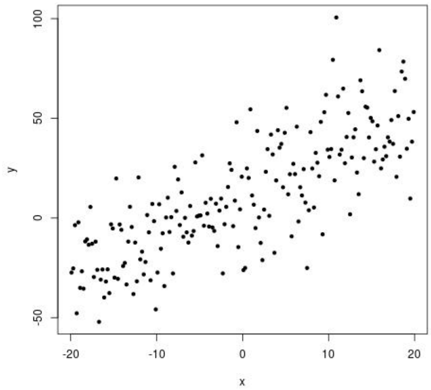

Figure 1: Plot for problem 8.

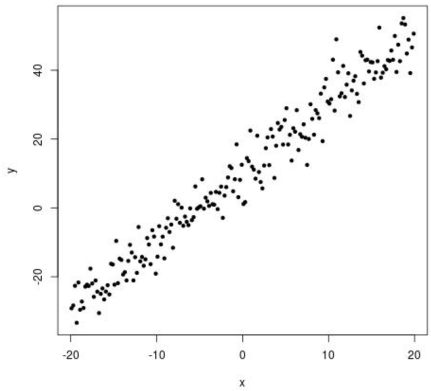

Figure 2: Plot for problem 8.

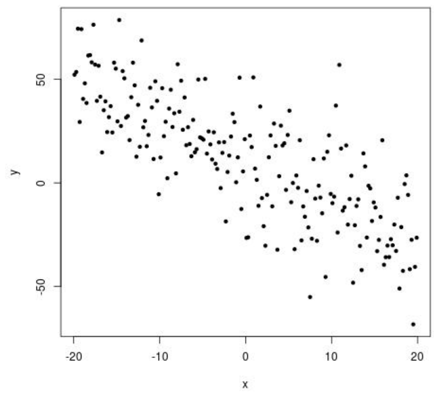

Figure 3: Plot for problem 8.

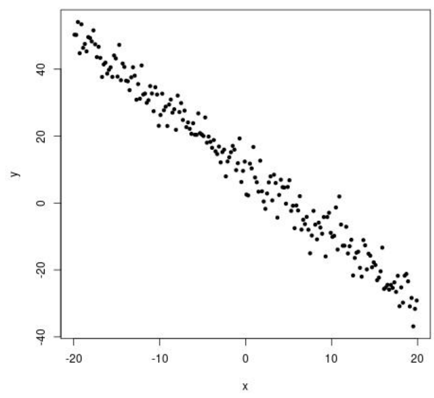

Figure 4: Plot for problem 8.