UPPSALA UNIVERSITY

Department of Mathematics

Rolf Larsson

Exam in Mathematical Statistics

Inference 1 1MS035

2022–01–12

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

Permitted aids: Pocket calculator. Dictionary. Formelsamling för stokastik.

1. We have a random sample 1.7, 1.2, 2.1 from a random variable X with density function
$$f_X(x) = \begin{cases} 2\theta^{-2}x, & 0 \le x \le \theta, \\ 0, & \text{otherwise,} \end{cases}$$
where $\theta > 0$. Estimate θ using

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

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

   (c) maximum likelihood. (2p)

Please turn the page!

2. A statistician wants to investigate the proportion of people (who are assumed to be employed) that take their car to work.

   (a) Suppose that the proportion of people living in average size cities that take their car to work is p. A random sample of 1000 people from such cities is taken, among which x take their car to work. Estimate p by

   $$p_1^* = \frac{x}{1000}.$$

   Show that p_1^* is unbiased for p.   (1p)

   (b) Suppose that the proportion of people who live in big cities that take their car to work is $p - a$, while the proportion that does so of people that live in small cities or at the countyside that do so is $p + a$, where $a > 0$.

   Assume that in a new investigation, the sample is stratified so that a random sample of 250 people who live in big cities is taken, and a random sample of 750 people who live in small cities or at the countryside is taken. (None of the samples contains any people living in average size cities.)

   In the big city sample, y take their car to work, and in the other sample, the number of people taking their car to work is z.

   The estimate
   $$p_2^* = \frac{3y + z}{1500}$$
   is proposed.

   Show that p_2^* is unbiased for p.   (1p)

   (c) Assume that
   $$p < \frac{1}{2}, \quad a > \frac{1}{2}\sqrt{p(1 - p)}.$$

   Then, show that p_2^* is more efficient than p_1^*.   (3p)

3. In his factory at the North Pole, Santa Claus wants to make sure that the proportion of defect toy cars is not too high. He is satisfied if this proportion is lower than 1%.

   He checks if this can be true by standing at the production line, counting the number of produced toy cars until the first defective one comes along. He then observes that the first defective toy car is the 350th one.

   (a) Perform a suitable hypothesis test to conclude if Santa can be satisfied. Choose the 5% level.   (2p)

   (b) Calculate the power of the test in (a) if in fact, 0.1% of the toy cars are defect.   (3p)

Please turn the page!

4. A researcher wants to find out if cows weight differently at farms that produce ecologic milk compared to farms that produce non ecologic milk. From each type of farm, ten cows are randomly selected. Their weights in tons are given in the table below.

Ecologic	1.51	1.23	0.85	1.44	0.97	1.12	1.37	1.22	0.88	1.12
Non ecologic	1.33	0.99	1.45	1.02	0.78	1.05	1.26	1.02	1.28	1.36

Hint: The sample mean and variance of the weights for the ecologic sample are 1.171 and 0.0518, respectively. For the non ecologic sample, the corresponding numbers are 1.154 and 0.0446, respectively.

   (a) Calculate a 95% confidence interval for the difference of mean weights between cows on ecologic and non ecologic farms. Make sure to carefully specify all your assumptions. (4p)

   (b) Can you conclude that cows weight differently at farms that produce ecologic milk compared to farms that produce non ecologic milk? (1p)

5. In the distant ski resort Svartlien, the number of avalanches (Swedish: laviner) per year is supposed to be Poisson distributed with parameter λ. In the last ten years, the numbers of avalanches have been 3, 0, 4, 2, 2, 6, 1, 5, 6, 1.

   (a) Estimate λ. (1p)

   (b) Gudrun says that, according to her experience from 'old days', the mean number of avalanches per year in Svartlien is 2, no more, no less. She thinks it still ought to be this way. Check if Gudrun is right by performing a suitable hypothesis test. (4p)

6. Zerblatt performs 100 independent measurements of the acceleration due to gravity g, in m/s^2, at the planet QZ34. The measurements have mean 15.20 and standard deviation 0.8.

   (a) Calculate a 99% confidence interval for the true value of g at QZ34. (3p)

   (b) A law in physics say that $G = m \cdot g$, where G is the weight of an object measured in Newton (N), and where m is its mass in kg. On the planet QZ34, Zerblatt finds a rock with weight 8.0 N. Calculate a 99% confidence interval for the mass of this rock. (2p)

Please turn the page!

7. At the latest traditional Svartlien New Years eve Slalom competiton (NYS), eight participants got results (in seconds) in two runs on the same track (and with the same snow conditions), according to the table below.

Participant no	1	2	3	4	5	6	7	8
Time run 1	14.5	12.3	24.3	36.2	27.0	17.0	26.6	33.1
Time run 2	13.8	12.7	23.7	29.8	26.5	14.2	25.9	31.5

For potential competitors of this event, does the skill improve (the running time get shorter) from run 1 to run 2? Try to answer this question by performing a suitable hypothesis test.

It is not allowed to assume that the run times are normally distributed. (5p)

8. The numbers of people that emigrated from Sweden during the years 2009-2020 are given in the following table (data from Statistics Sweden).

Year	2009	2010	2011	2012	2013	2014
Number	39 240	48 853	51 179	51 747	50 715	51 237

Year	2015	2016	2017	2018	2019	2020
Number	55 830	45 878	45 620	46 981	47 718	48 937

Is there a trend in this material? Try to answer this question by performing a suitable hypothesis test. (5p)

GOOD LUCK!