

Course of Study Bachelor Computer Science	Exercises Statistics WS 2022/23
Sheet VII	

Statistical Inference

- In an urn there is an unknown number N of balls numbered from 1 to N . The number of N should be estimated. A ball from the urn is used for this purpose and his number is noted. Describe the random variable X = the number of the drawn ball.
 - Determine the distribution of X depending on N . Calculate the expected value and variance of X .
 - Show that $T(X) = 2X - 1$ is an unbiased estimator for N is.
 - Calculate for $N = 4$ and $N = 5$ the probability for N to be exactly estimated at T .
 - Calculate the variance of T .

Maximum Likelihood Estimation

- A ticket inspector checks for Frankfurt S-Bahn lines the tickets from the passengers. He keeps checking until he sees a passenger without valid ticket. He then collects the increased fare and starts after a break with a new check of the tickets.

For 10 such check runs, he shall have

42 50 40 64 30 36 68 42 46 48

until he have found a non valid ticket.

Determine a maximum likelihood estimator based on the given numbers for p share of nonvalid tickets among all checked ticktes.

- A device consists of the components K_1, K_2 and K_3 . The device becomes defective as soon as one or more of the components is defective. The lifetimes L_1, L_2 and L_3 (in h) of the three components are independent random variables.

The distribution function of L_1 is $F_1(x) = \begin{cases} 1 - e^{-\lambda x} & \text{für } x \geq 0 \\ 0 & \text{sonst} \end{cases}$

The distribution functions of L_2 and L_3 are $F_2(x) = \begin{cases} 1 - e^{-\lambda \sqrt[3]{x}} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$.
 λ is an unknown parameter > 0 .

- (a) Calculate the distribution function and density for the lifetime S of the device.
- (b) When measuring the lifetime of randomly from production of the devices removed resulted in following values in hours:

82.2 94.0 122.5 95.8 106.4

Use a maximum likelihood estimator to determine the an estimate for λ .

3. To determine the number of N of red deers living in a precinct region 7 red deer were caught and marked in a trapping action. Afterwards the animals were again released. After a certain time, another trapping action was started. Thereby 3 red deer were caught, whereby 2 already were marked. It is assumed that between is no influx or outflow of red deer in the region and that the animals were able to pass the region within a short period of time.
 - (a) Determine a maximum likelihood estimator for the total number N of the red deer living in the region.
 - (b) A third trapping action started, where 8 red deers were caught. 4 of them were marked. What is no the maximum likelihood estimation of N ?

Confidence Intervals

1. Strictly speaking, what is the correct interpretation of a 95% confidence interval for the mean?
 - ☐ If repeated samples were taken and the 95% confidence interval was computed for each sample, 95% of the intervals would contain the population mean.
 - ☐ A 95% confidence interval has a 0.95 probability of containing the population mean.

- ☐ 95% of the population distribution is contained in the confidence interval.
2. A population is known to be normally distributed with a standard deviation of 2.8.
- (a) Compute the 95% confidence interval on the mean based on the following sample of nine: 8, 9, 10, 13, 14, 16, 17, 20, 21.
 - (b) Now compute the 99% confidence interval using the same data.
3. You take a sample of 22 from a population of test scores, and the mean of your sample is 60.
- (a) You know the standard deviation of the population is 10. What is the 99% confidence interval on the population mean?
 - (b) Now assume that you do not know the population standard deviation, but the standard deviation in your sample is 10. What is the 99% confidence interval on the mean now?

Hint: Assume that the test scores follow a normal distribution.

4. Calculate for the below given sample from a normally distributed population the 95% confidence intervals
- (a) for the mean, if the standard deviation is 2
 - (b) for the mean, if the standard deviation is unknown
 - (c) for the variance, if the mean is 250
 - (d) for the variance, if the mean is unknown

x_i : 247.4, 249.0, 248.5, 247.5, 250.6, 252.2, 253.4, 248.3, 251.4, 246.9, 249.8, 250.6, 252.7, 250.6, 250.6, 252.5, 249.4, 250.6, 247.0, 249.4

5. At a telemarketing firm, the length of a telephone solicitation (in seconds) is a normally distributed random variable with mean μ and standard deviation σ , both unknown. A sample of 51 calls has mean length 300 and standard deviation 60.
- (a) Construct the 95% confidence upper bound for μ .
 - (b) Construct the 95% confidence lower bound for σ .

6. At a certain farm the weight of a peach (in ounces) at harvest time is a normally distributed random variable with standard deviation 0.5. How many peaches must be sampled to estimate the mean weight with a margin of error ± 0.2 and with 95% confidence.
7. You read about a survey in a newspaper and find that 70% of the 250 people sampled prefer candidate A.
 - (a) Compute the 95% confidence interval.
 - (b) You are surprised by this survey because you thought that more like 50% of the population preferred this candidate. Based on this sample, is 50% a possible population proportion?
8. A researcher was interested in knowing how many people in the city supported a new tax. He sampled 100 people from the city and found that 40% of these people supported the tax. What is the upper limit of the 95% (one-side) confidence interval on the population proportion?
9. An advertising agency wants to construct a 99% confidence lower bound for the proportion of dentists who recommend a certain brand of toothpaste. The margin of error is to be 0.02. How large should the sample be?
10. The interval $[45.6, 47.8]$ is a symmetric 99% confidence interval for the unknown parameter μ based on a sample x_1, \dots, x_{10} from a normal distribution $N(\mu, \sigma^2)$ with unknown σ . Calculate the sample mean \bar{x} and the sample standard deviation s .
11. The waiting time at the pay desk of a certain supermarket is normally distributed with mean waiting time μ and known standard deviation $\sigma = 1,8$ minutes. A confidence interval for the mean waiting time (in minutes) for this supermarket is $[5.12; 8.32]$. If the sample size is $n = 10$, what is then the confidence level?