Exercises:
Estimation and Inference

Statistics for Data Science
Master Program in Advanced Analytics
2025/2026

1. Consider the population density function of a Poisson distribution:
mass

$$f(x \mid \lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \tag{1}$$

and the following sample: (95, 100, 80, 70, 110, 98, 97, 90, 70).
Obtain the maximum likelihood point estimator and point estimate for λ.

2. Consider a random variable X whose distribution depends on the parameters α and θ, for which we have $E(X)$ = $\alpha\theta$ and Var(X) = $\alpha\theta^2$. Knowing that in a sample of 320 observations we have $\sum_{i=1}^{320} x_i = 22.2$ and $\sum_{i=1}^{320} x_i^2 = 535.8$, present a plausible estimate for the unknown parameters.
Hint: Use the method of moments and consider Var(X) = E(X^2) - (E(X))^2.

3. Classify each of the following statements as true or false.

F (a) The primary goal of point estimation is to find a range of possible values for a population parameter.

T (b) Both the method of moments and MLE can be used to estimate parameters for discrete and continuous probability distributions.

F (c) For estimating the parameter of a Poisson distribution using the method of moments, we need to solve two equations.

T (d) Maximum likelihood estimation (MLE) aims to find parameter values that maximize the likelihood function, which measures the probability of observing the given data for a given set of parameter values.

F (e) The parameter space, Ω, defines the range of possible values for the observed values of a random sample.

4. For each of the following propositions, state whether it is a statistical hypothesis:

   (a) $\mu = 3$

   (b) $\bar{x} = 4$

   (c) $\sigma < 3$

   (d) $\bar{X} < 3$

5. An auditor assumes that the average value of the accounts to be received in a certain company is €750. To support his reasoning, he proposes to take a sample of 36 accounts and calculate their average. He will only reject the proposed value if the sample average clearly contradicts it, thus giving the benefit of the doubt to the procedure to be carried out.

   (a) Define the hypothesis and critical region for a test with a significance level of 0.05.

   (b) Assuming that in the sample it was observed $\bar{x} = 800$ and $s^2 = 62000$, test if it is appropriate to reject the proposed value, based on the observed sample, for the variance.

6. A factory produces a certain synthetic material whose breaking strength follows a random variable with a mean of 250 kg. Due to the high cost of the current production process, the factory administration is considering a new, more cost-effective process. However, the factory aims to ensure that the new process maintains a breaking strength at least as high as the current process. The factory administration decided to collect a sample of 25 units manufactured by the new process and proceed to test $H_0 : \mu \geq 250$ against $H_1 : \mu < 250$ with a significance level of 0.025.

   (a) Given the proposed test, do you think the company is more inclined to adopt the new process or maintain the old one?

(b) Assuming that in the sample it was observed a mean strength of the material of 245 kg and a corrected standard deviation of 6 kg, calculate the test value and conclude whether or not to adopt the new manufacturing process.

7. Consider a study to determine whether a sales tax on soda will reduce consumption of soda in the US below the current per-capita level of about 50 gallons of soda per year. The hypotheses for the test are $H_0 : \mu = 50$ vs $H_1 : \mu < 50$, where μ represents the average annual consumption of soda in communities where the sales tax is implemented.

   (a) Suppose sample results give a p-value of 0.02. Interpret this p-value in terms of random chance and in the context of taxes and soda consumption.

   (b) Now suppose sample results give a p-value of 0.41. Interpret this p-value in terms of random chance and in the context of taxes and soda consumption.

   (c) Which p-value, 0.02 or 0.41, gives stronger evidence that a sales tax will reduce soda consumption?

8. Researchers conducted a study of smartphone use among adults. A cell phone company claimed that iPhone smartphones are more popular with young individuals (less than 25 years old) than with older individuals (more than 25 years old). The results of the survey indicate that of the 232 older individuals cell phone owners randomly sampled, 5% have an iPhone. Of the 1,343 young individuals cell phone owners randomly sampled, 10% own an iPhone. Test at the 5% level of significance. Is the proportion of young iPhone owners greater than the proportion of older iPhone owners?

9. Classify the statements below as true or false, justifying briefly.

   (a) In a test with significance level $\alpha = 0.05$, a p-value of 0.078 leads to not rejecting H_0.

   (b) When in a hypothesis test a p-value of 0.05 is obtained, one is more confident in rejecting H_0 than when the p-value is 0.01.

   (c) If in the test $H_0 : \mu = \mu_0$ against $H_0 : \mu < \mu_0$, referring to a population with known variance and mean μ_0, the p-value is 0.05, then $P(Z > Z_{\text{obs}} \mid \mu = \mu_0)$ is 0.95.

   (d) In a hypothesis test, the p-value does not depend on the test's significance level.

10. Consider the dataset (already available in R!) `mtcars`.

    (a) Test the hypothesis that the average miles per gallon for car in this dataset is 20 mpg against its two-sided alternative.

    (b) Test the hypothesis that average horsepower is greater than 90.

    (c) Test the hypothesis that the average weight is 2.5.

    (d) Find a 95% confidence interval for the average horsepower (`hp`).

    (e) Test that the average miles per gallon is the same for cars for each transmission type (i.e., manual vs automatic).

11. Classify each of the following statements as true or false.

    (a) A narrower confidence interval indicates greater precision in estimating the population parameter.

    (b) A Type II error occurs when you incorrectly reject the null hypothesis when it is true.

    (c) The p-value is the probability of making a Type I error when the alternative hypothesis is true.

    (d) Increasing the sample size generally leads to wider confidence intervals.

    (e) The power of a statistical test is the probability of correctly rejecting the null hypothesis when it is true.

    (f) The critical region is the range of sample values that leads to the rejection of the null hypothesis.

12. Consider a situation in which a restaurant chain will test for arsenic levels in a sample of chickens from a supplier. If there is evidence that the average level of arsenic is over 80 ppb, the chain will permanently cancel its relationship with the supplier.

    (a) What are the underlying hypothesis?

    (b) What would it mean for analysts at the restaurant chain to make a Type I error in the context of this situation?

    (c) What would it mean to make a Type II error in this situation?

13. If the power of a statistical test is increased, for example by increasing the sample size, how does the probability of a Type II error change?

14. A study was conducted by a retail store to determine if the majority of their customers were teenagers. With $\hat{p}$ = 0.48, the null hypothesis was not rejected and the company concluded that they did not have evidence that the majority of their customers were teenagers. But, in reality, the proportion of all of their customers (i.e., the population) who are teenagers is actually p = 0.53 . Did this research study result in a Type I error, Type II error, or correct decision?