Statistical data analysis

Statistical Fundamentals – Sample Exam Questions.

1. Define the concept of *expected value* of a random variable and *arithmetic mean*. Explain the difference between them.

2. Consider a random vector $\mathbf{X}$. Define the *covariance* and *correlation matrix*. What properties do these matrices have? What are their applications?

3. What is the *cumulative distribution function*? What is a *quantile* and *quantile function*? What is the *probability density function*? Define the terms formally and describe the relationship between them.

4. Explain in your own words the meaning of the *p-value* in a statistical test. What is the interpretation of the *significance level* α? Assume that the p-value of a certain test is 0.045. What is the probability that the null hypothesis of this test is false? How does this relate to α?

5. Explain the concept of a *confidence interval*. Describe as precisely as possible how you would calculate a 99% confidence interval for the estimate of the mean of a normal distribution from a small sample of size m.

6. Explain the meaning of the terms *Type I error* and *Type II error* used to describe specific errors in the process of statistical hypothesis testing. Explain how these errors are related. Describe how the occurrence of these errors can be influenced.

7. Formulate the *central limit theorem*. Where can it be applied?

8. Motivate the introduction of the *Student's t-distribution*. To which distribution does the t-distribution asymptotically approach as the number of degrees of freedom increases? Explain under what circumstances and how it differs from this distribution. What is it used for?

9. Define the term *likelihood function*. What is the purpose of the maximum likelihood method? Name at least two methods used for maximizing likelihood. Explain why the logarithm of likelihood is often used. How can likelihood be utilized in hypothesis testing?

10. Formally define the *multivariate normal distribution*. How many parameters do we generally need to describe this distribution in d dimensions? Can the number of these parameters be limited in any way? What consequences might this limitation have?

11. Define at least two commonly used distributions of discrete random variables. Name their parameters. Provide examples to indicate when these distributions can be used.

12. Explain the difference between parametric and non-parametric statistical tests. Name the basic advantages and disadvantages of both approaches. Provide at least one example of a parametric test and one example of a non-parametric test.

13. For the following statistical problem, assign an appropriate test. Eleven cars of a certain brand were tested. Check whether their right and left front tires wear down comparably. (Assume that the tire wear [mm] follows a normal distribution. From the following tests, choose the best one and explain your choice: two-sample t-test, Friedman test, one-sample t-test, one-sample Wilcoxon test, test of an alternative distribution parameter, test of variance of a normal distribution, paired t-test). What would happen if you used the second best test from the given list?