Biostatistics Course requirements

This document is a slightly modified version of a document that was originally developed by Alexander Ploner who taught the course in the past.

Biostatistics is an advanced-level course in the Master's Programme in Biomedicine at Karolinska Institutet, not an introductory course. Although it starts with a short recapitulation of elementary statistical tools and concepts, the general expectation is that the students are already familiar with most of them. The knowledge acquired in the Biomedicine Bachelor Programme at KI suffices.

The students are expected to be familiar with the following concepts:

- Analyzing biomedical data: Descriptive statistics: graphical (bar plots, histograms, boxplots) and numerical (mean, standard deviation, median, quantiles)
- Elementary probability: definition, elementary calculations, conditional probability) and distributions (normal, binomial)
- Sampling and inference: sampling distribution and standard errors (as general concepts and for the mean), central limit theorem, confidence intervals (of means)
- Hypothesis testing: test for means, tests for proportions
- Simple linear regression (one dependent, one independent variable): definition, parameter estimation, confidence intervals for parameters, hypothesis tests of parameters and the model, prediction

Self-assessment test

Answer the following questions and rate yourself based on the guide at the end of this document.

1. The sample variance is the average of squared differences between the observations and the mean.

(a) True (b) False

2. The sample mean is always larger than the sample median.

(a) True (b) False

3. The larger the sample, the closer to normal is the distribution of the data.

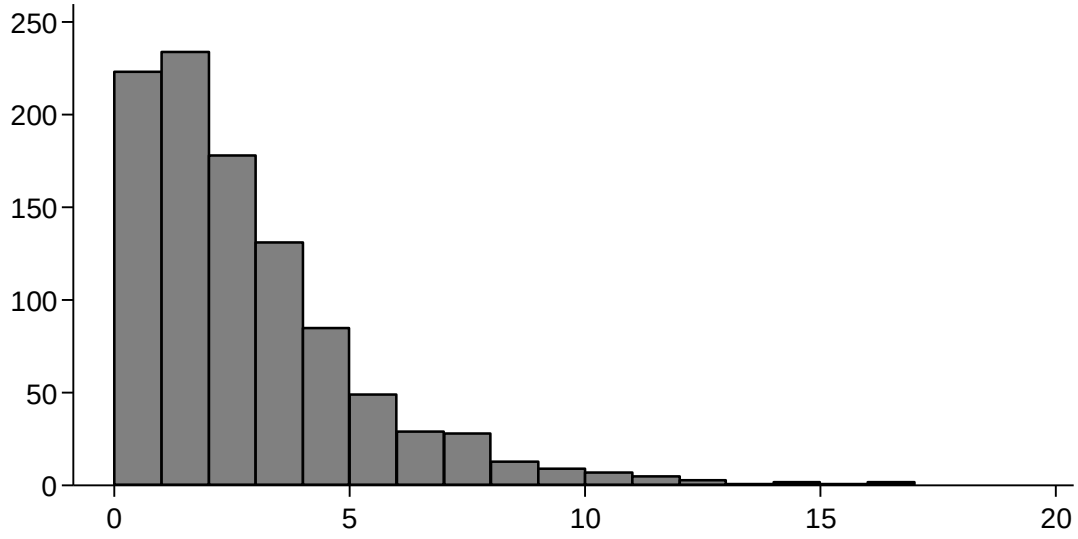
(a) True (b) False

4. The larger the sample, the closer to normal is the distribution of the sample mean.

(a) True (b) False

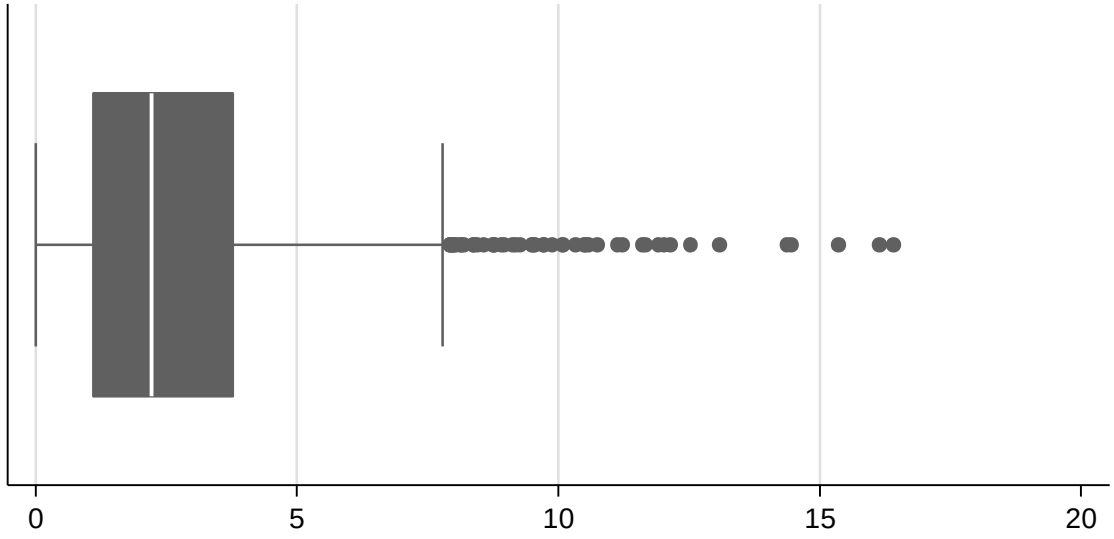
5. The median of the below histogram is

(a) smaller than the mean (b) approximately equal to the mean (c) larger than the mean



6. The data in the below box plot are

(a) normally distributed (b) symmetric (c) skewed



7. The probability of getting 10 heads in a row when throwing a fair coin is about

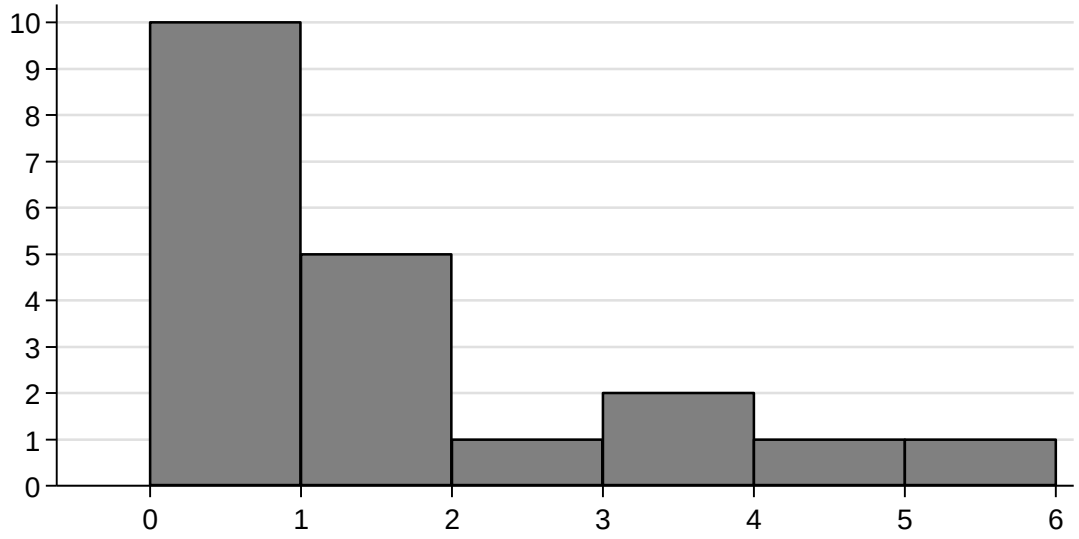
(a) 0.1 (b) 0.01 (c) 0.001 (d) 0.00001 (e) 0.00000001

8. The standard error of the mean is equal to the standard deviation of the sample of mean when repeating the underlying experiment an infinite number of times.

(a) True (b) False

9. A possible value for the median of the below histogram is

(a) 0.0 (b) 0.9 (c) 1.5 (d) 3.0 (e) 5.1



10. The correlation coefficient between variables X and Y is the scaled slope of the regression line

(a) of Y on X (b) of X on Y (c) both (d) neither

11. What is the most appropriate graphical tool for checking the normal distribution of a data set?

(a) boxplot (b) histogram (c) quantile plot

12. An approximate 95% confidence for the mean can be calculated as the sample mean plus or minus

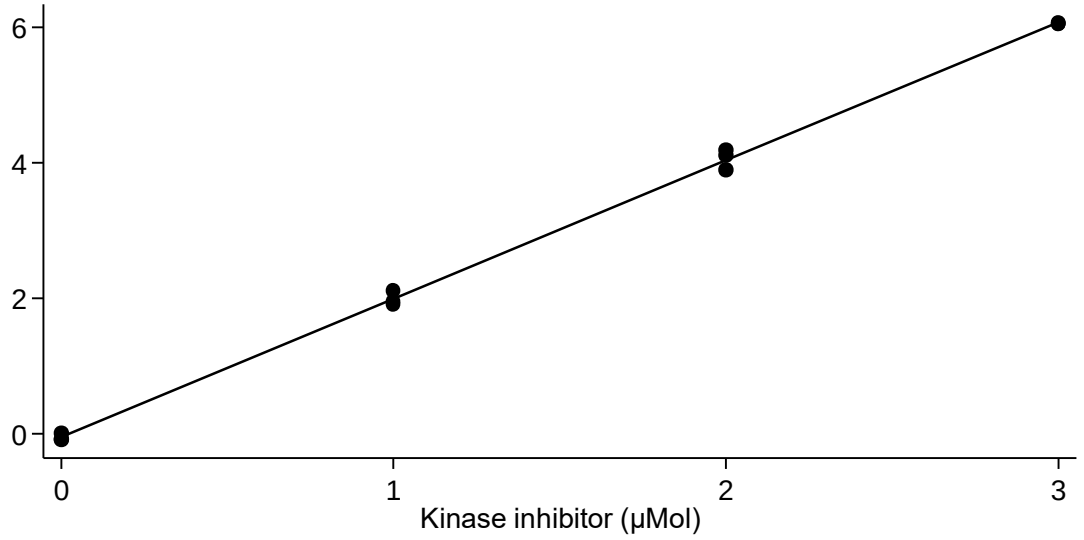
(a) two times (b) three times (c) 1.5 times the standard error

13. Researchers compared the average proliferation index of lymphocytes extracted from the colon of patients with irritable bowel disease (IBS, n=8) and healthy controls (Ctrl, n=7). They use a t-test and obtain a p-value of 0.033. This test is

(a) suitable with statistically significant result
(b) suitable with statistically non-significant result
(c) unsuitable for comparing two group means

14. Researchers studied the effect of adding a kinase inhibitor to an epithelial cell culture on its proliferation index for four different concentrations of the inhibitor and three replicates each. The total number of measures was 12. The plot below shows both the raw data and the fitted regression line. What is the correct 95% confidence interval for the slope of the regression line?

(a) [–4.3, –1.5] (b) [–0.05, 0.21] (c) [0.05, 1.21] (d) [1.98, 2.08]



Rate yourself

After going through the above questions, place yourself in one of the following groups.

1. I recognize all or most of the terms and concepts, I understand the questions, and I am confident that I have answered all or most questions correctly. You are well prepared for the course.

2. I recognize most of the terms and concepts, I understand most of the questions, and I am fairly sure about the majority of my answers. With my lecture notes/stats book, I am confident that I could figure out the rest in short time. You are obviously well qualified to take the course. The lectures of the introductory module should fill in any gaps.

3. It has been a while since my last statistics class, and while I recognize a majority of the terms and concepts, I am not confident at all about my answers. Some preparatory reading of your lecture notes from previous courses or from the literature listed at the course web page would be desirable.

4. I have not had any formal statistics training am completely lost with the questions. Please contact me at matteo.bottai@ki.se