Universität Bielefeld
Tanja Pasurek
Ehsan Abedi WiSe 2024/25

Foundations of Statistics

Homework 13

Exercise 1. By throwing a coin 20 times we would like to check if it is a fair one. As a test statistic we use the number of heads.

(a) Formulate the hypotheses of a suitable test problem.

(b) What is the exact distribution of the test statistic assuming H_0? Plot it with `R` and mark the critical region at level $\alpha = 0.05$.

(c) Derive an approximately normal distributed test statistic using the CLT. For that statistic calculate the critical region under the above conditions and compare it with (b).

(d) We now throw the coin 20 times and observe 4 times head. Perform the test in (b) and (c) and calculate the associated p-values.

Exercise 2. A market research institute conducts annual surveys on living costs. In the past the price of a basic shopping cart was 600 Euro in average. Years of experience show that the standard deviation of that price is 15. We assume that those costs of the shopping cart are normally distributed.

(a) We would like to show that the shopping cart got more expensive in the last year. Formulate suitable hypotheses for a test problem.

(b) Last year 40 shopping carts gave an average price of 605 Euro. Should we reject H_0 of (a) at level 0.05?

(c) How big has the sample size to be, such that an observed increase of 5 Euro of the average price at level 0.01 can be called signifant?

Exercise 3. The following data represent the height to width ratio of rectangles used for decorating purposes by some ancient peoples.

0.693	0.749	0.654	0.670	0.662	0.672	0.615	0.606	0.690	0.628
0.668	0.611	0.606	0.609	0.601	0.553	0.570	0.844	0.576	0.933

Is it reasonable that they followed the *golden ratio* $(\sqrt{5}-1)/2$?

To examine this, perform an empirical bootstrap for the data. That is, for 10 000 bootstrap samples $x_1^*, \ldots, x_n^*$ we compute the standardized mean

$$t^* = \frac{\bar{x}_n^* - \bar{x}_n}{s_n^*/\sqrt{n}}.$$

From that collection of t^* calculate a confidence interval at level $\alpha = 5\%$. Calculate the corresponding p-value. What is your test decision?

Exercise 4. (Distribution of p-value).

It is important to note that the p-value is a function of the chosen test statistic $T(X_1, ..., X_n)$ and is therefore a random variable. Here we would like to study its (empirical) distribution using an example.

(a) In `R` draw a sample of the standard normal distribution of size 100. Use the built-in function `t.test()` to test the hypothesis $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$ at the significance level 5%.

(b) Repeat (a) a thousand times and store the resulting p-values in a vector using the command `t.test(x,mu=0)$p.value`. Now plot the empirical distribution function of those p-values and explain the result.

Exercise 5. One is given a number t, which is the realization of a random variable T with an $\mathcal{N}(\mu, 1)$ distribution. To test $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$, one uses T as the test statistic. One decides to reject H_0 in favor of H_1 if $|t| \geq 2$.

(a) Compute the probability of committing a type I error.

(b) Compute the probability of committing a type II error if the true value of μ is 1.

Exercise 6. (Is normal body temperature 98.6°F (37°C)?) The data `normtemp (Using R)` contains measurements of 130 healthy, randomly selected individuals. The variable `temperature` contains normal body temperature. Does the data appear to come from a normal distribution? If so, perform a t-test to see if the commonly assumed value of 98.6°F is correct.

The following exercise will be solved in the tutorial class.

Exercise 7*. (Is π a normal number?) The `pi2000 (Using R)` data set contains the first two thousand digits of π. Perform a chi-squared significance test to see if the digits appear with equal probabilities.