

Additional Tasks WS 23

Descriptive Statistics

1. Import “corona.csv”.
2. Specify type and scale of all variables
3. New cases and new deaths: october, germany
4. Number of new cases and new deaths in every country per month
5. Max, min, mean and median of new cases and new deaths per month
6. 7 Day incidences in germany 2020
7. Linear regression new deaths = $a + b * \text{new cases in Germany}$
 - (a) Parameter a, b
 - (b) Scatterplot
 - (c) Coefficient of correlation, coefficient of determination
 - (d) Prediction for 20000 new cases
 - (e) Interpretation of the coefficients
8. Import csv-file “corona_march_2020.csv”; Why is the dataset not tidy?
Make the dat set tidy.

Probability

1. 100 Students in the graduating classes can choose between friday, saturday and sunday for their graduating ceremony. If all students choose one day and the preferences of the days are
friday: 0.4, saturday: 0.5, sunday: 0.1
what is the probability that
 - (a) more than 40 students sign up for saturday
 - (b) 30 students sign up for friday, 50 students sign up for saturday, 20 students sign up for sunday
 - (c) that the 20th student is first student choosing sunday
 - (d) The room for the ceremony has 80 places. What is maximum number graduating students if the probability that the room is big enough is 0.9?
2. A snack bar offers sausage (3 Euro), hamburger (4 Euro), chips (2 Euro), sandwich (2.5). The owner of the snack bar assumes that a customer orders only one dish and the probabilities of the dishes are sausage: 0.2, hamburger: 0.4, chips: 0.25, sandwich: 0.15.
 - (a) What is the probability that the turnover of 100 customers ist at least 300 Euro?
 - (b) What is the maximum amount of the turnover of 100 customers, which will be achieved with a probaility of 0.95?

Inferential Statistics

1. Consider the sample

0.92,0.83,0.81,0.70,0.88,0.73,1.05,0.91,0.83,0.67,0.94,0.90,0.91,
0.83,0.84,0.96,0.87,0.91,0.98,0.84,0.88,0.76,0.99,0.89,0.82

These are values of independent normally distributed random variables with $\sigma = 0.2$.

- (a) Determine a confidence interval of the expected value at the 99% level.
- (b) Find the length of the confidence interval.
- (c) The length of the confidence interval should be 0.15. Find
 - i. how many sample values are necessary.
 - ii. an appropriate confidence level
- (d) Assume now that the population variance is unknown. Conduct a suitable statistical test at a 5% level, to check whether the mean is greater 0.9.

2. A company produces chocolate bars with a standard weight of 100 gr. As a measure of quality controls he weighs 15 bars and obtains the following results:

98.32, 97.26, 99.85, 99.52, 95.73, 95.56, 100.49, 98.19, 95.16, 98.26,
96.46, 100.23, 99.76, 98.58, 97.43

- (a) What is an appropriate hypothesis regarding the expected weight for a two-sided-test?
- (b) If weights can be assumed to be normally distributed, which test should be used to test these hypothesis?
- (c) Conduct the test that was suggested to be used in b) at a 5% level. What is your test decision. Specify the p-value.
- (d) Based on the sample, the producer changes the settings in production. To check whether the correction has led to an improvement, he again takes 15 chocolate bars and weighs them.

100.14, 100.05, 96.51, 98.70, 98.22, 101.06, 103.55, 100.16, 100.60,
102.85, 103.15, 100.66, 102.52, 102.09, 100.84

What is an appropriate hypothesis for comparing the expected weights of the two samples?

- (e) Before testing the hypothesis using a test, first verify that the variances of the populations of the two samples are equal using an appropriate test at the 10% level.
- (f) Provide an appropriate test to test the hypothesis and perform at the 5% level. What is your test decision. Specify the p-value.