**Exercise 5.1**

As hinted in the Assigment Description, the sign test will be used for the tests in this exercise, by using the binomial test in R. Denote for each observation in the grades sample. Note that the sample size for the given sample.

a)

The sign test assumes that the underlying distribution has a unique median, however, since , we delete from the sample, and apply the sign test to the rest of the data.

The **test statistic** is

The distribution of the test statistic under is

The test score for the sample is 16. Using the corresponding binomial test (see R code), this gives an **p-value** of 0.941.

**Conclusion:** Since this p-value is greater than the significance level, we do not reject the null hypothesis. Therefore we conclude that we do not have evidence that shows that the median from the underlying distribution is (with significant probability) greater than 6.

b)

The sign test assumes that the underlying distribution has a unique median. Since none of the sample data is equal to 6.5, we do not have to adjust the data and sign test.

Testing this is the same as testing both one-sided variants for . is now rejected, if either of the one-sided tests rejects.

The **test statistic** is

The distribution of the test statistic under is

*Right-tailed part*

Note: for the sake of brevity, we will not list the test statistic and distribution again.

The test score for the sample is 8. Using the corresponding binomial test, this gives an p-value of 1.000.

Since this p-value is greater than the significance level, we do not reject the right-tailed part of the (original) null hypothesis.

*Left-tailed part*

The test score for the sample is 8. Using the corresponding binomial test, this gives an p-value of 0.000.

Since this p-value is lower than the significance level, we do reject the left-tailed part of the (original) null hypothesis.

**Overall conclusion:** Since one of the one-tailed versions (with ) of the two-tailed test rejects the null hypothesis, we reject the original null hypothesis ). Therefore we conclude that (with significant probability) the median of the underlying distribution is not equal to 6.5.

c)

We will use the one-sided binomial test for this:

Since we have a total of (=42) observations, the distribution of the binomial test statistic under is

We have a total of observed successes (amount of grades larger or equal than 7).

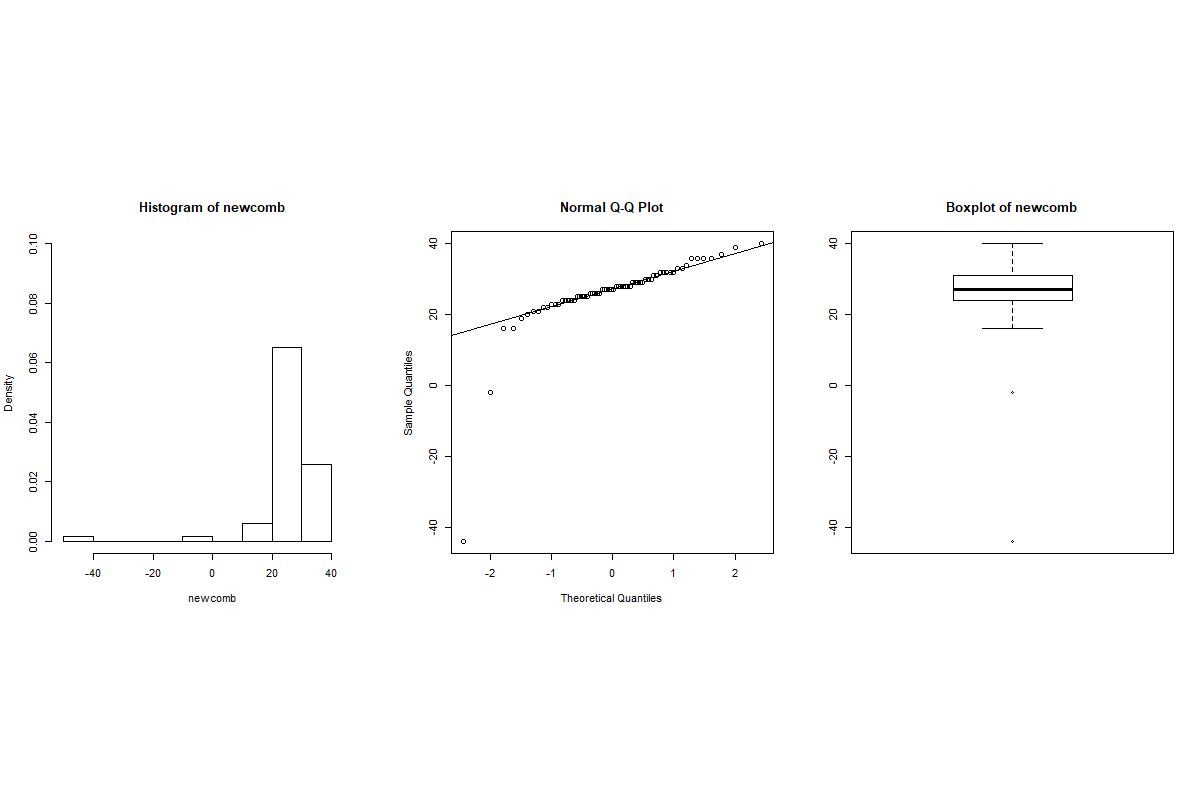
Using the binomial test in R, the corresponding one-sided **p-value** is 0.000 (rounded).

**Conclusion:** Since this p-value is lower than the significance level, we reject the null hypothesis. Therefore we conclude that the probability to get a grade of at least 7 is (probably) lower than 35%.

**Exercise 5.3**

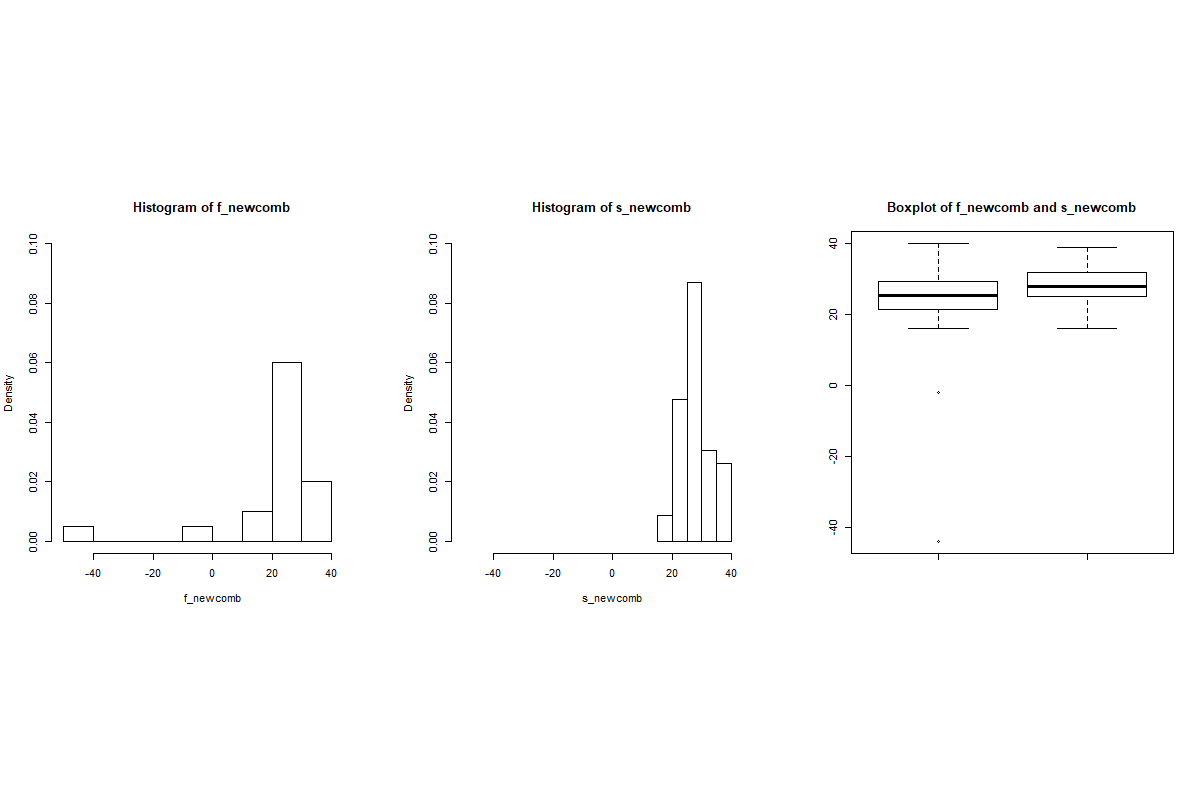
**Whole dataset (newcomb)**

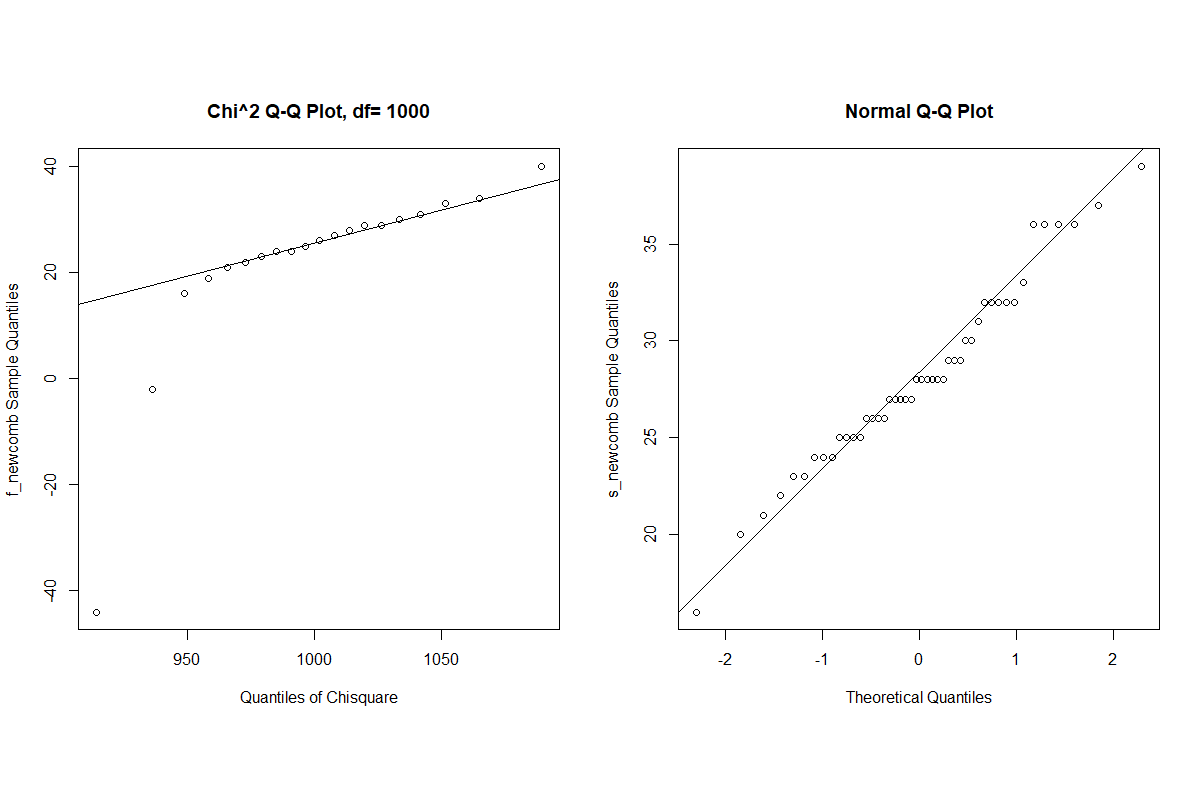
See below for a graphical summary of the data as a whole.



**Comparing First 20 observations (f\_newcomb) with last 46 observations (s\_newcomb)**

See below for a graphical summary (separated for improved readability)





**Estimate for the difference**

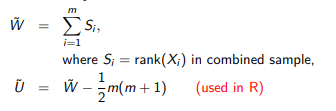
Our estimate of choice is: (f\_newcomb) (s\_newcomb). The computed value is -2.5. The reason for this is that the median is a better alternative than e.g. the mean for skewed distributions (f\_newcomb).

Test for (stochastic difference)

We have performed different tests for testing the difference between the observations (see the R code). However, we think a very suitable test is the Wilcoxon two-sample test. The reason for this being that other two-sample tests are less appropriate for various reasons (t-test: not appropriate because of non-normal data, median test: less powerful, KS-test: not appropriate because of large ties, permutation test: not appropriate because of dependent observations).

We choose a one-sided left-tailed test, inspired by the estimate above. Denote the underlying distribution for f\_newcomb and the underlying distribution for s\_newcomb.

The **test statistic** is 

The distribution of the test statistic under is 

The computed test score (called W in R) is 348. This corresponds to a (non-exact, because of ties) **p-value** of 0.059.

**Conclusion:** Since this p-value is greater than the significance level, we do not reject the null hypothesis. Therefore we conclude that we do not have evidence that shows that a significant (left-sided) difference, i.e. that the underlying distribution of f\_newcomb is stochastically smaller than s\_newcomb. However, since the p-value is rather close to the significance level, from our conclusion it does not follow that there is absolutely no “(stochastic) difference”.