

Assignment 1

Group 6 – Kaan Kara, Tommy Maaiveld, Krishnakanth Sasi

Exercise 1

For x_1 , the data looks similar to samples generated from a normal distribution in both graphs and seems normally distributed (the histogram is less relevant here because of the small sample size).

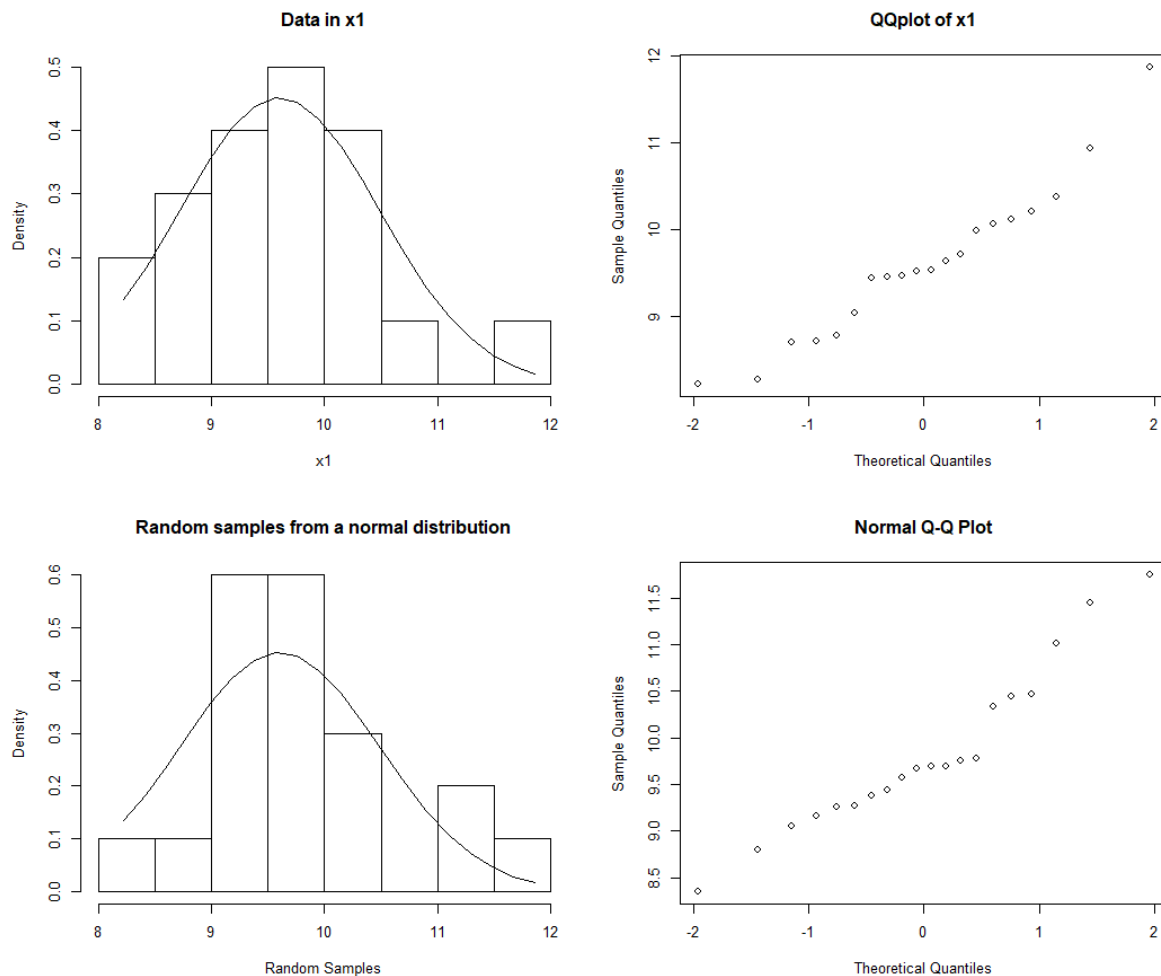


Figure 1. Comparison between data in vector x_1 (top) and random samples from a normal distribution (bottom).

For x_2 , the data appears to be uniformly distributed.

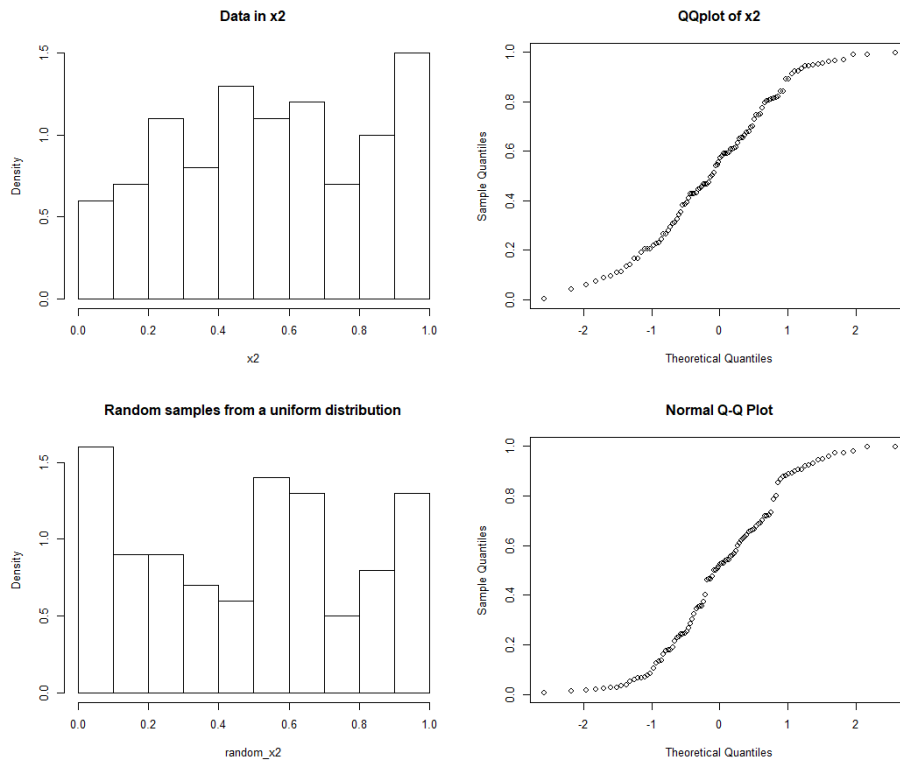


Figure 2. Comparison between data in vector `x2` (top) and random samples from a uniform distribution (bottom).

For `x3`, the data appears to be normally distributed.

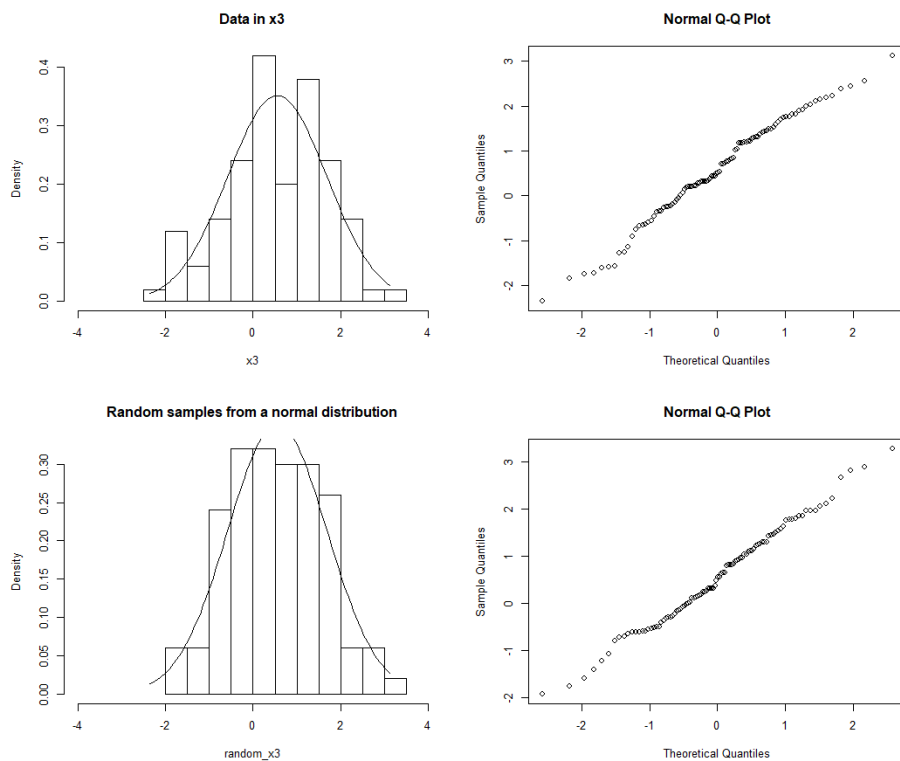


Figure 3. Comparison between data in vector `x3` (top) and random samples from a normal distribution (bottom).

For x4, the data seems normally distributed; this is difficult to see in the histogram.

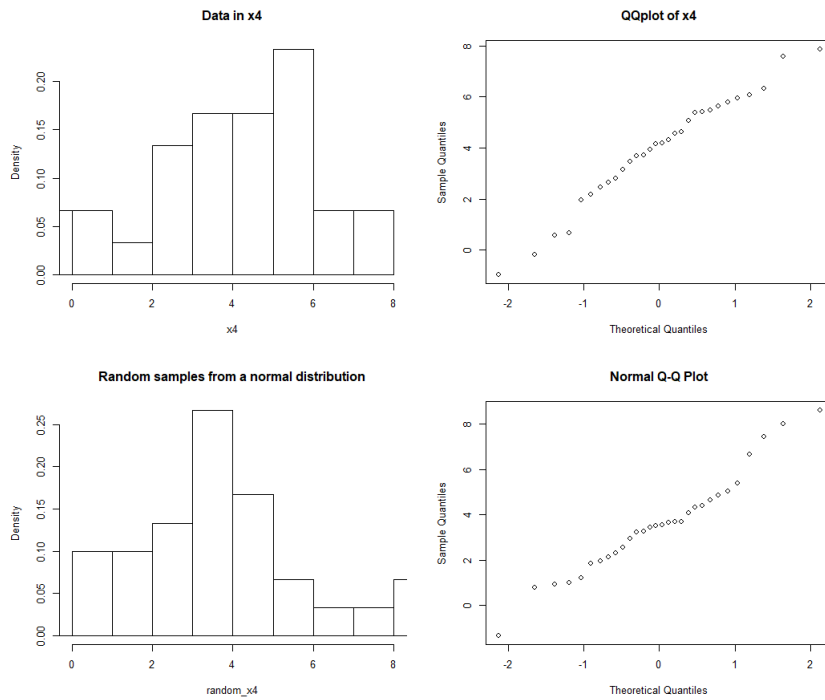


Figure 4. Comparison between data in vector x4 (top) and random samples from a normal distribution (bottom).

For x4, the data seems to follow an exponential distribution with a right-tailed skew. The distribution is close to the chi-squared distribution (shown in Figure 5).

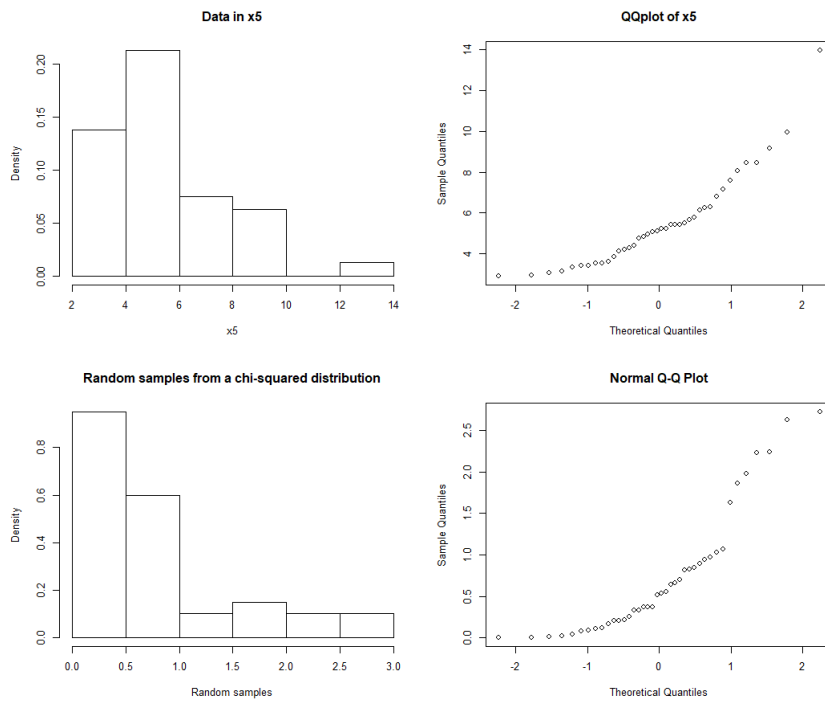


Figure 5. Comparison between data in vector x4 (top) and random samples from a chi-squared distribution (bottom).

Exercise 2

1. 4.3% of the p-values are smaller than 5%, and 7.9% of the p-values are smaller than 10%. The histogram shows that the p-values are normally distributed

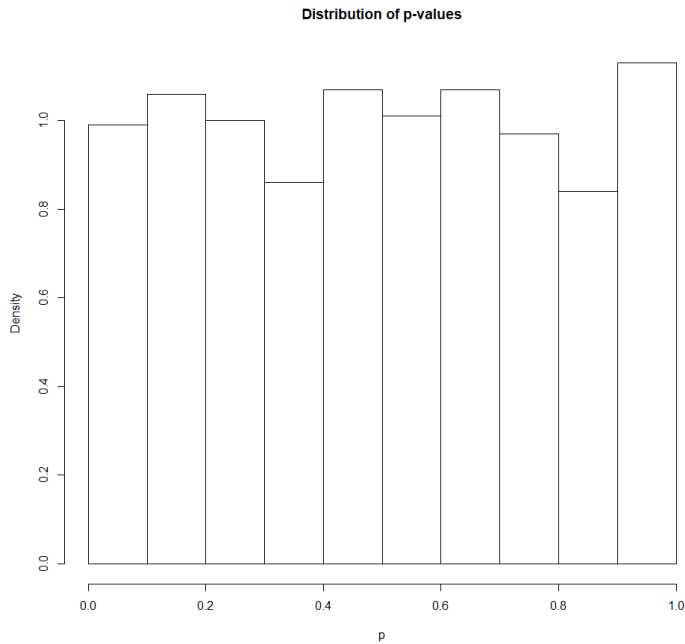


Figure 6. Histogram of the distribution of the p-values for question 1.

2. 5.1% of the p-values are smaller than 5%, and 10.9% are smaller than 10%. The p-values are normally distributed.

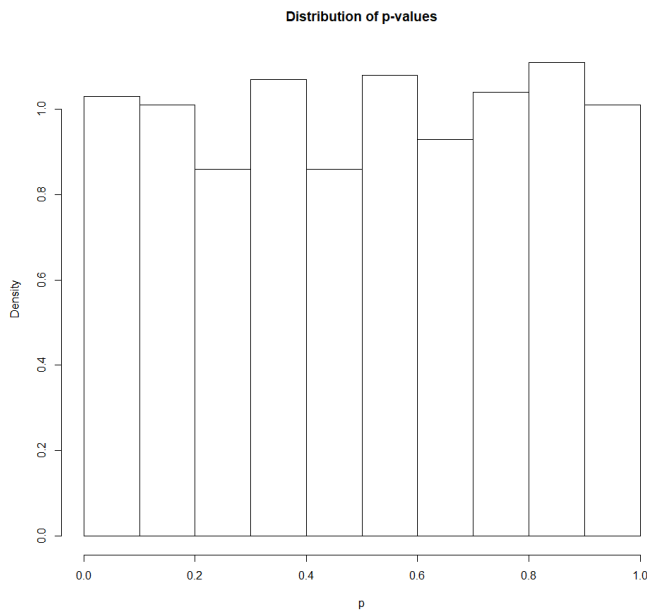


Figure 7. Histogram of the distribution of the p-values for question 2.

3. 89.2% of the p-values are smaller than 5%, and 95.5% are smaller than 10%. The p-values now follow a right-tailed distribution, with most of the p-values below 0.05.

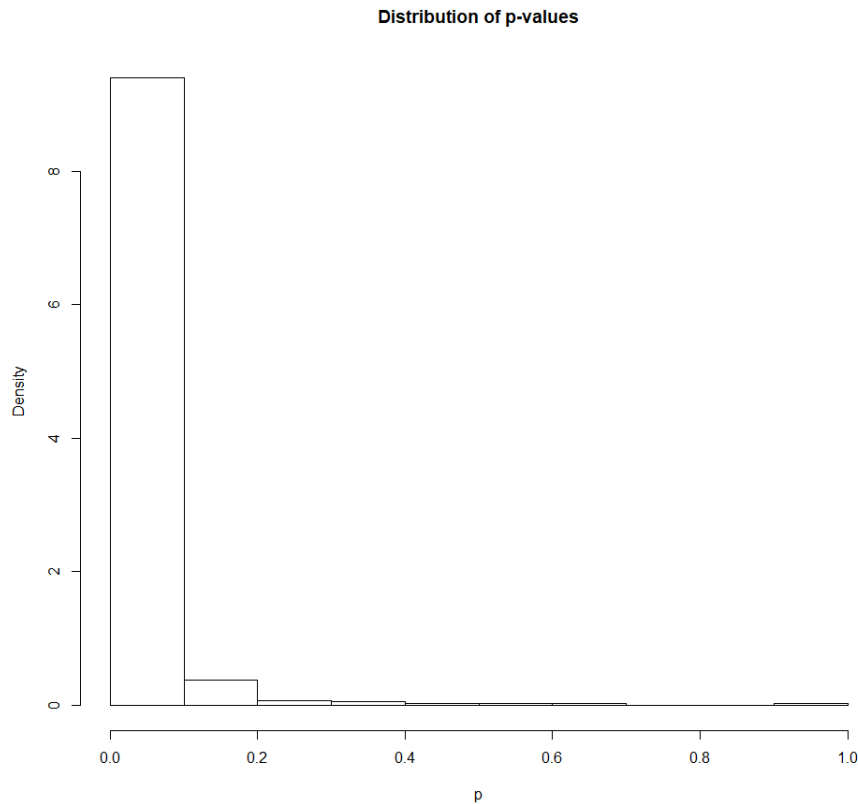


Figure 8. Histogram of the distribution of the p-values for question 3.

4. The p-values are uniformly distributed for equal means of the populations, with around 5% being smaller than 5% and around 10% being smaller than 10%. This seems to be independent of the standard deviation. However, with significantly different means, the test frequently rejects the null hypothesis. A vast proportion (around 90%) of the p-values are now rejected for $\alpha < 0.05$. This reflects the power of the test statistic, since the test accurately detects a difference # in means in 90% of the tests.

Exercise 3

1,2,3.

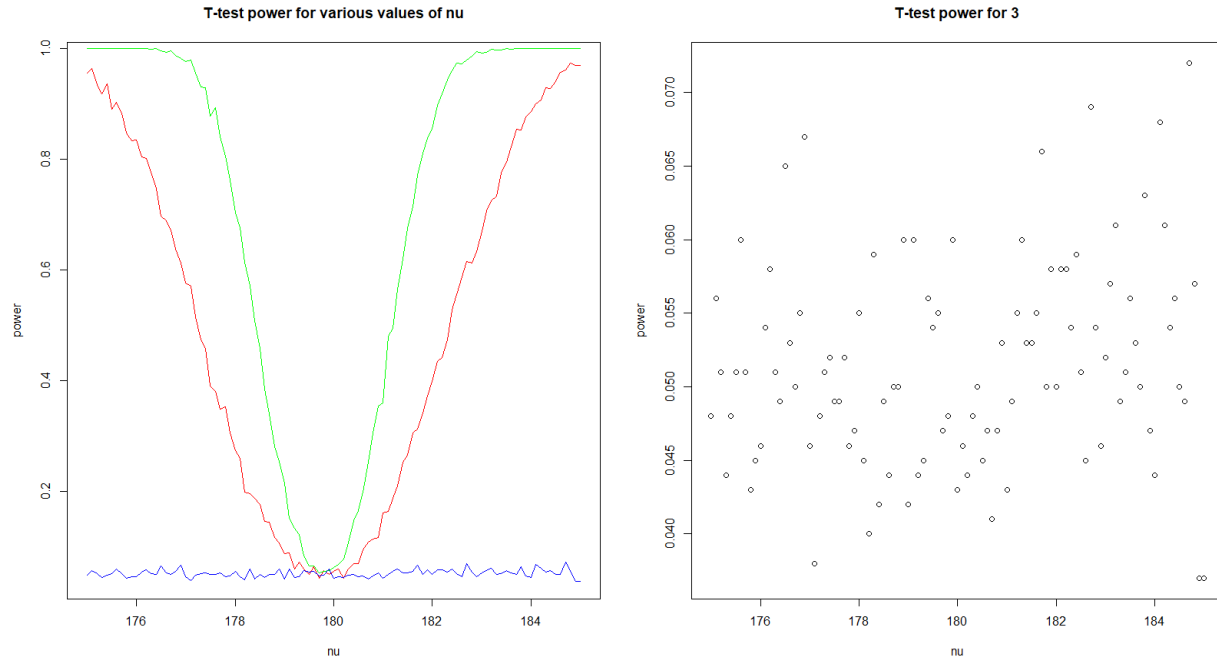


Figure 9. Composite graph of the t -test power for the three parameter sets, (left), and scatterplot of the results for #3 (right).

4. In the left graph, the red line shows the test power for the samples of length 30 (question 1). As ν approaches μ (180) and the likelihood of the null hypothesis increases, the probability of correctly rejecting the null hypothesis decreases. This is reflected in the power graph. For the parameter set of question 2, the green line shows that the increased sample sizes have increased the test's power, making the curve sharper. The blue line, representing the parameter set of question 3, shows the results for the dataset with $\sigma = 100$. The increased standard deviation means that a difference between means becomes more difficult to detect with a test. Therefore, the probability of correctly rejecting the null decreases, evening out at around 5% (uniform). With these parameters, for all values of ν within the given range, the probability of a deviation between μ and σ is not statistically significant.