Exercises for the course

"An introduction to R"

Sheet 07

Exercise 30: *The one sample Wilcoxon rank test is used to test the mean of a sample if the sample length is short and normality of the data can not be assumed but if the data is more or less symmetric around some point. If the data are not symmetric, then the null hypothesis needs to be changed into "true distribution is symmetric about 'mu' ".*
Recall that 0.5 is the true mean of the uniform distribution (unit interval). Enter `set.seed(1000)`. Sample 100 values from the uniform distribution (unit interval). Perform a one sample Wilcoxon test to test the null hypothesis that the distribution is symmetric about 0.5. Repeat the sampling and the testing for several times. What are the p-values and what are the test results?
Now we consider a short sample size. Set the seed to 2000. Sample 20 values from the uniform distribution on the interval $[0, 2.2]$. Test the null hypothesis that the sample is symmetric about 1. Note that the true mean is 10 per cent higher than 1. Repeat the test with new samples of the uniform distribution on $[0, 2.2]$. Does the Wilcoxon test detect the 10 per cent difference for most of the samples?
Recall `rent` from Exercise 22. Someone claims that the average net rent per square meter is 9.40 in Munich. Can you test this statement with the Wilcoxon one sample test? What is the p-value of the Wilcoxon one sample test. Formulate (carefully) a reply to the above statement. (3 points)

Exercise 31: *The null hypothesis of the two sample Wilcoxon test (or Mann-Whitney test) is that the distributions of the two samples differ by a location shift.* The true mean of a binomial distribution with 10 trials and success probability 1/2 is 5. Set the seed to 100. Sample 100 values from this binomial distribution. Then sample 100 values from the binomial distribution with 12 trials and success probability 1/2. Note that the mean hereof is 20 per cent higher than 5. What is the result of the two sample Wilcoxon test? Repeat the sampling and testing to convince yourself that the two sample Wilcoxon test reliably detects this 20 per cent difference. Next set the seed to 111 and repeat the above sampling and testing for 20 samples instead of 100. Does the Wilcoxon test reliably detect the 20 per cent difference in the mean?
Next sample 100 values from a standard normal distribution and 100 values from a normal distribution with mean 0 and standard deviation 100. These distributions are truly different. What is the result of the Wilcoxon test? Repeat sampling and testing several times.
Recall `heartbeats` from Exercise 12. Use the Wilcoxon test to test the null hypothesis that the treatment group (heartbeat group) and the control group are samples from the same distribution. Calculate the sample means of the treatment group and of the control group. Can you infer from the Wilcoxon test that these two means are significantly different? (3 points)

Exercise 32: Set the seed to 2222. Generate 20 values from a normal distribution with mean 1 and variance 1. Use the t-test for the null hypothesis that the true mean is equal to 1. Then sample 20 values from the normal distribution with mean 1.3 and variance 1. Again test the null hypothesis 'mean= 1' with the t-test. Repeat sampling and testing several times.
Recall `rent` from Exercise 22. Someone claims that the average net rent per square meter is 8.40 in Munich. Can you test this statement with the t-test? (Argue that the sample is long enough or test for normality of the data). Calculate the mean of `nmqm`. What is the p-value of the t-test.
(2 points)

Exercise 33: *The two sample t-test is used to test whether two samples have the same mean.* Set the seed to 2222. Generate 200 values from a standard normal distribution. Then generate 200 values from a normal distribution with mean 0 and standard deviation 100. Use the t-test for the null hypothesis that the true difference in means is equal to 0. Note that the t-test only tests for the mean and not for equality of the underlying distributions.
Apply the t-test to the two vectors `1:5` and `6:10`. Then increase the mean of the second vector by replacing it with `c(6:10,100)`. Apply the t-test to `1:5` and `c(6:10,100)`. Comparing with the previous test, the last test is surprising. Why should the t-test not be applied to these vectors? (2 points)

Exercise 34: Recall `rent` from Exercise 22. Someone claims that the average net rent per square meter is 8.40 in Munich. Can you test this statement with the t-test? Calculate the mean of `nmqm`. What is the p-value of the t-test. Formulate (carefully) a reply to the above statement. (1 points)

Exercise 35: The built-in data set `sleep` shows the effect of two soporific drugs (increase in hours of sleep compared to the control group) on 10 patients. We wish to know: Did the patients of group 1 sleep significantly longer than the control group? To answer this question, Justify why we may apply the t-test. Then apply the one sample t-test with `mu=0` to the subvector of `extra` which corresponds to group 1. Answer the analogous question for group 2. Are the increases in sleeping hours of the two groups significantly different? For the last test, note that the effect of the two drugs has been measured on each of the 10 patients. So this is a paired test (`paired=TRUE`). What would be the p-value if you apply an unpaired t-test? (3 points)

Exercise 36: Recall `heartbeats` from Exercise 12. Is the increase in weight of the heartbeat group compared to the control group significant? Use the t-test to answer this question. Then use the t-test to answer this question for every weight class separately. Justify each usage of the t-test.

Next we ask is the variance the same for both groups? Answer this question with the F-test implemented as `var.test()`. Then use this F-test to answer this question for every weight class separately. Justify each usage of the F-test by testing for normality with the Shapiro-Wilk test. (5 points)