Programming with R

Exercises

Antonio Gasparrini

01 November 2023

1. Using a loop, transform the three categorical variables in the dataset ESOPH in the package DATASET from ordered to unordered factors (see `help(factor)`).

2. Repeat the previous exercise avoiding a loop. For instance, use the functions in the `apply` family.

3. Load the dataset BIRTHS in the package EPI (remember to install/load the package first). Then, using different types of loops, identify the first record with a low birth weight. Critically compare the solutions using different loops.

4. Repeat the previous exercise avoiding a loop. For instance, consider the function `which()` (and similar).

5. Create a function `myttest()` to perform a one-sample t-test. Remember that the test is based on the statistic $t = (\bar{x} - \mu_0)/(s/\sqrt{n})$, with $\bar{x}$ and s as the sample mean and standard deviation of n observations, and μ as the assumed population mean. The distribution function `pt` (see `help(pt)`) can be used to derive the *p*-value of the two-sided test to be returned.

6. Use the function `myttest()` to test if the average birth weight of the 500 babies in the dataset BIRTHS from the package EPI is different from 3 kilograms. Compare the result with that returned by the standard function `t.test()` in the package STATS (see `help(t.test)`).

7. Modify the function created above, so that it can handle missing data through an argument `na.rm`, and so that it returns a named vector including the t statistic, the degrees of freedom and the *p*-value. Save the new function in the object `myttest2()`.

8 Apply the function `myttest2()` to test if the average gestational period is equal to 38.5 weeks. Compare the results with `t.test()`.

9. Using a loop, create a simulation study to check if the test obtained using the function `myttest()` has a nominal rejection rate. At each iteration, sample 10 observations from a standard normal distribution and apply the function `myttest()` to test the null hypothesis with a significance level $\alpha = 0.05$. Use 10,000 replicates and check the result.

10. Repeat the previous exercise avoiding a loop. For instance, consider the function `sapply()`.