Statistical Computing and Simulation

Spring 2025

Assignment 3, Due April 25/2025

1. Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) both can be used to reduce the data dimensionality. Use the mortality data, 17 5-age groups for ages 0~4, 5~9, …, 80~84, in Taiwan area to demonstrate how these two methods work. The data of the years 2001-2015 are used as the "training" (in-sample) data and the years 2016-2020 are used as the "testing" (out-sample) data. You only need to perform one set of data, according to your gender.

2. (a) Write a small program to perform the "Permutation test" and test your result on the correlation of DDT vs. eggshell thickness in class, and the following data:

X	585	1002	472	493	408	690	291
Y	0.1	0.2	0.5	1.0	1.5	2.0	3.0

Check your answer with other correlation tests, such as regular Pearson and Spearman correlation coefficients.

(b) Simulate a set of two correlated normal distribution variables, with zero mean and variance 1. Let the correlation coefficient be 0.2 and 0.8. (Use Cholesky!) Then convert the data back to Uniform(0,1) and record only the first decimal number. (亦即只取小數第一位，0至9的整數) Suppose the sample size is 10. Apply the permutation test, Pearson and Spearman correlation coefficients, and records the p-values of these three methods. (10,000 simulation runs)

3. Using simulation to construct critical values of the Mann-Whitney-Wilcoxon test in the case that $2 \le n_1, n_2 \le 10$, where n_1 and n_2 are the number of observations in two populations. (Note: The number of replications shall be at least 10,000.)

4. This assignment is to test parametric vs. nonparametric bootstrap, i.e., sensitivity of distribution assumption. Suppose 25 observations are drawn from N(0,1) and t(5). The goal is to give a 95% confidence interval for mean via both parametric and nonparametric bootstrap simulations. Assuming that observations are all from normal distribution for the parametric bootstrap. Conduct the at least 500 bootstrap

simulations each case (parametric vs. nonparametric, normal vs. t) for 1,000 times and comment on the results.

5. Similar to what we saw in class, use simulation to evaluate the type I error and testing powers of one-way ANOVA. Suppose there are 3 treatments, each with 25 observations. However, assume that the treatment with larger mean has variance 2, instead of 1. (Note: This assignment is to check the influence of the constant variance assumption. Also, as an extended study, you could also check the normality and independence assumptions.)

6. The block bootstrap can be used in prediction for dependent data. Use the built-in data "sunspot.year" in R, which is can be modeled as an AR(2) model, compare the difference of prediction via block bootstrap and AR(2) model. As a check, you can leave the final 10 observations as "testing" data.