Statistical Computing and Simulation

Spring 2017

Assignment 4, Due May 5/2017

1. Given the following data, use one of the orthogonalization methods introduced in class to perform regression analysis, including the parameter estimates and their standard errors. (You may use the functions of matrix computation built in S-Plus and R, but not the function “*lm*” or “*glm*”.) Compare your results with those from statistical software, such as SAS, SPSS, and Mintab.

-----------------------------------------------------------------------------------------------

Obs. #    y Obs. #    y

-----------------------------------------------------------------------------------------------

 1 80 27 89 42 11 58 18 89 14

 2 80 27 88 37 12 58 17 88 13

 3 75 25 90 37 13 58 18 82 11

 4 62 24 87 28 14 58 19 93 12

 5 62 22 87 18 15 50 18 89 8

 6 62 23 87 18 16 50 18 86 7

 7 62 24 93 19 17 50 19 72 8

 8 62 24 93 20 18 50 19 79 8

 9 58 23 87 15 19 50 20 80 9

10 58 18 80 14 20 56 20 82 15

 21 70 20 91 15

= air flow;

= cooling water inlet temperature;

= acid concentration;

y = stack loss.

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

1. 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.
2. 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 200 bootstrap simulations each case (parametric vs. nonparametric, normal vs. t) for 1,000 times and comment on the results.
3. To compare teaching, twenty schoolchildren were divided into two groups: ten taught by conventional methods and ten taught by an entirely new approach. The following are the test results:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Conventional  | 65 | 79 | 90 | 75 | 61 | 85 | 98 | 80 | 97 | 75 |
| New | 90 | 98 | 73 | 79 | 84 | 81 | 98 | 90 | 83 | 88 |

Are the two teaching methods equivalent in result? You need to use permutation test, (parametric and non-parametric) bootstrap, and parametric test, and then compare their differences in testing.