# Statistical Computing and Simulation

Spring 2026

Assignment 2, Due April 10/2026

1. We can use "arima.sim" in R to generate random numbers from ARIMA models.
   (a) We generate 100 random numbers from AR(2) with parameter values $(\phi_1, \phi_2) = (\theta, \theta)$ and apply correlation between $x_i$ vs. $x_{i+1}$ and $x_i$ vs. $x_{i+2}$ as a tool for verifying independence. You need to repeat the simulation at least 1,000 times and try different θ values, such as θ = 0, 0.05, 0.10, 0.15, and 0.20.
   (b) Using ARIMA random numbers to evaluate the type-1 and type-2 errors of various independence tests, e.g., Gap, Up-and-down, and Permutation tests.
2. (a) Using Rejection Method (or Inversion) described in the class, to generate random numbers from Cauchy distribution. (Note: You need to check the goodness-of-fit and independence.)

   (b) Compare with the methods of "Ratio of Uniform" (in-class) and "Ratio of Normal" and give your suggestion (such as choosing the "best" method) for creating random numbers from Cauchy distribution.
3. Write a program to generate random numbers from Poisson distribution. This program has the function for choosing the starting points, such as from starting from 0, mean, or median. In addition, this program can record the numbers of steps needed for generating a random number. Similar to what we saw in class, if λ = 10, compare the numbers of steps needed if starting from 0 and mean.
4. Given the following data, use the orthogonalization methods such as Cholesky or QR to perform regression analysis, including the parameter estimates and their standard errors via the sweep operator and compare them with those derived via the function "lm" or "glm" in R. (Note: You may use the Gauss elimination to solve the linear equation.)

| $x_1$<br>Reactor temperature ($^0$C) | $x_2$<br>Ratio of $H_2$ to n-heptane (mole ratio) | $x_3$<br>Contact time (sec.) | y<br>Conversion of n-heptane to acetylene(%) |
|---|---|---|---|
| 1300 | 7.5 | 0.0120 | 49.0 |
| 1300 | 9.0 | 0.0120 | 50.2 |
| 1300 | 11.0 | 0.0115 | 50.5 |
| 1300 | 13.5 | 0.0130 | 48.5 |

| | | | |
|---|---|---|---|
| 1300 | 17.0 | 0.0135 | 47.5 |
| 1300 | 23.0 | 0.0120 | 44.5 |
| 1200 | 5.3 | 0.0400 | 28.0 |
| 1200 | 7.5 | 0.0380 | 31.5 |
| 1200 | 11.0 | 0.0320 | 34.5 |
| 1200 | 13.5 | 0.0260 | 35.0 |
| 1200 | 17.0 | 0.0340 | 38.0 |
| 1200 | 23.0 | 0.0410 | 38.5 |
| 1100 | 5.3 | 0.0840 | 15.0 |
| 1100 | 7.5 | 0.0980 | 17.0 |
| 1100 | 11.0 | 0.0920 | 20.5 |
| 1100 | 17.0 | 0.0860 | 29.5 |

It is anticipated that an equation of the following form would fit the data:

$$E(Y)=\beta_0+\sum\beta_i x_i+\sum\beta_{ii}x_i^2+\sum_{i<j}\beta_{ij}x_i x_j\text{, and}\,Var(Y)=\sigma^2.$$

5. Figure a way to find the parameters of AR(1) and AR(2) models for the data “lynx” in R. Also, apply statistical software (e.g., R, SAS, SPSS, & Minitab) to get estimates for the AR(1) and AR(2) model and compare them to those from your program.
6. 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 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.