# Lab Exercise – 6: Data Sampling and Correlation

**Note:**

* Prepare a PDF document and name the file as "Lab6_RegisterNo.pdf".
* PDF file should consist Question No, Code, and Result for each Question.
* File Should be headed with your Register number, Slot number, Lab Exercise number.

* * *

**1. a.** Generate 100 random numbers for the age attribute and plot the equal-width(uniform) and equal frequency histograms.

**b.** Add two attributes ID (1, 2, 3, …, 100), Category (youth, middle_aged, Senior) for each age value. Develop user defined functions to perform sampling of the age attribute: SRSWOR, SRSWR, and stratified sampling. Use samples of size 10 and the strata "youth," "middle-aged," and "senior."

**2.** Develop user defined function to calculate chi-square correlation test for Nominal Data and to decide whether the two nominal attributes are independent or not.

Example Data: (Level of significance:0.05)

| | High School | Bachelors | Masters | Ph.d. | Total |
|---|---|---|---|---|---|
| **Female** | 90 | 84 | 76 | 66 | 316 |
| **Male** | 60 | 64 | 73 | 83 | 280 |
| **Total** | 150 | 148 | 149 | 149 | 596 |

| df | 0.5 | 0.10 | 0.05 | 0.02 | 0.01 | 0.001 |
|---|---|---|---|---|---|---|
| 1 | 0.455 | 2.706 | 3.841 | 5.412 | 6.635 | 10.827 |
| 2 | 1.386 | 4.605 | 5.991 | 7.824 | 9.210 | 13.815 |
| 3 | 2.366 | 6.251 | 7.815 | 9.837 | 11.345 | 16.268 |
| 4 | 3.357 | 7.779 | 9.488 | 11.668 | 13.277 | 18.465 |
| 5 | 4.351 | 9.236 | 11.070 | 13.388 | 15.086 | 20.517 |

Apply your function all pairs of Nominal attributes and formulate a chi-square correlation matrix and label each cell as 'D'(Dependent) or 'I'(Independent).

**3.** Develop user defined function to calculate the correlation coefficient for two numerical attributes and should print these two attributes are correlated (positive / negative) or not. Calculate the covariance.

Data:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *age* | 23 | 23 | 27 | 27 | 39 | 41 | 47 | 49 | 50 |
| *%fat* | 9.5 | 26.5 | 7.8 | 17.8 | 31.4 | 25.9 | 27.4 | 27.2 | 31.2 |
| *age* | 52 | 54 | 54 | 56 | 57 | 58 | 58 | 60 | 61 |
| *%fat* | 34.6 | 42.5 | 28.8 | 33.4 | 30.2 | 34.1 | 32.9 | 41.2 | 35.7 |

Draw the boxplots for *age* and *%fat*. Draw the scatter plot on these two variables.