# Lab Nr. 5, Probability and Statistics

## Statistical measures; Correlation and regression; Confidence intervals for means and variances

**A. Correlation and Regression** The frequency distributions of two characteristics $X$ and $Y$ are:

$$X = \begin{pmatrix} 20 & 21 & 22 & 23 & 24 & 25 & 26 & 27 \\ 2 & 1 & 3 & 6 & 5 & 9 & 2 & 2 \end{pmatrix},$$

$$Y = \begin{pmatrix} 75 & 76 & 77 & 78 & 79 & 80 & 81 & 82 \\ 3 & 2 & 2 & 5 & 8 & 8 & 1 & 1 \end{pmatrix}.$$

Find
a) the means $\overline{X}$, $\overline{Y}$ (mean);
b) the variances $\sigma_X^2$, $\sigma_Y^2$ (var);
c) the covariance $\text{cov}(X, Y)$ (cov);
d) the correlation coefficient $\rho_{XY}$ (corrcoef).

**B. Confidence intervals**

**1.** In a study of the size of various computer systems, the random variable $X$, the number of files stored (in hundreds of thousands), is considered. These data are obtained:

| | | | | | |
|---|---|---|---|---|---|
| 7 | 7 | 4 | 5 | 9 | 9 |
| 4 | 12 | 8 | 1 | 8 | 7 |
| 3 | 13 | 2 | 1 | 17 | 7 |
| 12 | 5 | 6 | 2 | 1 | 13 |
| 14 | 10 | 2 | 4 | 9 | 11 |
| 3 | 5 | 12 | 6 | 10 | 7 |

a) Assuming that past experience indicates that $\sigma = 5$, find a $100(1-\alpha)\%$ confidence interval for the average number of files stored.
b) Assuming nothing is known about $\sigma$, find a $100(1-\alpha)\%$ confidence interval for the average number of files stored.
c) Assuming the number of files stored are approximately normally distributed, find $100(1-\alpha)\%$ confidence intervals for the variance and for the standard deviation.

**2.** It is thought that the gas mileage obtained by a particular model of automobile will be higher if unleaded premium gasoline is used in the vehicle rather than regular unleaded gasoline. To gather evidence in this matter, 10 cars are randomly selected from the assembly line and tested using a specified brand of premium gasoline; 10 others are randomly selected and tested using the brand's regular gasoline. Tests are conducted under identical controlled conditions and gas mileages for both types of gas are assumed independent and (approximately) normally distributed. These data result:

| **Premium** | | **Regular** | |
|---|---|---|---|
| 22.4 | 21.7 | 17.7 | 14.8 |
| 24.5 | 23.4 | 19.6 | 19.6 |
| 21.6 | 23.3 | 12.1 | 14.8 |
| 22.4 | 21.6 | 15.4 | 12.6 |
| 24.8 | 20.0 | 14.0 | 12.2 |

a) Assuming $\sigma_1 = \sigma_2$, find a $100(1-\alpha)\%$ confidence interval for the difference of the true means.
b) Assuming $\sigma_1 \neq \sigma_2$, find a $100(1-\alpha)\%$ confidence interval for the difference of the true means.
( c) Find a $100(1-\alpha)\%$ confidence interval for the ratio of the variances.)