Department of Inter Disciplinary Studies,
Faculty of Engineering,
University of Jaffna, Sri Lanka
MC3020: Probability and Statistics

Tutorial-06 **July 2023**

1. The data regarding the production of wheat in tons (X) and the price of the kilo of flour in pesetas (Y) in the decade of the 80's in Spain were:

Wheat Production (X)	30	28	32	25	25	25	22	24	35	40
Flour Price (Y)	25	30	27	40	42	40	50	45	30	25

   (a) Fit the regression line using the method of least squares.
   (b) Compute the mean squared error.
   (c) Compute a 95% confidence interval for the slope of the regression line.
   (d) Test the hypothesis that the price of flour depends linearly on the production of wheat, using a 0.05 significance level.
   (e) Compute a 95% confidence interval for the intercept of the regression line.
   (f) Test the hypothesis that the regression line passes through the origin, using a 0.05 significance level.
   (g) If in a given year wheat production is 30 tons, what will be the price of a Kg of flour.

2. A random sample of eight drivers insured with a company and having similar auto insurance policies was selected. The following table lists their driving experiences (in years) and monthly auto insurance premiums.

Driving Experience (years)	5	2	12	9	15	6	25	16
Monthly Auto Insurance Premium	64	87	50	71	44	56	42	60

   (a) Does the insurance premium depend on the driving experience or does the driving experience depend on the insurance premium? and denote X, Y Variables
   (b) Find the least squares regression line by choosing appropriate dependent and independent variables based on your answer in part a.
   (c) Interpret the meaning of the values of α and β calculated in part b.
   (d) Calculate r and r^2 and explain what they mean.
   (e) Test the significance of the correlation coefficient (use α=0.05)
   (f) Test at the 5% significance level whether β is negative.

3. The following data represent the relationship between the number of alignment errors and the number of missing rivets for 10 different aircrafts.

   (a) Plot a scatter diagram.
   (b) Estimate the regression coefficients.
   (c) Test the hypothesis that $\beta_1 = 1$.

Number of Missing Rivets	Alignment Errors
13	7
15	7
10	5
22	12
30	15
7	2
25	13
16	9
20	11
15	8

(d) Estimate the expected number of alignment errors of a plane having 24 missing rivets.

(e) Compute a 90 percent confidence interval estimate for the quantity in (d).

4. A doctor wanted to determine whether there was a relation between a male's age and his HDL (so-called "good") cholesterol. He randomly selected 11 of his patients and determined their HDL cholesterol. He obtained the following data:

Age	38	42	46	32	55	52	61	61	26	38	66
HDL	47	54	64	46	45	50	62	58	37	44	62

Use R program to answer the following questions.

(a) Draw a scatter diagram of the data treating age as the predictor variable and HDL as the response variable. What type of relation, if any, appears to exist between age and HDL cholesterol?

(b) Compute the linear correlation coefficient between age and HDL cholesterol. Comment on the strength of the linear relationship using the correlation coefficient.

(c) Find the least-squares regression line treating age as the predictor variable and HDL as the response variable.

(d) Interpret the intercept and the slope coefficient in the regression line in the context of the problem.

(e) Use the regression line and predict HDL for a male of age 43.

(f) Compute the coefficient of determination and interpret the value.

5. In the questions below use the data in the accompanying table. The data come from a study of ice cream consumption that spanned the springs and summers of three years. The ice cream consumption is in pints per capita per week, price of the ice cream is in dollars, family income of consumers is in dollars per week, and temperature is in degrees Fahrenheit. Use R program to answer the following questions.

Consumption	0.386	0.374	0.393	0.425	0.406	0.344	0.327	0.288	0.269	0.256
Price	1.35	1.41	1.39	1.40	1.36	1.31	1.38	1.34	1.33	1.39
Income	351	356	365	360	342	351	369	356	342	356
Temperature	41	56	63	68	69	65	61	47	32	24

(a) Use a 0.05 significance level to test for a linear correlation between consumption and price.

(b) Find the equation of the regression that expresses consumption in terms of price.

(c) What percentage of the variation in consumption can be explained by the variation in price ?

(d) Test whether the linear model in predicting consumption using price as a predicting variable is significant or not. Use 5% level of significance.

(e) Find the equation of the multiple regression that expresses consumption in terms of price, income, and temperature.

(f) Identify the significant variables in the equation in 2(e).

(g) What percentage of the variation in consumption can be explained by the variation in price, income, and temperature?

(h) Test whether the model in predicting consumption using price, income and temperature as a predicting variable is significant or not. Use 5% level of significance.

6. A civil engineering student wants to investigate the relationship between the compressive strength of concrete (dependent variable) and the curing time (in days) (independent variable). The student collects data from various concrete samples and obtains the following results:

Curing Time (days)	Compressive Strength (MPa)
2	15
4	25
6	35
8	45
10	55

(a) Using simple linear regression, determine the equation of the regression line that best fits the data.

(b) Interpret the slope of the regression line in the context of the problem.

(c) Predict the compressive strength of concrete at 12 days of curing time using the regression equation.

(d) Calculate the coefficient of determination (R^2) and interpret its meaning in the context of the problem.

(e) Calculate a 95% confidence interval for the average compressive strength of concrete at 12 days of curing time.

(f) Calculate a 95% prediction interval for an individual compressive strength measurement at 12 days of curing time.