Department of Inter Disciplinary Studies,
Faculty of Engineering,
University of Jaffna, Sri Lanka
MC 3020 : Probability and Statistics

Tutorial-06 **July 2024**

1. Suppose you are working as a civil engineer and you are tasked with predicting the compressive strength of concrete based on the age of the concrete sample. You collected data from various concrete samples and measured their age (in days) and compressive strength (in megapascals). The dataset is as follows:

Age(days)	7	14	21	28	35	42	49	56
Compressive Strength (MPa)	15	20	23	27	30	33	37	40

   (a) Plot the data points on a scatter plot.

   (b) Calculate the equation of the least squares regression line for predicting compressive strength based on the age of the concrete.

   (c) Interpret the slope and intercept coefficients of the regression line in the context of the problem.

   (d) Use the regression line to predict the compressive strength of a concrete sample that is 30 days old.

   (e) Calculate a 95% prediction interval for an individual compressive strength measurement at 30 days of age.

   (f) Calculate the coefficient of determination (R-squared) for the regression line and interpret its meaning.

   (g) Perform a hypothesis test to determine whether the regression line is statistically significant at a 90% confidence level. State your hypotheses, the test statistic, and the conclusion clearly.

   (h) Discuss any assumptions or limitations of the linear regression model in this context.

2. Discuss any assumptions or limitations of the linear regression model in this context.

Operating Voltage (V)	110	125	130	141	145	160
Power Consumption (W)	75	68	70	80	90	100

   (a) Plot the data points on a scatter plot. what type of relation, if any, appears to exist between Operating Voltage and Power Consumption.

   (b) Estimate the correlation coefficient and interpret it.

   (c) Calculate the coefficient of determination (R-squared) for the regression line and interpret its meaning.

   (d) Test the significance of the correlation coefficient (use α =0.05).

   (e) Find the equation of the least squares regression line for predicting power consumption based on operating voltage.

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

   (g) Use the regression line to predict the power consumption for an appliance with an operating voltage of 135V.

   (h) Compute a 90% confidence interval for the slope of the regression line.

3. The table shows data on the number of visitors to the UK in a month, ν(1000s), and the amount of money they spent, m(£ millions), for each of 8 months

Number of visitors ν(1000s)	2450	2480	2540	2420	2350	2290	2400	2460
Amount of money spent m(£ millions)	1370	1350	1400	1330	1270	1210	1330	1350

(a) Find the correlation coefficient between m and ν.

(b) Fit a regression model of the form $m = a+b\nu$ to these data.

(c) Interpret the value of b.

(d) Compute a 95% confidence interval for the intercept of the regression line.

(e) Test the hypothesis that the regression line passes through the origin, using a 0.1 significance level.

(f) Estimate the amount of money spent when the no.of visitors to the UK in a month is 2500000.

(g) Compute a 90% confidence interval estimate for the amount of money spent in part(f).

4. Suppose you are conducting an experiment to study the relationship between the power consumption of a household appliance and various factors such as voltage (V), current (A), and operating time (hours). You collected data from multiple households and obtained the following dataset:

Household	**Voltage(V)**	**Current(A)**	**Operating Time (hours)**	**Power Consumption (W)**
1	220	5	4	800
2	240	6	5	960
3	230	4	3	690
4	220	5	6	1080
5	250	7	4	1400
6	230	6	5	1380

To answer this question using the output of the given R programs (without necessarily employing any formulas),

(a) Find the multiple linear regression line to determine the relationship between the power consumption (dependent variable) and the voltage, current, and operating time (independent variables).

(b) Interpret the coefficients of the regression equation in the context of the problem.

(c) Use the regression equation to predict the power consumption for a household with the following specifications: voltage = 235 V, current= 5.5 A, and operating time = 4 hours.

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

(e) Perform a hypothesis test to determine whether the regression equation is statistically significant at a 95% confidence level. State your hypotheses, the test statistic, and the conclusion.

(f) Perform the hypothesis test for testing the significance of the coefficient of the voltage variable and state the conclusion.

```
> summary(model)

Call:
lm(formula = PowerConsumption ~ Voltage + Current + OperatingtTime)

Residuals:
      1       2       3       4       5       6
-151.55 -205.49   75.77   53.94   53.94  173.38

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)     412.606   3658.272   0.113    0.920
Voltage          -4.113     17.775  -0.231    0.839
Current         258.944    199.670   1.297    0.324
OperatingtTime   37.254    139.850   0.266    0.815

Residual standard error: 231.1 on 2 degrees of freedom
Multiple R-squared:  0.753,     Adjusted R-squared:  0.3825
F-statistic: 2.033 on 3 and 2 DF,  p-value: 0.3466
```

Figure 1

5. Suppose you are a civil engineer tasked with predicting the flexural strength of concrete beams based on their dimensions. You collected data from various concrete beams and measured their width (in inches), depth (in inches), and length (in feet), as well as their corresponding flexural strength (in pounds per square inch, psi). The dataset is as follows:

Width (inches)	Depth (inches)	Length (feet)	Flexural Strength (psi)
8	6	2	3000
10	10	5	3500
12	14	6	4000
14	16	7	4200
16	18	10	4500
18	20	18	5000
20	22	20	5500
22	26	22	6000

Answer the following questions using the R programs output given Figure 2 (without necessarily employing any formulas):

(a) Find the equation of the multiple linear regression line for predicting flexural strength based on the dimensions of the concrete beams.

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

(c) Use the regression line to predict the flexural strength of a concrete beam with width 15 inches, depth 17 inches, and length 16 feet.

(d) Perform an individual coefficient test to determine the significance of the width variable in predicting flexural strength. State your hypotheses, the test statistic, and the conclusion clearly.

```
> summary(model)

Call:
lm(formula = Flexural_strength ~ Width + Depth + length)

Residuals:
       1        2        3        4        5        6        7        8
  37.113   -6.443   41.753  -12.629  -58.763 -134.278   65.464   67.784

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 2234.536    394.297   5.667  0.00478 **
Width          5.541     89.517   0.062  0.95361
Depth         98.711     48.191   2.048  0.10990
length        45.876     23.201   1.977  0.11916
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 91.77 on 4 degrees of freedom
Multiple R-squared:  0.9952,    Adjusted R-squared:  0.9917
F-statistic: 278.9 on 3 and 4 DF,  p-value: 4.239e-05
```

Figure 2

6. Suppose you are conducting an experiment to study the relationship between the efficiency of solar panels (dependent variable) and various factors such as temperature (°C), sunlight exposure (hours), and panel age (years) (independent variables). You collected data from multiple solar panel installations and obtained the following dataset:

Installation	**Temperature (°C)**	**Sunlight Exposure (hours)**	**Panel Age (years)**	**Efficiency (%)**
1	25	8	2	85
2	30	6	3	78
3	20	7	1	88
4	35	9	5	70
5	28	5	4	75
6	22	10	2	90

Based on the dataset, answer the following questions using R programs(without necessarily employing any formulas),

(a) Find the multiple linear regression line to determine the relationship between the efficiency of solar panels (dependent variable) and the temperature, sunlight exposure, and panel age (independent variables).

(b) Interpret the coefficients of the regression equation in the context of the problem.

(c) Use the regression equation to predict the efficiency of a solar panel installation with the following specifications: temperature = 27°C, sunlight exposure = 7 hours, and panel age = 3 years.

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

(e) Perform a hypothesis test to determine whether the regression equation is statistically significant at a 95% confidence level. State your hypotheses, the test statistic, and the conclusion.

(f) Perform the hypothesis test for testing the significance of the coefficient of the temperature variable and state the conclusion.

```
> summary(model2)

Call:
lm(formula = Efficiency ~ Temperature + SunlightExposure + PanelAge)

Residuals:
      1       2       3       4       5       6
 0.2459  1.1476 -1.7875 -1.1720  0.2905  1.2756

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)       99.3760     8.3549  11.894  0.00699 **
Temperature       -0.6861     0.4336  -1.582  0.25439
SunlightExposure   0.9561     0.4706   2.032  0.17927
PanelAge          -2.5591     1.6225  -1.577  0.25547
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.957 on 2 degrees of freedom
Multiple R-squared:  0.9755,    Adjusted R-squared:  0.9386
F-statistic:  26.5 on 3 and 2 DF,  p-value: 0.03659
```

Figure 3