DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SUBJECTCODE: 21MT2103RA
PROBABILITY STATISTICS AND QUEUING THEORY

Tutorial 8:
Correlation and Linear regression
Confidence limits and intervals, Statistical tests of significance: Null and Alternate Hypothesis

Date of the Session: //_____Time of the Session:_____ to_____

Learning outcomes:

- Student will be able to understand concept of liner regression and correlation
- Student will be able to understand concept of null and alternate hypothesis

PRE-TUTORIAL

1. Explain the concept of linear regression

Solution:

Linear Regression:

→ Linear regression attempts to model the relationship between two variables by fitting a linear equation to observed data

→ One variable is considered to an independent variable. & another is considered to be a dependent variable.

→ A linear regression line has an equation of the form

$$y = mx + c$$

The Linear regression model provides a slopped straight line representing the relationship between the variables.

$$y = a_0 + a_1 x + \varepsilon$$

y = Dependent Variable

x = Independent Variable

a_0 = intercept of the line

a_1 = Linear regression coefficient

ε = random error

2. Explain the concept of Null and alternate hypothesis.

Solution:

Null hyopthesis: It is a statement of no difference b/w sample means (or) propertions (or) (no diff) proportion & a population means (or) proportion. An other words the difference equals to '0.'

Alternate hypothesis: It is a claim about the population that is contradictory to H_0 & we conclude when we reject H_0.

A type II error appears when the null hypothesis is false but mistakely fails to be refused. It is losing to state that what is present and a miss. A type II error is also known as false negative, in an experiment checking for a condition with a final outcome of true or false

The rate level of type II error is represented by the Greek letter β (beta) and linked to the power of a test (which equals $1-\beta$).

IN-TUTORIAL:

1. Calculate the correlation coefficient between variables in the SAS built-in dataset called Fish, which contains various measurements for 159 different fish caught in a lake in Finland

Solution:

```
Proc corr data = sashelp.fish;
var Height width;
run;
```

The first table displays summary statistics for both Height and width.

The second table displays the Person correlation between two variables, including a p-value that tells us if the correlation is statistically significant.

from the output we can see:

- Pearson correlation coefficient 0.79288
- P-value < 0.0001

This tells us that there is a strong positive correlation between Height and that the correlation is statically significant since the p-value is less than $\alpha = 0.05$.

2 Determine a simple linear regression model using hours as the predictor variable and score as the response variable for a dataset that contains the total hours studied and final exam score for 15 students

Obs	hours	score
1	1	64
2	2	66
3	4	76
4	5	73
5	5	74
6	6	81
7	6	83
8	7	82
9	8	80
10	10	88
11	11	84
12	11	82
13	12	91
14	12	93
15	14	89

Solution:

Create the data

we'll create a dataset that contains the total hours studied and final exam score for 15. students.

we'll to fit a sample linear regression model using hours as the predictor variable and score as the responsible varible.

```
data exam_data;
input hours score;
datalines;
1  64
2  60
4  76
5  73
5  74
6  81
6  83
7  82
8  80
10 88
11 84
11 82
12 91
12 93
14 89;
run;
Proc print data = exam_data;
```

Probability Statistics and Queueing Theory - 21MT2103RA

Post Tutorial

1. Visualize Correlation with a Scatterplot using "sashelp.fish" Data set

Solution:

```
Proc corr data = sashelp.fish;
Var Height Width;
run;
```

From the plot we can see the strong positive correlation between Heights and width.
As Height increases, width tends to increase as well.
In the loop to left corner of the plot we can also see the total observations used, the correlation, and the p-value for the correlation coefficient.

(For Evaluators use only)

Comment of the Evaluator(if Any)	Evaluator's Observation Marks Secured:_____out of_____ Full Name of the Evaluator: Signature of the Evaluator: Date of Evaluation