

GUJARAT TECHNOLOGICAL UNIVERSITY

Master of Computer Applications (Integrated) Year II – (Semester-III) (w.e.f. July 2016) Practical List

Subject Name: Statistical Methods

Subject Code: 4430603

-
- (1) Create file (data.txt) using following data
12, 15, 17, 18, 14, 12, 6, 23, 22, 12, 17, 18, 15, 20, 25

- (2) **Generate mean , Median, Mode, coefficient of variation using separate command.**
The data below are the number of hours spent watching television per week for a sample of 34 households.

23.1 15.9 21.0 26.0 25.1 14.7 24.2 16.6 18.2 16.5 20.7 15.3 17.7 19.1 22.7
21.9 14.6 26.3 25.8 9.4 17.0 21.2 17.9 24.7 21.1 17.2 19.1 22.7 24.0 24.7
22.5 8.3 2.5 30.4

create file using data given and calculate

For each data set find the mean, median, standard deviation, quartiles and interquartile range using separate commands.

- (3) Use R to answer the following questions. Write down the answers.
- a. Generate numbers from 1-100.
 - b. If $x = 5$, $y = 3$ and $z = -4$ what is the value of $x^2 - 2xy + 3z + z^2$?
 - c. Find the mean value of the sample below:
14, 21, 23, 21, 16, 19, 22, 25, 16, 16

- (4) **A department of transportation's study on driving speed and mileage for mid-size automobiles in the following data.**

Driving Speed	30	50	40	55	30	25	60	25	50	55
Mileage	28	25	25	23	30	32	21	35	26	25

Compute and interpret the sample correlation coefficient.

The data below are the number of hours spent watching television per week for a sample of 34 households.

23.1 15.9 21.0 26.0 25.1 14.7 24.2 16.6 18.2 16.5 20.7 15.3 17.7 19.1 22.7
21.9 14.6 26.3 25.8 9.4 17.0 21.2 17.9 24.7 21.1 17.2 19.1 22.7 24.0 24.7
22.5 8.3 2.5 30.4

- (5) Obtain a box plot for the dataset.
The data below are the number of miles traveled to and from work each day by a sample of 12 company employees.

3.7 14.3 11.0 26.5 5.2 4.8 24.2 16.9 8.2 26.5 40.7 5.3

- (6) **create file using data given and** Obtain a scatter plot.
Create file using following data

DM	FOP	ERP
120	120	140
130	150	140
160	140	160
140	150	160
190	220	260

Plot histogram reading file data.

- (7) b) Consider a Poisson distribution with $\mu = 3$.

- a. Compute $f(2)$.
- b. Compute $f(1)$.
- c. Compute $P(x < 2)$.

- (8) a) If $n = 5$ and $p = 0.5$ find $p(x=3)$ using binomial distribution.

b) Consider a Poisson distribution with $\mu = 2.5$.

a. Compute $f(3)$.

b. Compute $f(2)$.

c. Compute $P(x \leq 2)$.

- (9) **Calculate one sample t –test for the following data.**

11 10 9 10 11 11 10 11 10 10 8 10 13 7 10 11 10 7 15 12

- (10) The data below are the number of miles traveled to and from work each day by a sample of 12 company employees.

3.7 14.3 11.0 26.5 5.2 4.8 24.2 16.9 8.2 26.5 40.7 5.3

Find a 95% confidence interval for the population mean for each sample. (3)

- (11) **Create file using following data**

DS	FOP
126	124
137	154
168	144
144	153

Create .csv file and read that file to calculate regression coefficients, also obtain regression line.

(12) Use R to answer the following questions. Write down the answers.

- Calculate $\sqrt{100 - 3 \times 4.5^2}$
- If $x = 3$, $y = 2$ and $z = -5$ what is the value of $x^3 - 2xy^2 + 3z + z^2$?
- Find the mean value of the sample below:
43, 52, 25, 44, 36, 37, 40, 43, 40, 32
- You recorded your car's mileage at your last eight fill-ups as
65311, 65624, 65908, 66219, 66499, 66821, 67145, 67447

Enter these numbers into the variable *petrol*. Use the function `diff()` on the data. What does it give?

- Interpret what both of these commands return: `mean(petrol)` and `mean(diff(petrol))`.

(13) A coin is given 10 independent tosses, each of which lands heads with probability 'a'. Let the random variable N denote the total number of heads obtained.

- Given *a* and any value (or vector of values!) of *x*, how we can calculate probability in R.
- Set *a* = 0.3 (not a very fair coin!) and verify that the above equation and the R command agree for *x*=2.
- Generate the entire probability function of the random variable N
- Use `sum` to verify that the sum of these probabilities is 1, and plot a bar chart to show their distribution.
- Calculate the mean and standard deviation of this distribution.

(14) Consider the results below that give data on highway traffic flow. It shows the relationship between traffic flow, *x* (in thousands of vehicles per day) and lead content, *y* (micro-grams per gram) of the bark of nearby trees.

<i>x</i>	8.3	8.3	12.1	12.1	17.0	17.0	17.0	24.3	24.3	24.3	33.6
<i>y</i>	227	312	362	521	640	539	728	945	738	759	1263

- Enter these data into two columns in R.
- Plot the data with labeling the diagram appropriately.
- Find the regression line of *y* on *x*. This has the equation $y = a + bx$. print the values *a* and *b*.
- Write down the equation of the best straight line.
- Find the residuals.