Sch Id : 1912160
Name : Subhajit Ghimire
Branch : CSE-B
Subject : Applied Probability Lab
Subject Code : CS216

Q.2. Answer all by examples and with R language specified functions / library if required, along with R code snippet.

1. Dependent and independent variable:

→ In regression, "dependent variable" represents the output or effect, or is tested to see if it is the effect.

The "independent variables" represent the inputs or causes, or are tested to see if they are the cause.

For example, in Q.1.,

X = height , is independent variable

Y = weight, is dependent variable.

Their relation is initiated by function lm(Y~X)

2. Dummy Variables:

→ A dummy variable is a variable that indicates whether an observation has a particular characteristic. A dummy variable can only assume the values 0 and 1, where 0 indicates the absence of the property, and 1 indicates the presence of the same.

A dummy variable is created using ifelse() function.

For example,

```
dataf$Disc_A <- ifelse (dataf$discipline == 'A', 1, 0)
dataf$Disc_B <- ifelse (dataf$discipline == 'B', 1, 0)
```

where, discipline is column name of the table.

3. Least Square Variables :

→ R is used to investigate the least squares linear regression model between two variables, the explanator (input) variable and the response (output) variable.

The command for simple linear regression is
lm(response variable ~ explanatory variable)
For example:

```
> year <- c (2000, 2001, 2002, 2003, 2004)
> rate <- c ( 9.34, 8.50, 7.62, 6.93, 6.60)
> cor (year, rate)
> fit <- lm (rate ~ year)
> plot (year, rate)
> abline (fit)
> summary (fit)
```

4. Mean Value by any example.

→ Mean is calculated by taking the sum of the values and dividing with the number of values in a data series.

The function mean() is used to calculate mean in R.

for example:

```
> x <- c (12, 7, 3, 4.2, 18, 2, 54, -21, 8, -5)
> print (mean(x))
```

Output will give 8.22 as mean.

mean() function has other attributes like "trim", which is used to drop some observations from both end of the sorted vector, and "na.rm", which is used to remove the missing values from the input vector.
for example: > print (mean(x, trim = 0.3))