UC San Diego Extension
Advanced Web Analytics:
Harnessing the Predictive Power
Winter 2016

Homework#4

Date Given: Feb 22, 2016 Due Date: Feb 28, 2016

==========

Linear Regression

The values of 2 variables X and Y are given below. Here X is the predictor variable and Y is the response variable.

X	25	30	35	40	45
Y	5	260	480	745	1100

1. Compute the correlation between X and Y variables. Verify your answer using Excel.

2. Compute regression equation using Excel by using a scatter plot. Compute the value of RSquare.

3. Compute regression equation using the following parameters.
   - Correlation between X and Y variables
   - Mean and the standard deviation of X and Y variables

4. Compute regression equation using Excel/Solver by minimizing the square of the difference. Compute the values of J, S, and RSquare.

5. Compute regression equation using KNIME software.

6. Prediction: Predict the value of Y variable when the value of X = 40 and 56 using the regression equation computed.

Multiple Regression

The data for the last 3 columns (New Visits, Page per Session, Average Session Duration) was captured by Google Analytics for a certain period. The revenues generated for the same period by the business were merged with the Google Analytics data. The data for the 25 time periods is displayed in the following table.

Period ID	Revenue $ (in thousands)	New Visits	Page Per Session	Average Session Duration (in seconds)
1	130.58	20	2	235.57
2	188.79	18	4	355.98
3	181.99	20	2	315.85
4	97.95	20	3	155.92
5	157.56	17	2	319.3
6	136.46	15	4	219.04
7	163.93	21	1	310.18
8	142.33	15	2	250.08
9	128.87	22	2	227.83
10	208.18	18	3	416.37
11	160.64	22	5	255.85
12	198.69	17	4	377.24
13	160.51	15	3	317.16
14	177.88	18	1	337.99
15	142.68	22	3	220.43
16	157.87	22	4	253.6
17	195.9	20	3	354.93
18	137.91	21	1	196.03
19	180.08	15	1	316.85
20	167.04	17	4	332.64
21	132.21	22	4	204.8
22	191.14	17	3	362.53
23	186.81	16	6	359.11
24	145.64	22	5	272.52
25	148.29	16	5	262.76

There are 3 predictor variables (New Visits, Page per Session, Average Session Duration) and 1 response variable (Revenues in thousand).

7. Compute regression equation using Excel/Solver by minimizing the square of the difference. Compute the value of RSquare.

8. Compute regression equation using KNIME software. Compute the value of RSquare.

9. Predict the revenues for a period where:
   New Visits = 32
   Page per Session = 7
   Average Session Duration = 457

Answer to the questions.

1. Correlation = 0.9959
2. Regression equation: y = 53.5x -1354.5, R Square = 0.9919
3. Regression equation: y = 53.5x -1354.5
4. Regression equation: y = 53.5x -1354.5, R Square = 0.9919
5. Regression equation: y = 53.5x -1354.5, R Square = 0.9919
6. X = 40, Y = 785.498; X = 56, Y = 1641.496
7.
8. Revenues =
   0.9801*NewVisits + 0.1858*PagesPerSession + 0.4008*AvgSessionDuration + 25.9573

   RSquare = 0.9083
9. $241,787