## BN5035 MEng in Internet of Things Technologies

## MIOT H6023 Research Project (Part 1: Research Methods)

###### Assignment 5 [Regression Modelling]

###### Due Date: December 11, 2018 CA Weighting: 2% of Final Grade

**Author: VIMAL JASWAL Student Number: B00122875**

#### Submission Date: 08/12/2018 Deferred Submission Date (when approved):

**[This assignment will be graded out of 100% and subsequently weighted to 1% of the Final Grade for the Module]**

(*Please read ALL instructions carefully*)

In this Assignment 5 [Regression Modelling]:

1. You are required to use this template for your electronic submission/upload to Moodle, and the submission MUST be in MSWord File Format. You may locate your responses after each question, or consolidate all responses with clear numbering at the end of this template.
2. Rename the electronic template file with your ITB Student Number & Assignment Name ONLY.

For example, **B0001254 Assignment 1 Part B**.

**Instructions on Submission of Assignments:**

This Assignment will contribute to the final examination grade as specified, and it is required that the submission:

1. Should be in neat typed format in MSWord;
2. MUST be submitted on or before the specified date, except with an approved deferral date;
3. Should reflect independent work.

**Instructor:**

Dr Philip Owende

Office: Room F011

Ext: 1128

Email: philip.owende@itb.ie

**Mark Awarded**

What do you honestly consider will be a fair score for this assignment that you are handing in?

**Self assessment**

**[Mark out of 100]**

**98**

**Question 1**

A study was performed on wear of a bearing, y, and its relationship to Lubrication Oil Viscosity (*x*1) and service load (*x*2). The following data was obtained:

|  |  |  |
| --- | --- | --- |
| *y* | *x*1 | *x*2 |
| 193 | 1.6 | 851 |
| 230 | 15.5 | 816 |
| 172 | 22.0 | 1058 |
| 91 | 43.0 | 1201 |
| 113 | 33.0 | 1357 |
| 125 | 40.0 | 1115 |

1. Fit a multiple linear regression model on the data.

**[5 Marks]**

1. Test for significance of the regression

**[5 Marks]**

1. Compute *t* statistics for each model parameter.

**[5 Marks]**

1. What conclusions can you draw from Part (c) above?

**[5 Marks]**

**[5 Marks for appropriate graphical presentation and detail]**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |  |  |
| Multiple R | 0.928326347 |  |  |  |  |  |  |  |
| R Square | 0.861789807 |  |  |  |  |  |  |  |
| Adjusted R Square | 0.769649679 |  |  |  |  |  |  |  |
| Standard Error | 25.49785768 |  |  |  |  |  |  |  |
| Observations | 6 |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |  |  |
| Regression | 2 | 12161.58 | 6080.789 | 9.353035 | 0.051382 |  |  |  |
| Residual | 3 | 1950.422 | 650.1407 |  |  |  |  |  |
| Total | 5 | 14112 |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 350.9942706 | 74.75307 | 4.695382 | 0.018269 | 113.0966 | 588.8919 | 113.0966 | 588.8919 |
| X Variable 1 | -1.271994448 | 1.16914 | -1.08797 | 0.3562 | -4.99272 | 2.448731 | -4.99272 | 2.448731 |
| X Variable 2 | -0.153904233 | 0.08953 | -1.71903 | 0.184101 | -0.43883 | 0.131019 | -0.43883 | 0.131019 |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| RESIDUAL OUTPUT |  |  |  | PROBABILITY OUTPUT | | |  |  |
|  |  |  |  |  |  |  |  |  |
| *Observation* | *Predicted Y* | *Residuals* |  | *Percentile* | *Y* |  |  |  |
| 1 | 217.9865769 | -24.9866 |  | 8.333333 | 91 |  |  |  |
| 2 | 205.6925023 | 24.3075 |  | 25 | 113 |  |  |  |
| 3 | 160.1797139 | 11.82029 |  | 41.66667 | 125 |  |  |  |
| 4 | 111.4595251 | -20.4595 |  | 58.33333 | 172 |  |  |  |
| 5 | 100.1704092 | 12.82959 |  | 75 | 193 |  |  |  |
| 6 | 128.5112725 | -3.51127 |  | 91.66667 | 230 |  |  |  |
|  |  |  |  |  |  |  |  |  |

**S 25.4979**

**R-square 86.18%**

**R-square (adjusted) 76.96%**

**Regression Equation**

**Y (wear of bearing) = b0 +b1 x1(oil viscosity) +b2 x2 (service load)**

**Y=351.0 -1.27 x1 – 0.1539 x2**

**wear = 351.0 - 1.27 viscosity - 0.1539 load**

**Y depend upon x1 and x2, two variances.**

|  |  |  |
| --- | --- | --- |
| SUMMARY OUTPUT |  |  |
|  |  |  |
| *Regression Statistics* | |  |
| Multiple R | 0.928326347 |  |
| R Square | 0.861789807 |  |
| Adjusted R Square | 0.769649679 |  |
| Standard Error | 25.49785768 |  |
| Observations | 6 |  |

**(b)**

Test for significance of regression.

First examine the normal probability plot to ensure that there is no reason to doubt the normality assumption. Based on the graph below, accept the assumption of normality and proceed to test for significance.

Regression Analysis: wear versus viscosity, load or (Y versus (x1,x2))

Analysis of Variance

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 2 | 12161.58 | 6080.789 | 9.353035 | 0.051382 |
| Residual | 3 | 1950.422 | 650.1407 |  |  |
| Total | 5 | 14112 |  |  |  |

(c) t statistics:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 350.9942706 | 74.75307 | 4.695382 | 0.018269 | 113.0966 | 588.8919 |
| X Variable 1 | -1.271994448 | 1.16914 | -1.08797 | 0.3562 | -4.99272 | 2.448731 |
| X Variable 2 | -0.153904233 | 0.08953 | -1.71903 | 0.184101 | -0.43883 | 0.131019 |

(d) **Since the t-values are relatively small for both viscosity and wear, but the t-value of the constant/intercept is high, we can conclude that that the two variables probably have a linearly dependent relationship.**

**Y=351.0 -1.27 x1 – 0.1539 x2**

**wear = 351.0 - 1.27 viscosity - 0.1539 load**

**we can conclude that viscosity and load has negative effect on wear and both variables reduces wear. It means wear directly depend on change in viscosity and load and is directly proportional to both variables and can be represented using above linear equation**

**Question 2**

Given the following data:

|  |  |  |
| --- | --- | --- |
| *y* | *x*1 | *x*2 |
| 26 | 1.0 | 1.0 |
| 24 | 1.0 | 1.0 |
| 175 | 1.5 | 4.0 |
| 160 | 1.5 | 4.0 |
| 163 | 1.5 | 4.0 |
| 55 | 0.5 | 2.0 |
| 62 | 1.5 | 2.0 |
| 100 | 0.5 | 3.0 |
| 26 | 1.0 | 1.5 |
| 30 | 0.5 | 1.5 |
| 70 | 1.0 | 2.5 |
| 71 | 0.5 | 2.5 |

1. Fit the Second-order polynomial regression model:

 **[20 Marks]**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *y* | *x*1 | *x*2 | *x1 square* | *x2 square* | *x1x2* |
| 26 | 1 | 1 | 1 | 1 | 1 |
| 24 | 1 | 1 | 1 | 1 | 1 |
| 175 | 1.5 | 4 | 2.25 | 16 | 6 |
| 160 | 1.5 | 4 | 2.25 | 16 | 6 |
| 163 | 1.5 | 4 | 2.25 | 16 | 6 |
| 55 | 0.5 | 2 | 0.25 | 4 | 1 |
| 62 | 1.5 | 2 | 2.25 | 4 | 3 |
| 100 | 0.5 | 3 | 0.25 | 9 | 1.5 |
| 26 | 1 | 1.5 | 1 | 2.25 | 1.5 |
| 30 | 0.5 | 1.5 | 0.25 | 2.25 | 0.75 |
| 70 | 1 | 2.5 | 1 | 6.25 | 2.5 |
| 71 | 0.5 | 2.5 | 0.25 | 6.25 | 1.25 |

|  |  |
| --- | --- |
| SUMMARY OUTPUT |  |
|  |  |
| *Regression Statistics* | |
| Multiple R | 0.997 |
| R Square | 0.994 |
| Adjusted R Square | 0.989 |
| Standard Error | 6.042 |
| Observations | 12.000 |

1. Test for significance of the regression.

**[10 Marks]**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 5 | 35092.600 | 7018.520 | 192.230 | 1.56E-06 |
| Residual | 6 | 219.066 | 36.511 |  |  |
| Total | 11 | 35311.667 |  |  |  |

1. Compute *t* statistics for each model parameter

**[10 Marks]**

**t value for constant/intercept, x1, x2, x1\*x2 are less than 1**

**whereas t value for x1 square and x2 square are greater than 1**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 24.410 | 26.592 | 0.918 | 0.394 | -40.658 | 89.477 | -40.658 | 89.477 |
| X Variable 1 | -38.033 | 40.452 | -0.940 | 0.383 | -137.015 | 60.950 | -137.015 | 60.950 |
| X Variable 2 | 0.720 | 11.687 | 0.062 | 0.953 | -27.877 | 29.317 | -27.877 | 29.317 |
| X Variable 3 | 34.975 | 21.556 | 1.623 | 0.156 | -17.770 | 87.721 | -17.770 | 87.721 |
| X Variable 4 | 11.066 | 3.158 | 3.504 | 0.013 | 3.338 | 18.793 | 3.338 | 18.793 |
| X Variable 5 | -9.986 | 8.742 | -1.142 | 0.297 | -31.378 | 11.405 | -31.378 | 11.405 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| RESIDUAL OUTPUT | |  |  | PROBABILITY OUTPUT | |
|  |  |  |  |  |  |
| *Observation* | *Predicted Y* | *Residuals* |  | *Percentile* | *Y* |
| 1 | 23.152 | 2.848 |  | 4.16666667 | 24 |
| 2 | 23.152 | 0.848 |  | 12.5 | 26 |
| 3 | 166.067 | 8.933 |  | 20.8333333 | 26 |
| 4 | 166.067 | -6.067 |  | 29.1666667 | 30 |
| 5 | 166.067 | -3.067 |  | 37.5 | 55 |
| 6 | 49.853 | 5.147 |  | 45.8333333 | 62 |
| 7 | 61.799 | 0.201 |  | 54.1666667 | 70 |
| 8 | 100.908 | -0.908 |  | 62.5 | 71 |
| 9 | 32.350 | -6.350 |  | 70.8333333 | 100 |
| 10 | 32.625 | -2.625 |  | 79.1666667 | 160 |
| 11 | 67.347 | 2.653 |  | 87.5 | 163 |
| 12 | 72.614 | -1.614 |  | 95.8333333 | 175 |

1. What conclusions can you draw from part (c) above?

**[10 Marks]**



**By using coefficients from the above t- stat table we can say the new equation by replacing beta values from table.**

**Y= 24.410 -38.033 x1 + 0.720 x2 +34.975 (x1\*x1) + 11.066(x2\*x2) -9.986(x1\*x2) + e**

**By looking at the normal probability graph we can say that the Y varies exponentially by change in variables x1, x2 and can be represented by above polynomial equation.**

**Question 3**

The data below provides a perspective of crime rate in 50 small cities in *Continent Placebo*.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total overall reported crime rate per 1 million residents,  (*y)* | Reported violent crime rate per 100,000 residents, (x*1*) | Annual police funding in €/resident,  (x*2*) | % of people 25 years+ with 4 yrs. of high school,  (*x3*) | % of 16-19 year-olds not in high school & not high school graduates, (*x4*) | % of 18- 24 year-olds in college,  (*x5*) | % of people 25 years+ with at least 4 years of college,  (*x6*) |
| 478 | 184 | 40 | 74 | 11 | 31 | 20 |
| 494 | 213 | 32 | 72 | 11 | 43 | 18 |
| 643 | 347 | 57 | 70 | 18 | 16 | 16 |
| 341 | 565 | 31 | 71 | 11 | 25 | 19 |
| 773 | 327 | 67 | 72 | 9 | 29 | 24 |
| 603 | 260 | 25 | 68 | 8 | 32 | 15 |
| 484 | 325 | 34 | 68 | 12 | 24 | 14 |
| 546 | 102 | 33 | 62 | 13 | 28 | 11 |
| 424 | 38 | 36 | 69 | 7 | 25 | 12 |
| 548 | 226 | 31 | 66 | 9 | 58 | 15 |
| 506 | 137 | 35 | 60 | 13 | 21 | 9 |
| 819 | 369 | 30 | 81 | 4 | 77 | 36 |
| 541 | 109 | 44 | 66 | 9 | 37 | 12 |
| 491 | 809 | 32 | 67 | 11 | 37 | 16 |
| 514 | 29 | 30 | 65 | 12 | 35 | 11 |
| 371 | 245 | 16 | 64 | 10 | 42 | 14 |
| 457 | 118 | 29 | 64 | 12 | 21 | 10 |
| 437 | 148 | 36 | 62 | 7 | 81 | 27 |
| 570 | 387 | 30 | 59 | 15 | 31 | 16 |
| 432 | 98 | 23 | 56 | 15 | 50 | 15 |
| 619 | 608 | 33 | 46 | 22 | 24 | 8 |
| 357 | 218 | 35 | 54 | 14 | 27 | 13 |
| 623 | 254 | 38 | 54 | 20 | 22 | 11 |
| 547 | 697 | 44 | 45 | 26 | 18 | 8 |
| 792 | 827 | 28 | 57 | 12 | 23 | 11 |
| 799 | 693 | 35 | 57 | 9 | 60 | 18 |
| 439 | 448 | 31 | 61 | 19 | 14 | 12 |
| 867 | 942 | 39 | 52 | 17 | 31 | 10 |
| 912 | 1017 | 27 | 44 | 21 | 24 | 9 |
| 462 | 216 | 36 | 43 | 18 | 23 | 8 |
| 859 | 673 | 38 | 48 | 19 | 22 | 10 |
| 805 | 989 | 46 | 57 | 14 | 25 | 12 |
| 652 | 630 | 29 | 47 | 19 | 25 | 9 |
| 776 | 404 | 32 | 50 | 19 | 21 | 9 |
| 919 | 692 | 39 | 48 | 16 | 32 | 11 |
| 732 | 1517 | 44 | 49 | 13 | 31 | 14 |
| 657 | 879 | 33 | 72 | 13 | 13 | 22 |
| 1419 | 631 | 43 | 59 | 14 | 21 | 13 |
| 989 | 1375 | 22 | 49 | 9 | 46 | 13 |
| 821 | 1139 | 30 | 54 | 13 | 27 | 12 |
| 1740 | 3545 | 86 | 62 | 22 | 18 | 15 |
| 815 | 706 | 30 | 47 | 17 | 39 | 11 |
| 760 | 451 | 32 | 45 | 34 | 15 | 10 |
| 936 | 433 | 43 | 48 | 26 | 23 | 12 |
| 863 | 601 | 20 | 69 | 23 | 7 | 12 |
| 783 | 1024 | 55 | 42 | 23 | 23 | 11 |
| 715 | 457 | 44 | 49 | 18 | 30 | 12 |
| 1504 | 1441 | 37 | 57 | 15 | 35 | 13 |
| 1324 | 1022 | 82 | 72 | 22 | 15 | 16 |
| 940 | 1244 | 66 | 67 | 26 | 18 | 16 |

**Reference:** Thomas GS (1990). The Rating Guide to Life in America's Small Cities..

1. Fit a multiple linear regression that can be used to inform of the significance of factors indicated by variables, *x1* through *x6* on the Total overall reported crime rate in *Continent Placebo* (*y*)

**[5 Marks]**

|  |  |
| --- | --- |
| SUMMARY OUTPUT | |
|  |  |
| *Regression Statistics* | |
| Multiple R | 0.783 |
| R Square | 0.613 |
| Adjusted R Square | 0.559 |
| Standard Error | 195.158 |
| Observations | 50.000 |

1. Test for significance of the regression.

**[5 Marks]**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 6 | 2595877 | 432646.2 | 11.35954 | 1.42E-07 |
| Residual | 43 | 1637723 | 38086.58 |  |  |
| Total | 49 | 4233600 |  |  |  |

1. Compute *t* statistics for each model parameter

**[5 Marks]**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 100.394 | 370.693 | 0.271 | 0.788 | -647.180 | 847.968 | -647.180 | 847.968 |
| X Variable 1 | 0.332 | 0.060 | 5.574 | 0.000 | 0.212 | 0.453 | 0.212 | 0.453 |
| X Variable 2 | 3.998 | 2.682 | 1.490 | 0.143 | -1.412 | 9.408 | -1.412 | 9.408 |
| X Variable 3 | 1.858 | 5.241 | 0.355 | 0.725 | -8.711 | 12.427 | -8.711 | 12.427 |
| X Variable 4 | 7.839 | 7.760 | 1.010 | 0.318 | -7.810 | 23.488 | -7.810 | 23.488 |
| X Variable 5 | 2.559 | 3.427 | 0.747 | 0.459 | -4.352 | 9.470 | -4.352 | 9.470 |
| X Variable 6 | -3.231 | 10.715 | -0.302 | 0.764 | -24.841 | 18.378 | -24.841 | 18.378 |

1. What conclusions can you draw from outputs of part (c) above?

**[10 Marks]**

**Regression equation for the multiple linear regression can be written as**

**Y= b0 + b1 x1 + b2 x2 + b3 x3 + b4 x4 + b5 x5 + b6 x6**

**Where b0 is constant, b1, b2, b3, b4, b5, b6 are variable co-efficient (b= beta here)**

**Y= 100.394 + 0.332 x1 + 3.998 x2 + 1.858 x3 +7.839 x4 +2.559 x5 -3.231 x6**

**As it is clear from equation that crime rate increases by effect of variables x1,x2,x3,x4,x5**

**Crime rate increases significantly by variables x2,x3,x4,x5**

**Crime rate reduces by effect of x6 variable/factor as its effect is negative.**

**Therefore it can be concluded that These variables have a linearly dependent relationship.**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| RESIDUAL OUTPUT | |  |  | PROBABILITY OUTPUT | |
|  |  |  |  |  |  |
| *Observation* | *Predicted Y* | *Residuals* |  | *Percentile* | *Y* |
| 1 | 559.882 | -81.882 |  | 1 | 341 |
| 2 | 570.986 | -76.986 |  | 3 | 357 |
| 3 | 704.005 | -61.005 |  | 5 | 371 |
| 4 | 632.824 | -291.824 |  | 7 | 424 |
| 5 | 677.921 | 95.079 |  | 9 | 432 |
| 6 | 509.218 | 93.782 |  | 11 | 437 |
| 7 | 580.919 | -96.919 |  | 13 | 439 |
| 8 | 519.430 | 26.570 |  | 15 | 457 |
| 9 | 465.220 | -41.220 |  | 17 | 462 |
| 10 | 592.558 | -44.558 |  | 19 | 478 |
| 11 | 523.893 | -17.893 |  | 21 | 484 |
| 12 | 605.521 | 213.479 |  | 23 | 491 |
| 13 | 561.610 | -20.610 |  | 25 | 494 |
| 14 | 750.879 | -259.879 |  | 27 | 506 |
| 15 | 498.821 | 15.179 |  | 29 | 514 |
| 16 | 505.314 | -134.314 |  | 31 | 541 |
| 17 | 489.952 | -32.952 |  | 33 | 546 |
| 18 | 583.595 | -146.595 |  | 35 | 547 |
| 19 | 603.776 | -33.776 |  | 37 | 548 |
| 20 | 526.018 | -94.018 |  | 39 | 570 |
| 21 | 727.874 | -108.874 |  | 41 | 603 |
| 22 | 549.932 | -192.932 |  | 43 | 619 |
| 23 | 614.592 | 8.408 |  | 45 | 623 |
| 24 | 815.577 | -268.577 |  | 47 | 643 |
| 25 | 710.461 | 81.539 |  | 49 | 652 |
| 26 | 742.455 | 56.545 |  | 51 | 657 |
| 27 | 632.544 | -193.544 |  | 53 | 715 |
| 28 | 846.266 | 20.734 |  | 55 | 732 |
| 29 | 825.025 | 86.975 |  | 57 | 760 |
| 30 | 570.105 | -108.105 |  | 59 | 773 |
| 31 | 738.086 | 120.914 |  | 61 | 776 |
| 32 | 853.831 | -48.831 |  | 63 | 783 |
| 33 | 696.862 | -44.862 |  | 65 | 792 |
| 34 | 629.087 | 146.913 |  | 67 | 799 |
| 35 | 747.239 | 171.761 |  | 69 | 805 |
| 36 | 1007.496 | -275.496 |  | 71 | 815 |
| 37 | 722.310 | -65.310 |  | 73 | 819 |
| 38 | 713.109 | 705.891 |  | 75 | 821 |
| 39 | 882.602 | 106.398 |  | 77 | 859 |
| 40 | 831.415 | -10.415 |  | 79 | 863 |
| 41 | 1907.605 | -167.605 |  | 81 | 867 |
| 42 | 739.800 | 75.200 |  | 83 | 912 |
| 43 | 734.416 | 25.584 |  | 85 | 919 |
| 44 | 729.285 | 206.715 |  | 87 | 936 |
| 45 | 667.719 | 195.281 |  | 89 | 940 |
| 46 | 942.241 | -159.241 |  | 91 | 989 |
| 47 | 698.317 | 16.683 |  | 93 | 1324 |
| 48 | 998.259 | 505.741 |  | 95 | 1419 |
| 49 | 1060.799 | 263.201 |  | 97 | 1504 |
| 50 | 1100.349 | -160.349 |  | 99 | 1740 |