Assignment 3

Hypothesis Testing

Statistical analysis for engineers

By

Vimal Jaswal

B00122875

Master of Engineering in Internet of Things Technologies.

Damian Cox

School of Informatics & Engineering, Department of Engineering
TU Dublin – Blanchardstown Campus

Blanchardstown, Dublin 15

Declaration

I declare that all the questions of the assignment have been solved by me using Microsoft excel. I will be available for explaining anything written in the assignment which seems unclear.

Signature ___Vimal Jaswal_____ Date _____02/09/2019_____

In Simple linear Regression

(SST) = Total Sum of Squares $= \sum(y_i - \bar{y})^2$

Regression Sum of Squares

$SSR = SST - SSE$

Coefficient of Multiple regression determination:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad 0 \le R^2 \le 1$$

value near 1 means good fitness.

multiple correlation coefficient

$$R = +\sqrt{R^2}$$

R is the measure of association between predictors (X's) and Response variable (Y)

Multiple Regression Model in Matrix Notation,

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & x_{13} & x_{14} \\ 1 & x_{21} & x_{22} & x_{23} & x_{24} \\ 1 & x_{31} & x_{32} & x_{33} & x_{34} \\ 1 & x_{41} & x_{42} & x_{43} & x_{44} \end{bmatrix} \quad \beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_4 \end{bmatrix}$$

For $[\ Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4\ ]$

Matrix X is design matrix containing observed data for independent explanatory variables.

Set of equations for Co-efficient can be written as

$$X\beta = Y$$

that fits as $Y = X\beta + \varepsilon$

β - column matrix of parameters and ε - error matrix. (residuals)

Pseudo inverse Solution (Ordinary least squares for coefficients.

$$b = (X^T X)^{-1} X^T Y$$

Hypothesis Tests for β_j for level of Significance α

Null $H_0: \beta_j = 0$

Alternate $H_a: \beta_j \neq 0$

Test Statistic is $T_0 = \dfrac{\hat{\beta}_j - 0}{S\sqrt{V_{jj}}} \sim T_{n-(k+1)}$

we reject H_0 (Null Hypothesis) at α if and only if $|t_0| \geq t_{\alpha/2, n-(k+1)}$

Another test based on p value

$H_0: \beta_j = 0$ for all $0 \leq j \leq k$

$H_a: \beta_j \neq 0$ for at least one $0 \leq j \leq k$

(when H_0 is true) Test Statistic $F_0 = \dfrac{MSR}{MSE} \sim f_{k, n-(k+1)}$

At level of significance α, we reject H_0 if and only if $p < \alpha$

By solving dataset in form of Matrix we observe.

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

	Coefficients	error	T-Stat	P-Value
Intercept, α	1.678	6.13	0.273	0.785
β_1	1.751	0.67	2.597	0.013
β_2	-1.07	0.605	-1.776	0.084
β_3	4.045	0.700	5.778	1.85
β_4	-2.377	0.628	-3.78	0.0006

Regression Statistics

R Square = 0.62

It should be close to 1 for good fitness. which means model includes unwanted variables

whole equation → $Y = 1.678 + 1.751X_1 - 1.07X_2 + 4.045X_3 - 2.377X_4$

Now look at p values from below table.

ANOVA

	df	SS	MS	F	Significance F (P value)
Regression	4	5727.41	1431.85	13.51	1.26E-06
Error	33	3496.56	105.95		
Total	37	9223.98			

P-value = 1.26E-06 which is greater than α - 0.05 means model is not useful as whole.

Now look at individual terms

p- value for β_1 and β_2 co-efficients is less than 0.05 level of significance and must be included in Model.

H_0: β_3 and β_4 is required to explain Y

H_a: β_3 and β_4 is not required.

As per the probability value it is clear that β_4 is required to explain Y as probability value is $P < \alpha$ or $P < 0.05$

From the Observed Statistics we can conclude that all four variables are Not required to explain 'Y'.

H_0: All variables required for Y

H_a: All variables not required.

Critical values from f- distribution for 0.05 α is 2.641 for 4 and 33 degree of freedom

The F-value obtained is 13.51 which exceeds this value

So, Null Hypotheses rejected.

All variables are not required to explain Y.

So, Significant variables equation is →

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2$$

$$Y = 1.67 + 1.75 X_1 - 1.07 X_2$$

(Reduced Model)

Regression equation

X matrix

Y		X1	X2	X3	X4
7.1	1	4.58	2.48	8.13	8.67
12.12	1	9.87	5.18	4.04	3.02
9.95	1	6.42	8.59	3.87	2.55
2.86	1	5.17	2.12	2.49	9.03
16.74	1	7.71	8.74	4.55	7.51
31.29	1	6.07	5.82	6.72	7.48
15.09	1	7.65	4.17	4.63	7.34
-9.75	1	5.74	6.65	4.84	8.72
49.12	1	5.15	0.3	8.07	7.15
19.54	1	9.72	9.9	1.81	2.75
32.4	1	9.9	8.85	8.41	2.65
-11.33	1	1.76	3.56	2.66	3.77
10.25	1	5.6	8.38	4.4	5.35
4.92	1	2.57	6.5	4.1	4.21
8.78	1	1.08	0.01	7.43	9.79
24.31	1	4.97	4.05	2.65	0.06
24.98	1	9.95	6.09	6.46	1.4
1.74	1	3.17	2	0.48	4.61
-5.88	1	4.33	6.25	0.59	6.75
3.65	1	3.66	1.88	5.42	7.19
4.26	1	3.99	0.12	3.36	6.21
3.8	1	8.28	3.42	3.22	2.62
33.98	1	6.16	5.41	5.26	1.78
-6.74	1	0.83	3.65	6.54	8.87
2.64	1	3.46	4.36	5	3.54
-5.99	1	1.7	8.91	2.87	6.02
-7.12	1	5.01	2.92	4.21	8.83
19.81	1	9.4	4.01	8.13	9.72
43.04	1	6.31	3.52	6.85	2.35
10.95	1	1.24	7.73	9.32	9.68
19.78	1	5.68	1.16	3.02	7.83
33.21	1	8.52	2.11	6.63	1.79
25.95	1	0.98	0.99	5.33	3.45
36.83	1	8.55	6.83	8.96	8.49
-10.89	1	5.79	0.77	1.13	9.49
31.67	1	4.65	0.3	9.05	6.25
7.19	1	6.78	0.4	0.13	1.54
13.13	1	9.26	7.81	8.21	9.04

XT*Y

503.38
3472.3158
2197.616
3325.4285
2363.5835

b

intercept	1.678839684
	1.751065665
	-1.075237798
	4.045148412
	-2.377164146

MMULT

XT*X

38	211.66	165.94	188.97	217.5
211.66	1459.801	1022.0045	1078.9326	1137.8177
165.94	1022.0045	1049.5094	846.6158	911.4987
188.97	1078.9326	846.6158	1185.9157	1164.4107
217.5	1137.8177	911.4987	1164.4107	1566.612

M inverse

INVERESE(XT*X)

0.354875313	-0.021359952	-0.00884909	-0.010026674	-0.02115428
-0.021359952	0.004287524	-0.001121498	-0.000708304	0.001030484
-0.00884909	-0.001121498	0.003458206	-0.000254805	0.000220402
-0.010026674	-0.000708304	-0.000254805	0.004625139	-0.001382975
-0.02115428	0.001030484	0.000220402	-0.001382975	0.003726518

1	-1.38778E-16	7.63278E-17	0	2.22045E-16
-1.77636E-14	1	5.55112E-16	4.44089E-16	8.88178E-16
0	-1.11022E-15	1	-4.44089E-16	0
-1.06581E-14	-4.44089E-16	1.11022E-16	1	8.88178E-16
0	-2.22045E-16	-5.55112E-17	0	1

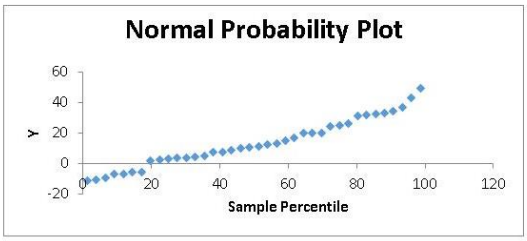
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.787989008
R Square	0.620926676
Adjusted R Sqı	0.574978395
Standard Erroı	10.29352022
Observations	38

ANOVA

	df	SS	MS	F	Significance F
Regression	4	5727.41799	1431.854497	13.51359951	1.26E-06
Residual	33	3496.566431	105.9565585		
Total	37	9223.984421			

	Coefficients	Standard Error	t Stat	P-value
Intercept	1.678839684	6.131995346	0.273783587	0.785957354
X1	1.751065665	0.674011352	2.597976518	0.013905846
X2	-1.075237798	0.605325993	-1.776295434	0.084905749
X3	4.045148412	0.700045612	5.778406922	1.85E-06
X4	-2.377164146	0.628370098	-3.783063764	0.000620043



RESIDUAL OUTPUT

Observation	Predicted Y	Residuals	andard Residuals
1	19.30917	-12.2091741	-1.255932382
2	22.55549	-10.4354899	-1.073477165
3	13.27734	-3.32734435	-0.342277001
4	-2.941028	5.801027654	0.596739665
5	6.3349	10.40509986	1.070351009
6	15.45213	15.8378662	1.629208397
7	11.8714	3.218597283	0.331090417
8	3.429272	-13.1792722	-1.355724355
9	26.02188	23.09811944	2.376055567
10	8.838861	10.70113903	1.100803943
11	37.21875	-4.81874841	-0.495694639
12	2.731055	-14.0610546	-1.446431482
13	7.555139	2.694860511	0.277214703
14	5.76728	-0.84728019	-0.087157953
15	10.34225	-1.56225393	-0.160705817
16	16.60394	7.706063602	0.792706755
17	35.35737	-10.3773738	-1.067498886
18	-3.937813	5.67781323	0.584064853
19	-11.1185	5.238502648	0.538873886
20	10.89919	-7.24918714	-0.745708823
21	7.366072	-3.10607247	-0.319515223
22	19.29756	-15.4975579	-1.594201594
23	23.6945	10.28550384	1.058048415
24	4.577431	-11.3174309	-1.164200604
25	14.86007	-12.2200711	-1.257053327
26	-7.62567	1.635669685	0.168257943
27	3.3517	-10.4716997	-1.077201996
28	23.60817	-3.79817445	-0.390710316
29	31.06616	11.97384214	1.231724269
30	10.22841	0.721592808	0.074228753
31	3.98077	15.79923025	1.625233997
32	36.89338	-3.68337754	-0.378901397
33	15.68982	10.26017665	1.05544306
34	25.36898	11.46101687	1.178970999
35	-6.998693	-3.89130674	-0.400290641
36	31.25004	0.4199591	0.04320032
37	9.986006	-2.79600628	-0.287619358
38	21.21721	-8.08720512	-0.831913991

PROBABILITY OUTPUT

Percentile	Y
1.3157895	-11.33
3.9473684	-10.89
6.5789474	-9.75
9.2105263	-7.12
11.842105	-6.74
14.473684	-5.99
17.105263	-5.88
19.736842	1.74
22.368421	2.64
25	2.86
27.631579	3.65
30.263158	3.8
32.894737	4.26
35.526316	4.92
38.157895	7.1
40.789474	7.19
43.421053	8.78
46.052632	9.95
48.684211	10.25
51.315789	10.95
53.947368	12.12
56.578947	13.13
59.210526	15.09
61.842105	16.74
64.473684	19.54
67.105263	19.78
69.736842	19.81
72.368421	24.31
75	24.98
77.631579	25.95
80.263158	31.29
82.894737	31.67
85.526316	32.4
88.157895	33.21
90.789474	33.98
93.421053	36.83
96.052632	43.04
98.684211	49.12