6.2 Multiple Linear Regression - Example

Thursday, 27 October 2022 20:54

Summary	• Implementing MLR in Excel

This is the data we're given

	Y	X1	X2
record	GPA at college	Entrance exam	interview
1	9.5	9.8	9.1
2	6.3	7.5	7.1
3	8.2	7.9	7.7
4	9.1	9.5	9.6
5	8.2	9.1	7.5
6	8.32	8.5	8.4
7	9.6	7.54	9.5
8	7.6	8.4	7.8
9	6.5	5.6	7.8
10	8.64	8	8.5
11	9.5	9.8	9.9
12	8.1	8	8.9
13	7.95	7.5	6.9
14	9.99	10	8.9
15	6.87	7.6	7.9

[Equation] GPA at college
[Equation] Entrance exam
[Equation] Interview

Response variable: Y
Explanatory variables: [Equation]

- We want to explain the variation in the response variable using these two explanatory variables.

Correlation Coefficients of [Equation] with [Equation] :

	GPA at college
GPA at college	1
Entrance exam	0.74665952
interview	0.763282985

Individual influence of [Equation] (Entrance exam) on [Equation] (GPA)

Y X1
GPA at college Entrance exam

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.74665952
R Square	0.557500439
Adjusted R Square	0.523462011
Standard Error	0.78574914
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	10.11215109	10.11215109	16.37856019	0.00138404
Residual	13	8.026222245	0.617401711		
Total	14	18.13837333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2.301862586	1.493803156	1.540941038	0.147313493	-0.925302932	5.529028103
Entrance exam	0.720234578	0.177965618	4.047043389	0.00138404	0.335763235	1.104705921

- Standard Error: s_e
- [Equation] values tells us that the regression is significant ([Equation]).
- Coefficient of Entrance exam: [Equation] Estimate of [Equation] (marginal slope).
- This SLR tells us that the entrance exam score is a good explanatory variable for Y (conclude from the P-value of Entrance exam).

Individual influence of [Equation] (interview) on [Equation] (GPA)

Y X2

GPA at college interview

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.763282985
R Square	0.582600915
Adjusted R Square	0.550493293
Standard Error	0.76313828
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	10.56743289	10.56743289	18.14525272	0.00093024
Residual	13	7.570940441	0.582380034		
Total	14	18.13837333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	0.470298787	1.846585182	0.254685672	0.802950357	-3.519005962	4.459603536
interview	0.934785006	0.219447292	4.259724488	0.00093024	0.460697953	1.408872058

- [Equation] values tells us that the regression is significant ([Equation]).
- Coefficient of interview: [Equation] Estimate of [Equation] (marginal slope).
- This SLR tells us that the interview score is a good explanatory variable for Y.

Multiple Linear Regression (MLR)

- Interpret:
  - Multiple R
  - R Square
  - Adjusted R Squared
- Regression df?
- Residual df?
- Total df?

Y X1 X2

GPA at college Entrance exam interview

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.860528625
R Square	0.740509514
Adjusted R Square	0.6972611
Standard Error	0.626281041
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	13.43163802	6.715819012	17.12223502	0.000305301
Residual	12	4.70673531	0.392227942		
Total	14	18.13837333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-0.704401949	1.576544	-0.446801326	0.662975358	-4.139396243	2.730592345
Entrance exam	0.455442321	0.168539069	2.702295228	0.019227553	0.088227236	0.822657406
interview	0.62250322	0.213981085	2.909150685	0.013101725	0.156278487	1.088727953

- Multiple R: [Equation] correlation coefficient between observed value(y) and the fitted/predicted value($\hat{y}$).
- R Square: explanatory power of this model.
  [Equation] these two explanatory variables are able to explain [Equation] in the variation of GPA.
- [Equation] (two explanatory variables)
- [Equation] .
- Adjusted R Square: [Equation] more realistic view of [Equation] . The explanatory power of the model is is actually [Equation] (and not [Equation]).

	df
Regression	2
Residual	12
Total	14

[Equation]

- We're estimating [Equation] parameters.
  [Equation] Regression [Equation] .
- Residual [Equation] .
- Total [Equation] .

- [Equation] values tells us that the regression is significant ([Equation]).

[Equation] for Entrance Exam

	[Equation] null hypothesis [Equation] : alternate hypothesis [Equation] null hypothesis is rejected. [Equation] is a good estimate of [Equation] . [Equation] for interview [Equation] null hypothesis [Equation] : alternate hypothesis [Equation] null hypothesis is rejected. [Equation] is a good estimate of [Equation] . *P-value* Intercept 0.662975358 Entrance exam 0.019227553 interview 0.013101725
☐? But the [Equation] for the intercept is [Equation] , so how are we [Equation] estimating ?	• [Equation] Regression Equation: [Equation] . • Here, [Equation] and [Equation] are partial slopes. ○ *e.g.*, keeping constant [Equation] , one unit increase in [Equation] is expected to increase [Equation] by 0.455 units.