

8.2: Software Solution

Model Development and Selection

There are many different reasons for creating a multiple linear regression model and its purpose directly influences how the model is created. Listed below are several of the more common uses for a regression model:

1. Describing the behavior of your response variable
2. Predicting a response or estimating the average response
3. Estimating the parameters (β_0 , β_1 , β_2 , ...)
4. Developing an accurate model of the process

Depending on your objective for creating a regression model, your methodology may vary when it comes to variable selection, retention, and elimination.

When the object is simple description of your response variable, you are typically less concerned about eliminating non-significant variables. The best representation of the response variable, in terms of minimal residual sums of squares, is the full model, which includes all predictor variables available from the data set. It is less important that the variables are causally related or that the model is realistic.

A common reason for creating a regression model is for prediction and estimating. A researcher wants to be able to define events within the x-space of data that were collected for this model, and it is assumed that the system will continue to function as it did when the data were collected. Any measurable predictor variables that contain information on the response variable should be included. For this reason, non-significant variables may be retained in the model. However, regression equations with fewer variables are easier to use and have an economic advantage in terms of data collection. Additionally, there is a greater confidence attached to models that contain only significant variables.

If the objective is to estimate the model parameters, you will be more cautious when considering variable elimination. You want to avoid introducing a bias by removing a variable that has predictive information about the response. However, there is a statistical advantage in terms of reduced variance of the parameter estimates if variables truly unrelated to the response variable are removed.

Building a realistic model of the process you are studying is often a primary goal of much research. It is important to identify the variables that are linked to the response through some causal relationship. While you can identify which variables have a strong correlation with the response, this only serves as an indicator of which variables require further study. The principal objective is to develop a model whose functional form realistically reflects the behavior of a system.

The following figure is a strategy for building a regression model.

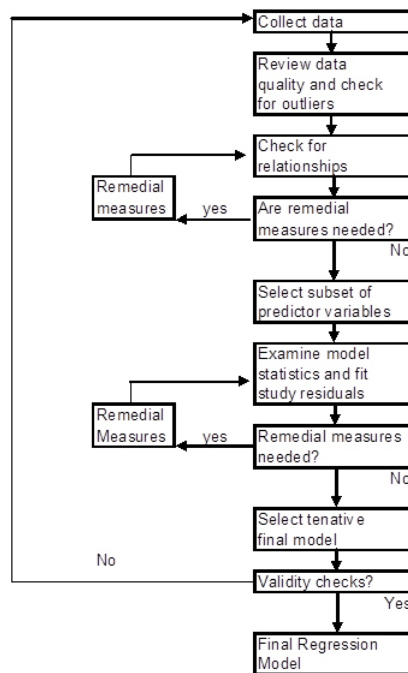


Figure 8.2.1 . Strategy for building a regression model.

Software Solutions

Minitab

The screenshot shows the Minitab software interface. The 'Stat' menu is open, and the 'Regression' option is selected. The 'General Regression...' option is highlighted. Below the menu, a worksheet titled 'Worksheet 1' is visible, containing data for a polynomial regression analysis. The data is organized into columns: C1 (CuFt), C2 (BA/ac), C3 (%BA Bspruce), C4 (SI), C5, C6, and C7. The data rows are numbered 1 through 11. The status bar at the bottom indicates the project name 'Project...' and the description 'Perform polynomial regression with continuous and categorical predictors'.

	C1	C2	C3	C4	C5	C6	C7
	CuFt	BA/ac	%BA Bspruce	SI			
1	55	51	79	45			
2	68	100	48	53			
3	60	63	67	44			
4	40	52	52	31			
5	45	67	52	29			
6	49	42	82	43			
7	62	81	80	42			
8	56	70	66	36			
9	93	108	96	63			
10	76	90	81	60			
11	94	110	78	56			

Minitab multiple regression.MPJ

File Edit Data Calc Stat Graph Editor Tools Window Help Assistant

General Regression

Response: CuFt

Model: BA/ac SI %BA Bspruce

Categorical predictors (optional):

Options... Box-Cox... Prediction...
Graphs... Results... Storage...

Select

Help OK Cancel

	C1	C2	C3	C4
	CuFt	BA/ac	%BA Bspruce	SI
1	55	51	79	45
2	68	100	48	53
3	60	63	67	44
4	40	52	52	31
5	45	67	52	29
6	49	42	82	43
7	62	81	80	42
8	56	70	65	36
9	93	108	96	63
10	76	90	81	60
11	94	110	78	56

General Regression - Graphs

Residuals for Plots:
☒ Regular ☐ Standardized ☐ Deleted

Residual Plots
☒ Individual plots
☐ Histogram of residuals
☒ Normal plot of residuals
☒ Residuals versus fits
☐ Residuals versus order
☐ Four in one

Residuals versus the variables:

Select

Help OK Cancel

Project...

Perform polynomial regression with continuous and categorical predictors

General Regression

Response: CuFt

Model: BA/ac SI %BA Bspruce

Categorical predictors (optional):

Options... Box-Cox... Prediction...
Graphs... Results... Storage...

Select
Help OK Cancel

	CuFt	BA/ac	%BA Bspruce	SI
1	55	51	79	45
2	68	100	48	53
3	60	63	67	44
4	40	52	52	31
5	45	67	52	29
6	49	42	82	43
7	62	81	80	42
8	56	70	65	36
9	93	108	96	63
10	76	90	81	60
11	94	110	78	56

General Regression - Results

☒ Regression equation
☐ Display all equations (if >50)
☒ Coefficient table
☐ Display variance inflation factors
☒ Display confidence intervals
☒ Summary of model
☒ Analysis of Variance table
☐ Fits and Diagnostics table
☒ Display only unusual observations
☐ Display all observations
☐ Durbin-Watson statistic
☐ Prediction tables

Help OK Cancel

Perform polynomial regression with continuous and categorical predictors

The output and plots are given in the previous example.

Excel

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

100% Arial

Reply with Changes... End Review...

G21

A	B	C	D	E	F	G	H	I	J
CuFt	BA/ac	%BA Bspr	SI						
55	51	79	45						
68	100	48	53						
60	63	67	44						
40	52	52	31						
45	67	52	29						
49	42	82	43						
62	81	80	42						
56	70	65	36						
93	108	96	63						
76	90	81	60						
94	110	78	56						
82	111	59	48						
86	94	84	53						
55	82	48	40						
71	65	93	35						
67	87	68	41						
73	108	51	54						
87	105	82	51						
80	100	70	45						
77	103	61	43						
64	55	96	51						
60	60	80	47						
65	70	76	40						
65	78	74	46						
83	85	96	55						
67	92	58	50						
61	82	58	38						
51	56	69	35						

Data Analysis

Analysis Tools

- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram
- Moving Average
- Random Number Generation
- Rank and Percentile
- Regression
- Sampling
- t-Test: Paired Two Sample for Means

OK Cancel Help

Microsoft Excel - Book1

File Edit View Insert Format Tools Data Window Help

100% Arial

Reply with Changes... End Review...

A	B	C	D	E	F	G	H	I	J	K
CuF1	BA/ac	%BA Bspr SI								
55	51	79	45							
68	100	48	53							
60	63	67	44							
40	52	52	31							
45	67	52	29							
49	42	82	43							
62	81	80	42							
56	70	65	36							
93	108	96	63							
76	90	81	60							
94	110	78	56							
82	111	59	48							
86	94	84	53							
55	82	48	40							
71	65	93	35							
67	87	68	41							
73	108	51	54							
87	105	82	51							
80	100	70	45							
77	103	61	43							
64	55	96	51							
60	60	80	47							
65	70	76	40							
65	78	74	46							
83	85	96	55							
67	92	58	50							
61	82	58	38							
51	56	69	35							

Regression

Input

Input Y Range: \$A\$2:\$A\$29

Input X Range: \$B\$2:\$D\$29

☐ Labels ☐ Constant is Zero

☐ Confidence Level: 95 %

Output options

☒ Output Range: \$G\$1

☐ New Worksheet Ply:

☐ New Workbook

Residuals

☐ Residuals ☒ Residual Plots

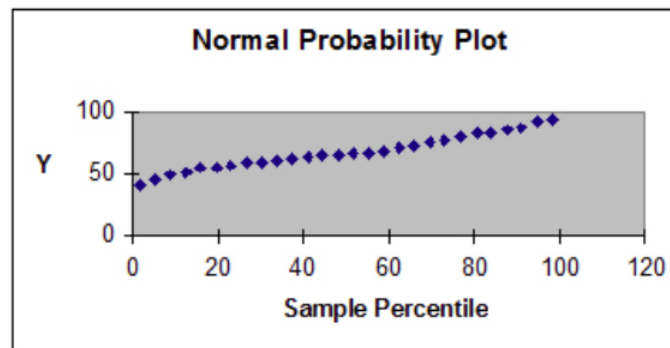
☐ Standardized Residuals ☐ Line Fit Plots

Normal Probability

☒ Normal Probability Plots

OK Cancel Help

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.977388							
R Square	0.955287							
Adjusted R Square	0.949697							
Standard Error	3.177363							
Observations	28							
ANOVA								
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>			
Regression	3	5176.562	1725.521	170.9175	2.52E-16			
Residual	24	242.2952	10.09563					
Total	27	5418.857						
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	-19.3858	4.153321	-4.66753	9.67E-05	-27.9578	-10.8137	-27.9578	-10.8137
X Variable 1	0.591004	0.042936	13.76471	6.95E-13	0.502388	0.67962	0.502388	0.67962
X Variable 2	0.489441	0.052453	9.331096	1.87E-09	0.381184	0.597696	0.381184	0.597696
X Variable 3	0.089988	0.112618	0.799059	0.432094	-0.14244	0.32242	-0.14244	0.32242



This page titled [8.2: Software Solution](#) is shared under a [CC BY-NC-SA 3.0](#) license and was authored, remixed, and/or curated by [Diane Kiernan \(OpenSUNY\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.