Gus Lipkin

Multiple Regression

Quick note: Each step is on its own page.

1. First, we run a best subsets regression to check everything.

Response is Heat Flux

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	Insulation	East	South	North	Time of Day
1	72.1	71.0	66.9	38.5	12.328				X	
1	39.4	37.1	26.3	112.7	18.154	X				
2	85.9	84.8	81.4	9.1	8.9321			X	X	
2	82.0	80.6	74.2	17.8	10.076				X	X
3	87.4	85.9	79.0	7.6	8.5978		X	X	X	
3	86.5	84.9	81.4	9.7	8.9110	X		X	X	
4	89.1	87.3	80.6	5.8	8.1698	X	X	X	X	
4	88.0	86.0	79.3	8.2	8.5550	X		X	X	X
5	89.9	87.7	78.8	6.0	8.0390	X	X	X	X	X

2. Then we run a regression with all five predictors as that had the highest R^2 value in the best subsets.

THERMALENERGYTEST.MTW

Regression Analysis: Heat Flux versus Insulation, East, South, North, Time of Day

Regression Equation

Heat Flux = 325.4 + 0.0675 Insulation + 2.55 East + 3.80 South - 22.95 North + 2.42 Time of Day

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	325.4	96.1	3.39	0.003	
Insulation	0.0675	0.0290	2.33	0.029	2.32
East	2.55	1.25	2.04	0.053	1.36
South	3.80	1.46	2.60	0.016	3.18
North	-22.95	2.70	-8.49	0.000	2.61
Time of Day	2.42	1.81	1.34	0.194	5.37

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
8.03902	89.88%	87.68%	78.82%

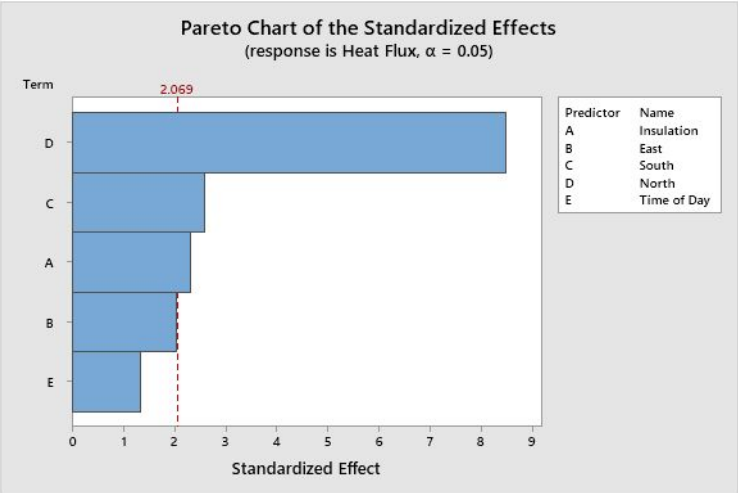
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	13195.5	2639.11	40.84	0.000
Insulation	1	350.6	350.56	5.42	0.029
East	1	270.1	270.13	4.18	0.053
South	1	437.2	437.15	6.76	0.016
North	1	4656.6	4656.56	72.05	0.000
Time of Day	1	115.5	115.50	1.79	0.194
Error	23	1486.4	64.63		
Total	28	14681.9			

Fits and Diagnostics for Unusual Observations

Obs	Heat Flux	Fit	Resid	Std Resid	
1	271.80	267.57	4.23	0.97	X
4	230.70	213.30	17.40	2.89	R
22	254.50	240.14	14.36	2.09	R

R Large residual
X Unusual X



3. Now that we have our initial model, we want to eliminate any predictors that have a high p-value. High p-values have less of an effect and by reducing the number of predictors, we have fewer things to measure which is less expensive for the company. This means eliminating East and Time of Day.

THERMALENERGYTEST.MTW

Regression Analysis: Heat Flux versus North, South, Insulation

Regression Equation

Heat Flux = 437.3 + 0.0291 Insulation + 4.693 South - 22.53 North

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	437.3	58.9	7.42	0.000	
Insulation	0.0291	0.0274	1.06	0.299	1.69
South	4.693	0.954	4.92	0.000	1.10
North	-22.53	2.50	-9.00	0.000	1.82

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
8.91102	86.48%	84.86%	81.43%

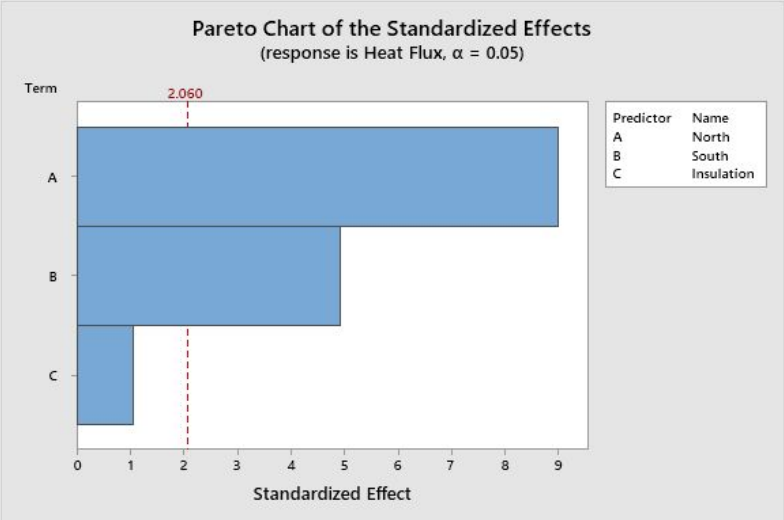
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	3	12696.8	4232.26	53.30	0.000
Insulation	1	89.2	89.17	1.12	0.299
South	1	1922.3	1922.28	24.21	0.000
North	1	6436.9	6436.92	81.06	0.000
Error	25	1985.2	79.41		
Total	28	14681.9			

Fits and Diagnostics for Unusual Observations

Obs	Heat Flux	Fit	Resid	Std Resid	
22	254.50	230.94	23.56	2.74	R

R Large residual



4. We want to repeat the process of eliminating variables with high p-values. This step we eliminate Insulation.

THERMALENERGYTEST.MTW

Regression Analysis: Heat Flux versus North, South

Regression Equation

Heat Flux = 483.7 + 4.796 South - 24.22 North

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	483.7	39.6	12.22	0.000	
South	4.796	0.951	5.04	0.000	1.09
North	-24.22	1.94	-12.48	0.000	1.09

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
8.93207	85.87%	84.78%	81.36%

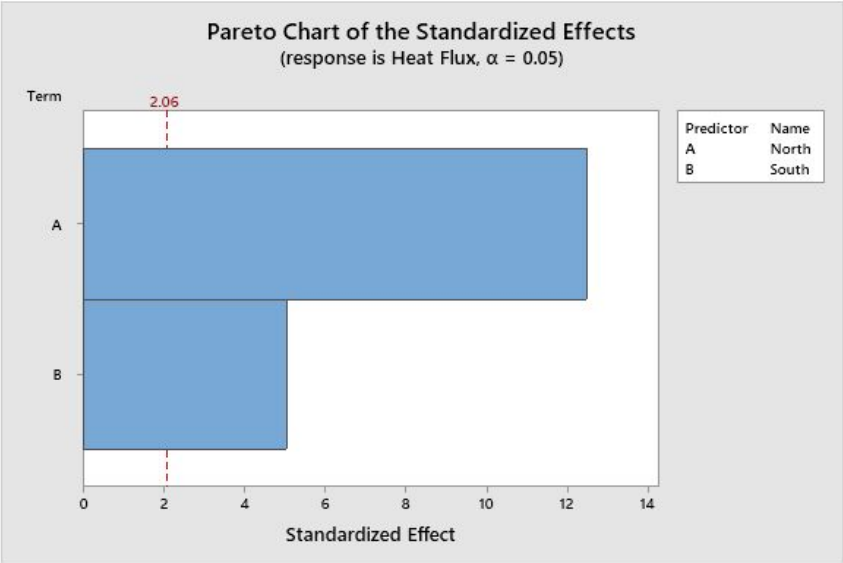
Analysis of Variance

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	12608	6303.8	79.01	0.000
South	1	2029	2028.9	25.43	0.000
North	1	12423	12423.1	155.71	0.000
Error	26	2074	79.8		
Total	28	14682			

Fits and Diagnostics for Unusual Observations

Obs	Heat Flux	Fit	Resid	Std Resid	
1	271.80	274.74	-2.94	-0.40	X
22	254.50	230.91	23.59	2.74	R

R Large residual
X Unusual X



5. Now that we've eliminated all high p-values, we have the final equation which is $Heat\ Flux = 483.7 + 4.796\ South - 24.22\ North$.