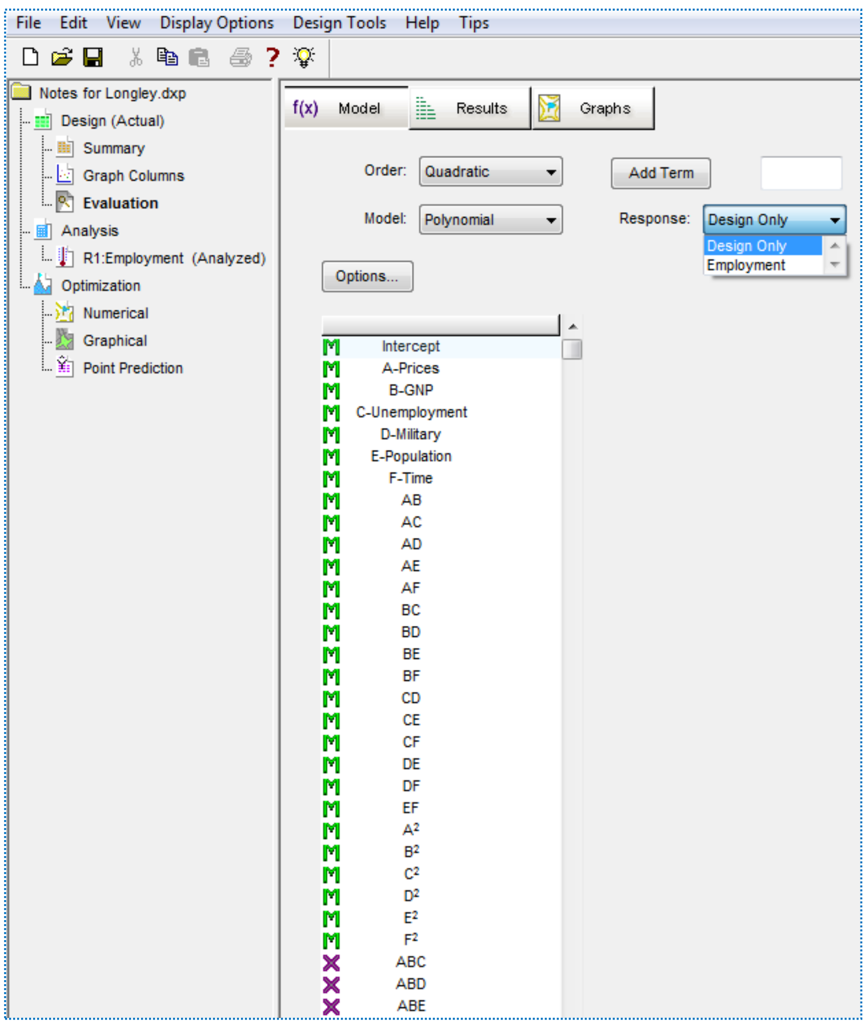
Historical Data RSM Tutorial
(Part 2 – Advanced Topics)

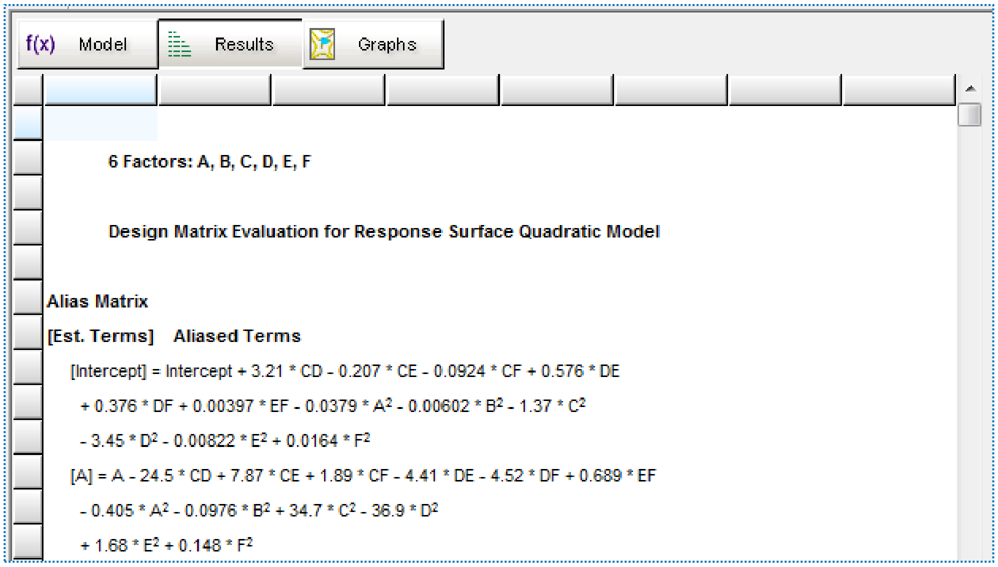
Design Evaluation

If you still have the Longley data active in Design-Expert software from Part 1 of this tutorial, continue on. If you exited the program, re-start it and use **Open Design** to open your data file (**Longley.dxp**). Under the **Design** branch of the program, click **Evaluation**. The software brings up a quadratic polynomial model by default, but, as you will see, the order must be downgraded to linear (we will get to the reason momentarily). The screen shot shows the Response field set at "Design Only" as opposed to the Employment response. In other words, it will evaluate the entire matrix of factors, regardless whether response data are present. The other option (response by response) comes in handy when experimenters end up with missing data, thus degrading the "designed-for" model.



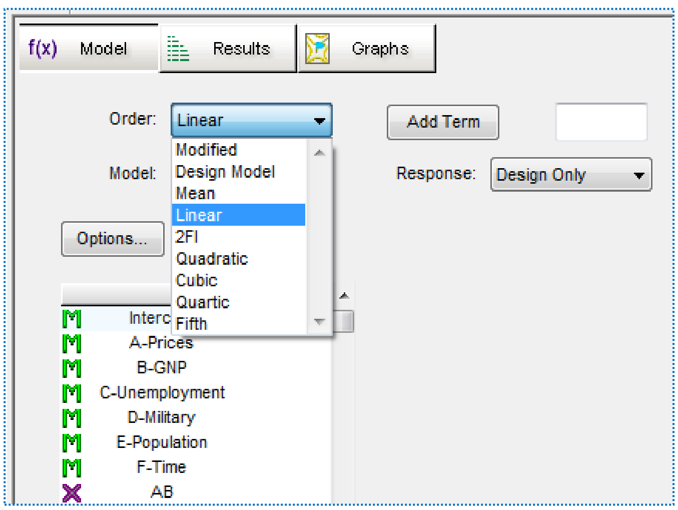
Design evaluation (design only)

Press the ***Results*** button.



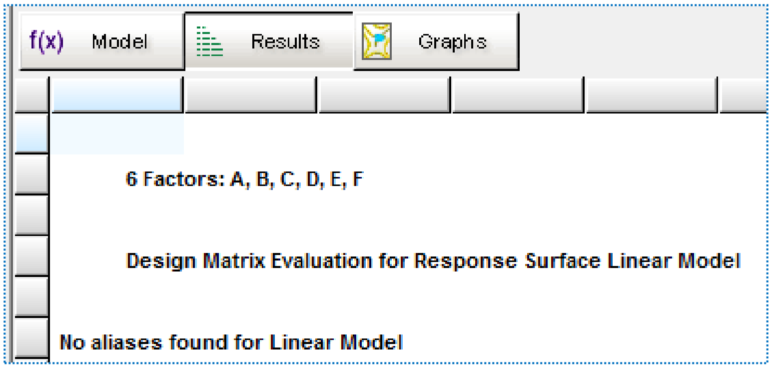
Results of evaluation for quadratic polynomial

This model is badly aliased. For example, the effect of A is confounded with -24.5 CD, etc. Go back to ***Model*** and reduce the **Order** to **Linear**.



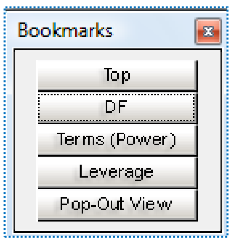
Re-setting order to linear

Press ***Results*** again and note "No aliases found…" Much better!



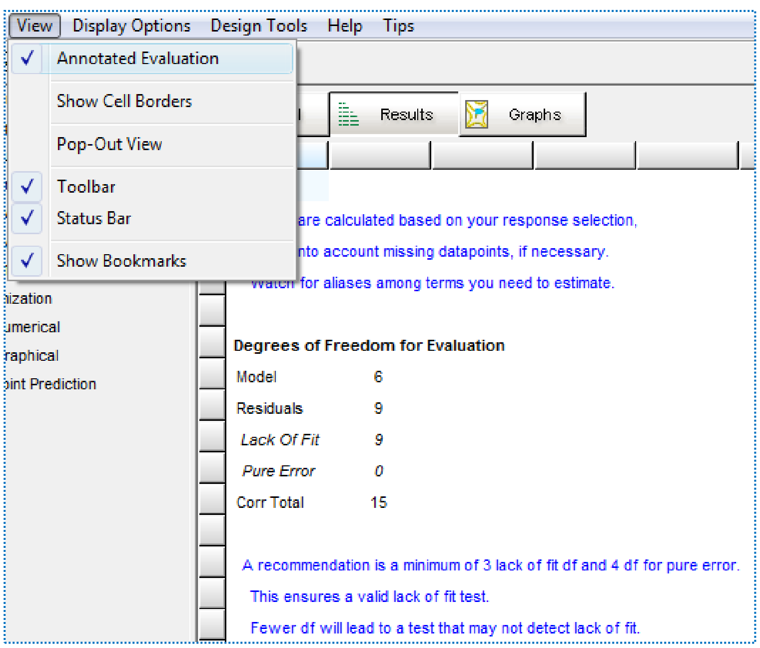
Results of evaluation for linear model

On **Bookmarks** click the **DF** option to bring up the accounting for degrees of freedom.



Bookmarking to evaluate degrees of freedom (DF)

Looking over the annotations provided by the software (activated via **View**, **Annotated Evaluation**), notice this design flunks the recommendation for pure error df. Of course this really is not a designed experiment, but rather historical data collected at happenstance.



Annotations for degrees of freedom

Study the next section of the evaluation by Design-Expert. Do any of the statistics pass the tests suggested for a good design? No!

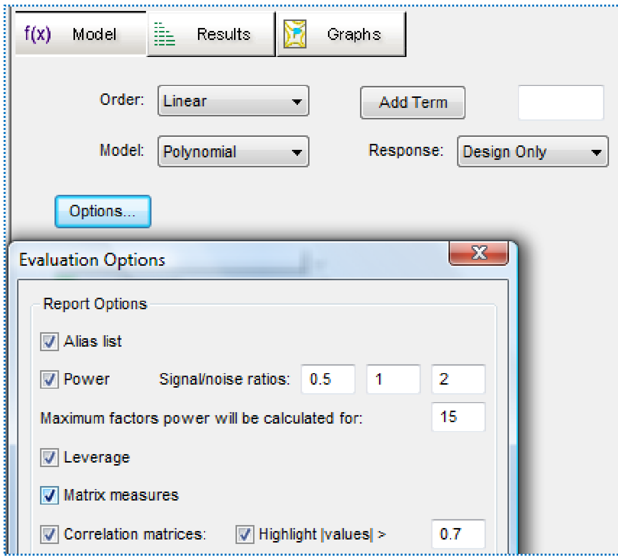
				Power at 5 % alpha level to detect signal/noise ratios of		
Term	StdErr**	VIF	Ri-Squared	0.5 Std. Dev.	1 Std. Dev.	2 Std. Dev.
A	4.72	135.53	0.9926	5.0 %	5.1 %	5.4 %
B	17.61	1788.51	0.9994	5.0 %	5.0 %	5.0 %
C	2.35	33.62	0.9703	5.1 %	5.4 %	6.7 %
D	0.75	3.59	0.7214	6.0 %	9.2 %	22.2 %
E	8.33	399.15	0.9975	5.0 %	5.0 %	5.1 %
F	11.21	758.98	0.9987	5.0 %	5.0 %	5.1 %

**Basis Std. Dev. = 1.0

Details on model terms, including power

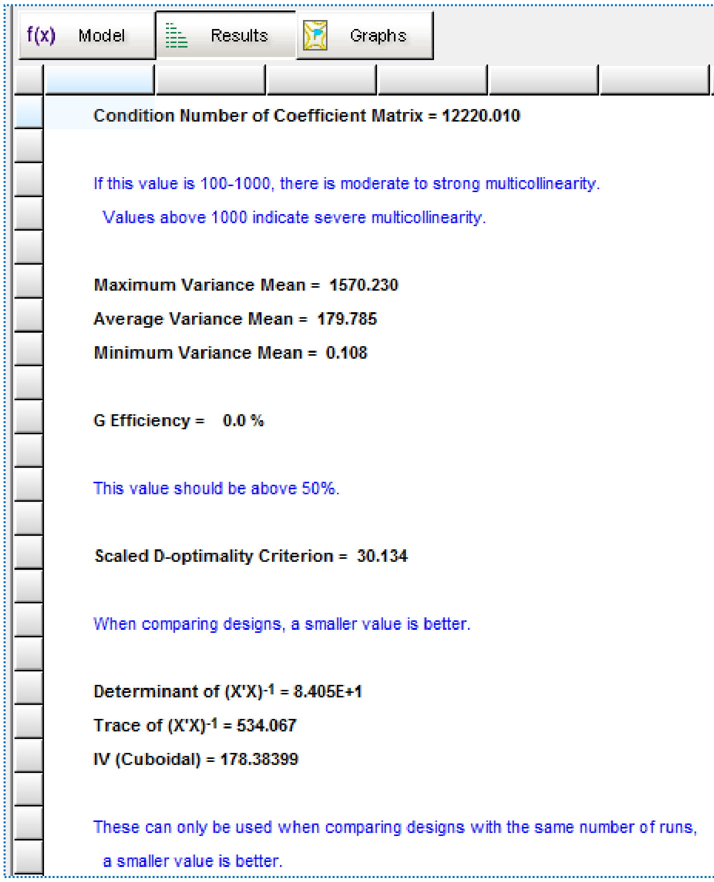
Scroll down or bookmark to the leverage report. These statistics come out surprisingly good – none exceeds twice the average.

More statistics are available by going back to ***Model***, selecting **Options,** and turning on (checkmarks) **Matrix Measure** and **Correlation Matrices**.



Turning on more options for report

Click **OK** and view the ***Results***. On **Bookmarks** choose **Matrix** to see new statistics.



Matrix measures for design evaluation

Notice the condition number (12,220) far exceeds the level considered to represent severe multicollinearity for a design matrix (1000 or fewer). Viewing specific correlations reveals why.

Correlation Matrix of Regression Coefficients

	Intercept	A	B	C	D	E	F
Intercept	1.000						
A	-0.527	1.000					
B	0.523	-0.649	1.000				
C	0.518	-0.555	0.946	1.000			
D	0.557	-0.349	0.469	0.619	1.000		
E	-0.082	0.659	-0.833	-0.758	-0.189	1.000	
F	-0.608	0.186	-0.802	-0.824	-0.549	0.388	1.000

Correlation Matrix of Factors [Pearson's r]

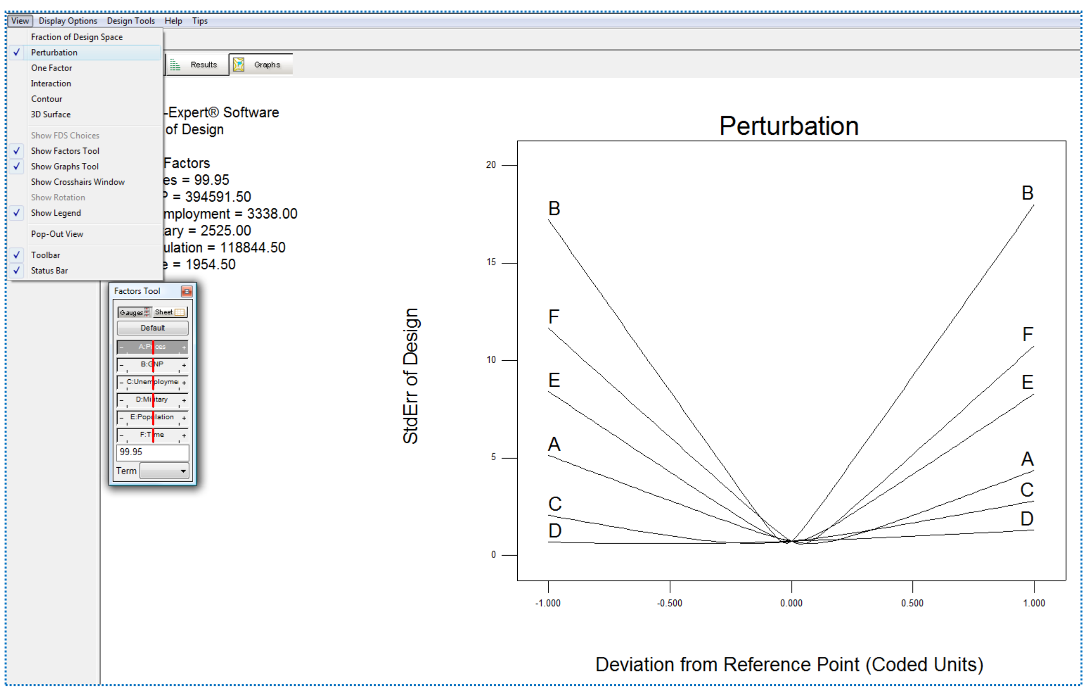
	A	B	C	D	E	F
A	1.000					
B	0.99	1.00				
C	0.62	0.60	1.00			
D	0.46	0.45	-0.18	1.00		
E	0.98	0.99	0.69	0.36	1.00	
F	0.99	1.00	0.67	0.42	0.99	1.00

Off-diagonal values close to zero are better.

Correlation matrices

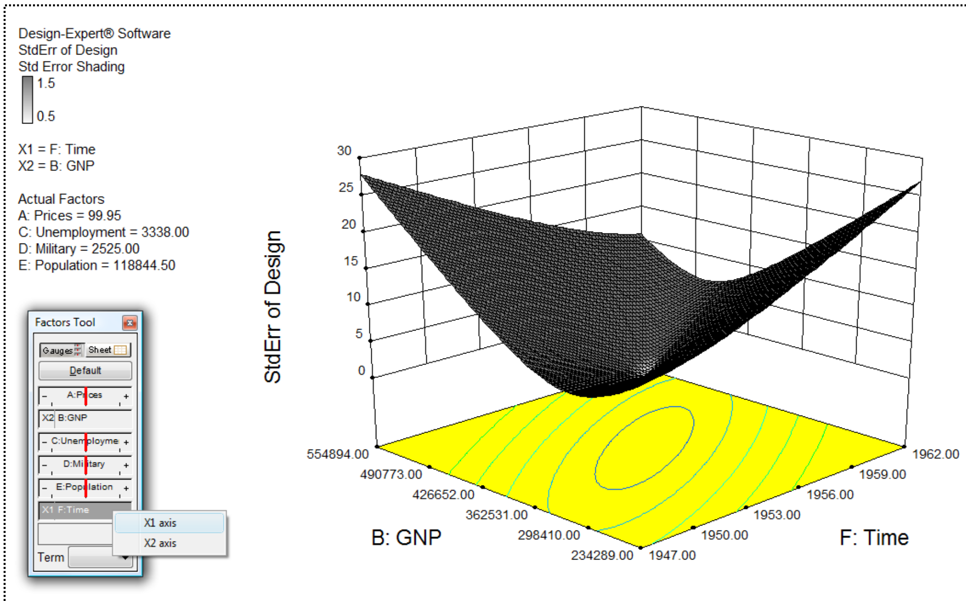
Notice many values are highlighted in blue for being unacceptably high. No wonder Longley picked this data to test regression software!

Now, just for fun, press the ***Graphs*** button and select **View**, **Perturbation** (or press this option on the floating Graphs Tool).



Perturbation plot for standard error

Notice factors B and F exhibit the most dramatic tracks for standard error. On the floating **Graphs Tool** select **3D Surface.** On the **Factors Tool,** right-click factor **F:Time** and change it to **X1 axis**.



3D view of standard error for factors B and F

There's no sense doing anything more. By now it's clear that this 'design' fails all the tests for a good experiment, but that's generally the nature of the beast for happenstance data.