



Stepwise Analysis in CO\$TAT

Steven Ikeler
ODASA-CE (Army)



Agenda

ODASA
Cost &
Economics

- Background
- Challenges in explaining methodology
- Benefits of the Stepwise Analysis tool
- Example of CO\$TAT Stepwise Analysis
- Advanced Topics
 - Handling multicollinearity
 - Missing data
 - Other advanced topics
- Developing best practices
- Questions



Background



Background and Environment

ODASA
Cost &
Economics

Assessing the needs of the organization:

- Lots of databases
- Many potential variables means multicollinear data
- May be incomplete data
- Statistical regressions are an acceptable methodology
- Repeatable results
- Decision-makers are informed about statistics



Background and Environment

ODASA
Cost &
Economics

Assessing the needs of the organization:

- Many first-uses of new CO\$TAT tools
- Review CO\$TAT results from a broad variety of organizations
- Informal best practices among teams
- Tend to be one or two CO\$TAT specialists on a team (more technical)



Challenges Explaining Methodology



Challenges Explaining Methodology (Lessons Learned)

ODASA
Cost &
Economics

- Decision Makers expect to be able to interpret the coefficients of the CER
- Must be intuitive
- Potentially many steps involved
- Non-linear progression adding variables
- Must be able to show why something didn't work
- Biases



Challenges Explaining Methodology (Guidance Challenges)

- Defense Acquisition University (DAU) recommends comparing all combinations of all variables
- CO\$TAT Rules of Thumb

The screenshot shows the CO\$TAT Help window. The title bar reads 'CO\$TAT Help'. The menu bar includes 'Hide', 'Back', 'Print', and 'Options'. The left pane has tabs for 'Contents', 'Index', 'Search', and 'Glossary'. Below the tabs is a search box with the text 'Type in the keyword to find:' and a search box containing 'General Rules'. A list of topics is shown below the search box, with 'General Rules' selected. The main pane displays the following content:

General Rules of Thumb for Regression Statistics

Statistics which are Comparable Within Model Forms:

Fit Measures	Symbol	Rule of Thumb
Coefficient of Determination	R^2	Greater than 0.7
Coefficient of Determination – Adjusted	R^2_{Adj}	Greater than 0.6
Standard Error of Estimate	SEE	Less than 0.25 when divided by \sqrt{y}
F Ratio	F_R	Greater than F table value for the test level of significance
Standard Error of Slope Coefficients	S_{bi}	Greater than t table value for the test level of significance
Standardized Residual	SRESID	Greater than 2
Leverage Value	LV	Greater than $3 * k / n$
Cook's D	D	Greater than 50th percentile of F distribution

Statistics which are Comparable Between Model Forms:

Predictive Measures	Symbol	Rule of Thumb
Coefficient of Variation Based upon the Standard Error	CV-SE	Less than 25%
Coefficient of Variation Based upon the MAD of the Residuals	CV-MARes	Less than 25%
Mean Absolute Deviation % Errors	MAD	Less than 25%



Typical Documentation



CO\$TAT has the Case utility. Can be exported to Excel.
 Important documentation for showing range of coefficients.

Calc	Type	Name	Status	Equation	DF	F-Prob	T-Prob Intercept	T-Prob b1	T-Prob b2	T-Prob b3	T-Prob b4	R ² Adj(%)	SE (fit space) ¹
<input type="checkbox"/>	Log Linear	Case 125	Calculated	\$/mile = 1.0376e-005 * var1P ^ 3.82	15	1.0000	0.9668	1.0000	0.9475	0.9996	1.0000	93.5826	0.5506
<input type="checkbox"/>	Log Linear	Case 71	Calculated	\$/mile = 0.0001523 * GVW ^ 0.5604	12	1.0000	0.9888	0.9296	0.9912	0.9999		90.6174	0.5576
<input type="checkbox"/>	Log Linear	Case 63	Calculated	\$/mile = 1.7483e-005 * GVW ^ 0.765	17	1.0000	0.9999	0.9976	0.9966	1.0000		92.1461	0.5759
<input type="checkbox"/>	Log Linear	Case 64	Calculated	\$/mile = 1.7483e-005 * GVW ^ 0.765	17	1.0000	0.9999	0.9976	0.9966	1.0000		92.1461	0.5759
<input type="checkbox"/>	Log Linear	Case 61	Calculated	\$/mile = 1.7149e-006 * GVW ^ 0.965	17	1.0000	1.0000	0.9997	0.9930	1.0000		92.9730	0.5768
<input type="checkbox"/>	Log Linear	Case 94	Calculated	\$/mile = 1.3445e-005 * GVW ^ 0.647	17	1.0000	1.0000	0.9928	0.9814	0.9319	0.9998	92.8923	0.5795
<input type="checkbox"/>	Log Linear	Case 103	Calculated	\$/mile = 6.8967e-007 * GVW ^ 0.653	17	1.0000	1.0000	0.9930	0.9794	0.9229	0.9999	92.8147	0.5826
<input type="checkbox"/>	Log Linear	Case 77	Calculated	\$/mile = 4.5910e-005 * GVW ^ 0.613	10	1.0000	0.9920	0.9270	0.6047	0.9636	0.9994	89.1117	0.5831
<input type="checkbox"/>	Log Linear	Case 102	Calculated	\$/mile = 3.6206e-008 * GVW ^ 0.657	17	1.0000	1.0000	0.9932	0.9782	0.9216	0.9999	92.7728	0.5843
<input type="checkbox"/>	Log Linear	Case 35	Calculated	Cost_mile = 1.9510e-007 * GVW ^ 1.	13	1.0000	0.9996	1.0000				72.2508	0.5873
<input type="checkbox"/>	Log Linear	Case 39	Calculated	Cost_mile = 1.9510e-007 * GVW ^ 1.	13	1.0000	0.9996	1.0000				72.2508	0.5873
<input type="checkbox"/>	Log Linear	Case 46	Calculated	Cost_mile = 1.9510e-007 * GVW ^ 1.	13	1.0000	0.9996	1.0000				72.2508	0.5873
<input type="checkbox"/>	Log Linear	Case 105	Calculated	\$/mile = 1.9510e-007 * GVW ^ 1.81	13	1.0000	0.9996	1.0000				72.2508	0.5873
<input type="checkbox"/>	Log Linear	Case 106	Calculated	\$/mile = 1.9510e-007 * GVW ^ 1.81	13	1.0000	0.9996	1.0000				72.2508	0.5873
<input type="checkbox"/>	Log Linear	Case 122	Calculated	\$/mile = 6.8457e-008 * var1P ^ 3.88	17	1.0000	1.0000	1.0000	0.9997	1.0000	0.9817	92.5298	0.5941
<input type="checkbox"/>	Log Linear	Case 59	Calculated	\$/mile = 2.5306e-006 * GVW ^ 0.924	16	1.0000	1.0000	0.9993	0.8538	0.9115	0.9988	92.4363	0.5978
<input type="checkbox"/>	Log Linear	Case 51	Calculated	Cost_mile = 1.8585e-007 * GVW ^ 2.	12	0.9997	0.9994	0.9457	0.3597			70.5033	0.6055
<input type="checkbox"/>	Log Linear	Case 92	Calculated	\$/mile = 2.7054e-005 * CW ^ 0.7995	16	1.0000	1.0000	0.9977	0.8695	0.7509	0.9611	92.0897	0.6113
<input type="checkbox"/>	Log Linear	Case 24	Calculated	Cost_mile = 9.8518e-008 * GVW ^ 1.	12	0.9998	0.9995	0.9999	0.5230			71.3532	0.6173
<input type="checkbox"/>	Log Linear	Case 87	Calculated	\$/mile = 0.0001176 * GVW ^ 0.8602	17	1.0000	0.9990	0.9985	0.9994	0.9997	0.9993	91.9288	0.6175
<input type="checkbox"/>	Log Linear	Case 127	Calculated	\$/mile = 0.0002267 * GVW ^ 0.811 *	16	1.0000	0.9969	0.9966	0.9989	0.9965	0.6051	91.8152	0.6218
<input type="checkbox"/>	Log Linear	Case 128	Calculated	\$/mile = 4.2187e-005 * GVW ^ 0.811	16	1.0000	0.9995	0.9966	0.9989	0.6051	0.9992	91.8152	0.6218
<input type="checkbox"/>	Log Linear	Case 89	Calculated	\$/mile = 1.9720e-005 * GVW ^ 0.87E	17	1.0000	0.9999	0.9984	0.8805	0.9971	0.9988	91.4183	0.6367
<input type="checkbox"/>	Log Linear	Case 99	Calculated	\$/mile = 1.9720e-005 * GVW ^ 0.87E	17	1.0000	0.9999	0.9984	0.8805	0.9971	0.9988	91.4183	0.6367
<input type="checkbox"/>	Log Linear	Case 100	Calculated	\$/mile = 1.9720e-005 * GVW ^ 0.87E	17	1.0000	0.9999	0.9984	0.8805	0.9971	0.9988	91.4183	0.6367
<input type="checkbox"/>	Log Linear	Case 101	Calculated	\$/mile = 1.9720e-005 * GVW ^ 0.87E	17	1.0000	0.9999	0.9984	0.8805	0.9971	0.9988	91.4183	0.6367
<input type="checkbox"/>	Log Linear	Case 22	Calculated	Cost_mile = 2.1636e-007 * Annual FC	11	0.9996	0.9987	0.1876	0.9991	0.0804		67.3290	0.6379
<input type="checkbox"/>	Log Linear	Case 9	Calculated	Cost_mile = 1.5363e-009 * GVW ^ 3.	4	0.8838	0.9412	0.8981	0.4501	0.5023		54.2086	0.6479



Typical Documentation



The documentation exported to ACEIT (shown below) or JIAT is shown here. As a stand-alone document, the file needs some further explanation.

A screenshot of a Microsoft Excel spreadsheet titled "AUV Example.xlsx" showing statistical analysis results. The ribbon includes "COSTAT" with various analysis options like Linear, Log Linear, Non Linear, Learning, Beta, and Univariate. The spreadsheet content is as follows:

Log Linear Analysis for Dataset Dataset, Case 5

I. Model Form and Equation Table

Model Form:	Unweighted Log-Linear model
Number of Observations Used:	22
Equation in Unit Space:	$\$/year = 2.9960e-027 * Weight^{0.8647} * Reliability^{11.83} * Mobility^{2.363} * Speed^{2.052}$

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	P-Value	Prob Not Zero
Intercept	-61.0725	14.9645		-4.0812	0.0008	0.9992
Weight	0.8647	0.2383	0.3104	3.6284	0.0021	0.9979
Reliability	11.8327	3.1232	0.2751	3.7886	0.0015	0.9985
Mobility	2.3627	0.6421	0.3840	3.6794	0.0019	0.9981
Speed	2.0519	0.4618	0.4846	4.4432	0.0004	0.9996

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
0.6490	92.78%	91.09%	0.9632

The spreadsheet also shows a taskbar at the bottom with several tabs: "Ridge Regression", "4a. Finished CER Ridge", "4. Finished CER", "3. WT MOB and Speed (SPD)", "2. WT and Mobility (MOB)", and "1. Wei". The status bar indicates "Ready" and "75%" zoom.



Benefits of the Stepwise Analysis Tool



Benefits of the Stepwise Analysis Tool



The screenshot shows the CO\$TAT Help window with the following content:

CO\$TAT Help

Hide Back Print Options

Contents | Index | Search | Glossary

- Getting Started
- Creating a Dataset
- Performing Analysis
 - Pairwise Analysis
 - Distribution Finder
 - Analyze Univariate
 - Analyze Beta
 - Analyze Learning
 - Analyze Linear
 - Analyze LogLinear
 - Analyze NonLinear
 - Analyze Summary Statistics
 - Applying Ridge Regression
 - Stepwise Analysis**
 - Generating Prediction Intervals
 - Performing Weighted Regression
 - Removing Outliers from the Data
 - Specifying Dummy Variables
- Evaluating Analysis
 - Available Statistics
 - Comparing Analysis Results
 - Differences between Fitted and Observed

Stepwise Analysis

Stepwise Analysis allows you to automate the process of performing regression. It is NOT [stepwise regression](#). Stepwise Analysis provides the evaluation of a single step towards improving a CER. The analyst first chooses a functional form and then identifies

1. One or more variables that should be in the CER and
2. Several others that may be useful to further improve the CER.

CO\$TAT adds one independent variable at a time into the equation and measures the improvement made by this additional variable.

The resulting Stepwise table lets you choose one or more of the new CERs and repeat the analysis the normal way to verify the new CER is in fact both acceptable and better than the simpler, original equation (analysts can compare them in the Case Analysis dialog).

If no dependent variables are identified initially, Stepwise Analysis begins with the independent variable that is most highly correlated with the dependent variable. For every independent variable selected for the analysis, CO\$TAT examines the impact of adding each variable into the regression equation.

Stepwise Analysis is available for both [Linear](#) and [Log-Linear](#) regression. When you select **Stepwise Analysis** in the Options section of the resulting dialog, CO\$TAT adds an additional column to the **Variables** section of the dialog where you can select the variables to be included in the Stepwise Analysis.

To perform Stepwise analysis, you must do the following:



Benefits of the Stepwise Analysis Tool

ODASA
Cost &
Economics

- Quicker way to generate results
- Can compare several cases at once
- More compact way to show steps
- Facilitates producing all the documentation in one file with multiple tabs
- Ability to quickly compare transformations



Example of Stepwise Analysis



Example of CO\$TAT Stepwise Analysis



Home Insert Page Layout

Criteria Prediction Intervals Report Styles

Dataset

	B	C	D
3	Observations	\$/year	Weight
4	System Name	\$/year	Weight
5	M121	39,674	60,8
6	M122	16,787	31,8
7	M123	1,903	14,7
8	M124	39,674	60,8
9	M125	16,787	31,8
10	M126	1,903	14,7
11	M127	8,237	12,3
12	M128	87,499	71,5
13	M129	39,152	33,0
14	M130	17,977	36,2
15	M131	14,731	33,0
16	M132	72,313	77,0
17	M133	25,085	28,5
18	M134	47,817	33,0
19	M135	89,117	66,0

Report New Dataset New Dat

Log Linear Model

Specifications Results

Case Name: Case 1

Dependent Variable: Name: \$/year Transform:

Weighting Variable:

Options: Ridge Parameter: Maximum Iterations: Method: MUPE Intercept (Non Origin) Stepwise Analysis

Report Precision: Precision: 4 Digits

Name	Not Used	Independ	Dummy	Stepwi
Weight	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Reliability	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Mobility	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Speed	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
HP_to_Wt_Ratio	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Maintainance_hours	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Availability	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Speed_XC	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
MR	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
MR_10	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
MR_20	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
MR_30	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
UA_15	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
\$/mile_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
GVW_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
NMC_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
WT_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
CT_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
ConRep_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Age_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
\$/mile_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
GVW_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
NMC_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
WT_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
CT_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
ConRep_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Age_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Var1	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>

OK Cancel Help



Example of CO\$TAT Stepwise Analysis



AUW Example.xlsx - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Approvalt Acrobat COSTAT

Criteria Prediction Intervals Report Styles Dataset

Pairwise Analysis Distribution Finder Cases

Linear Learning Update Report Examples About
 Log Linear Beta Clear Report Help Close
 Non Linear Univariate Copy To Clipboard Rules of Thumb Application

Q18

Log Linear Analysis for Dataset Dataset, Case 1

I. Model Form and Equation Table

Model Form:	Unweighted Log-Linear model
Number of Observations Used:	22
Equation in Unit Space:	$\$/\text{year} = 5.7043\text{e-}006 * \text{Weight}^{2.043}$

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	P-Value	Prob Not Zero
Intercept	-12.0743	4.3027		-2.8062	0.0109	0.9891
Weight	2.0428	0.4233	0.7335	4.8256	0.0001	0.9999

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
1.5139	53.80%	51.49%	0.7335

Analysis of Variance

Due To	DF	Sum of Sqr (SS)	Mean SQ = SS/DF	F-Stat	P-Value	Prob Not Zero
Regression	1	53.3727	53.3727	23.2861	0.0001	0.9999
Residual (Error)	20	45.8408	2.2920			
Total	21	99.2135				

Outlier Analysis Summary

4. Final CER 3. WT MOB and Speed (SPD) 2. WT and Mobility (MOB) 1. Weight (WT) Dataset Sheet2

Ready 75%



Example of CO\$TAT Stepwise Analysis



Here I will choose to add Mobility

AUW Example.xlsx - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Approvalt Acrobat COSTAT

fx Insert Function AutoSum Logical Lookup & Reference Define Name Trace Precedents Show Formulas
 Recently Used Text Math & Trig Use in Formula Trace Dependents Error Checking Watch Window
 Financial Date & Time More Functions Name Manager Create from Selection Remove Arrows Evaluate Formula Calculation Options
 Function Library Defined Names Formula Auditing Calculation

L59

66			19	5.7854	8.8911	-3.1057	0.3273	-2.1011	0.0467	0.1082	
67			20	6.5548	5.7143	0.8404	0.6851	0.6225	0.2048	0.0499	
68			21	5.8162	5.3251	0.4911	0.7571	0.3746	0.2501	0.0234	
69			22	4.2970	5.6104	-1.3134	0.7041	-0.9800	0.2163	0.1325	

SE = 15139, Mean = 8.6301, Coef. of Var. = 17.54% in Fit Space
 R denotes an observation with an unusual Dependent variable value.

Partial Correlation Coefficients of Independent Stepwise Variables

Ind Variable	Partial Correls	T-Statistic	P-Value	Equation	R-Squared (Adj)	MAD (%Error)
Original Equation				$\$/year = 5.7043e-006 * Weight ^ 2.043$	51.49%	336.93%
Mobility	0.7712	5.2802	0.0000	$\$/year = 0.06526 * Weight ^ 0.9716 * Mobility ^ 4$	79.30%	88.84%
Speed	0.7515	4.9654	0.0000	$\$/year = 3.4511e-005 * Weight ^ 0.9649 * Speed ^ 2.714$	77.77%	160.92%
Speed_XC	0.4131	1.9773	0.0626	$\$/year = 2.9857e-005 * Weight ^ 1.453 * Speed_XC ^ 1.589$	57.65%	274.91%
Availability	0.2280	1.0206	0.3202	$\$/year = 1.7980e-006 * Weight ^ 2.235 * Availability ^ 13.95$	51.59%	294.71%
HP_to_Wt_Ratio	0.2187	0.9770	0.3408	$\$/year = 1.5671e-008 * Weight ^ 2.017 * HP_to_Wt_Ratio ^ 2.152$	51.37%	304.62%
Reliability	0.1730	0.7658	0.4532	$\$/year = 4.6823e-017 * Weight ^ 2.166 * Reliability ^ 5.405$	50.46%	307.36%
Maintainance_hours	-0.5251	-2.6896	0.0145	$\$/year = 0.006409 * Weight ^ 1.426 * Maintainance_hours ^ (-0.6257)$	63.01%	196.09%

III. Predictive Measures (in Unit Space)

Percentage Error Table

Obs #	Name	Actuals	Predicted	Residuals	% Errors	Flags
1	M121	39673.5759	33856.3568	5817.2192	-14.6627	
2	M122	16787.0458	8997.2824	7789.7634	-46.4034	

4. Final CER 3. WT MOB and Speed (SPD) 2. Wt and Mobility (MOB) 1. Weight (WT) Dataset Sheet2

Ready 75%



Example of CO\$TAT Stepwise Analysis



Excel interface showing a dataset with columns: Observations, System Name, \$/year, and Weight.

Observations	System Name	\$/year	Weight
3			
4			
5	M121	39,674	60,8
6	M122	16,787	31,8
7	M123	1,903	14,7
8	M124	39,674	60,8
9	M125	16,787	31,8
10	M126	1,903	14,7
11	M127	8,237	12,3
12	M128	87,499	71,5
13	M129	39,152	33,0
14	M130	17,977	36,2
15	M131	14,731	33,0
16	M132	72,313	77,0
17	M133	25,085	28,5
18	M134	47,817	33,0
19	M135	89,117	66,0

Log Linear Model - Specifications tab

Case Name: Case 2

Dependent Variable Name: \$/year

Options: Stepwise Analysis, Intercept (Non Origin)

Report Precision: 4 Digits

Name	Not Used	Independence	Dummy	Stepwise
Weight	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Reliability	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Mobility	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Speed	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
HP_to_Wt_Ratio	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Maintenance_hours	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Availability	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Speed_XC	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
MR	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
MR_10	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
MR_20	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
MR_30	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
UA_15	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
\$/mile_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
GVW_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
NMC_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
WT_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
CT_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
ConRep_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Age_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
\$/mile_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
GVW_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
NMC_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
WT_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
CT_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
ConRep_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Age_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Var1	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>



Example of CO\$TAT Stepwise Analysis



AUW Example.xlsx - Microsoft Excel

Home Insert Page Layout Formulas Data Review View ApproveIt Acrobat CO\$TAT

Criteria Prediction Intervals Report Styles Dataset

Pairwise Distribution Analysis Cases

Linear Learning Update Report Examples About
Log Linear Beta Clear Report Help Close
Non Linear Univariate Copy To Clipboard Rules of Thumb Application

Q6

Log Linear Analysis for Dataset Dataset, Case 2

I. Model Form and Equation Table

Model Form:	Unweighted Log-Linear model
Number of Observations Used:	22
Equation in Unit Space:	$\$year = 0.06526 * weight^{0.9716} * Mobility^{0.4}$

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	P-Value	Prob Not Zero
Intercept	-2.7294	3.3212		-0.8218	0.4213	0.5787
Weight	0.9716	0.3430	0.3488	2.8330	0.0106	0.9894
Mobility	4.0004	0.7576	0.6502	5.2802	0.0000	1.0000

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
0.9889	81.27%	79.30%	0.9015

Analysis of Variance

Due To	DF	Sum of Sqr (SS)	Mean SQ = SS/DF	F-Stat	P-Value	Prob Not Zero
Regression	2	80.6347	40.3174	41.2314	0.0000	1.0000
Residual (Error)	19	18.5788	0.9778			
Total	21	99.2135				

Further Analysis of Variance

(SS explained by each variable when entered in the order given)

Ready 4. Final CER 3. WT MOB and Speed (SPD) 2. Wt and Mobility (MOB) 1. Weight (WT) Dataset Sheet2 75%



Example of CO\$TAT Stepwise Analysis



Next I might add speed

A screenshot of Microsoft Excel showing a CO\$TAT analysis. The window title is "AUW Example.xlsx - Microsoft Excel". The ribbon includes "Home", "Insert", "Page Layout", "Formulas", "Data", "Review", "View", "Approval", "Acrobat", and "COSTAT". The "COSTAT" ribbon has sections for "Dataset", "Analysis", and "Application".

The active cell is T95. The data table below shows the results of a stepwise regression analysis:

	D	E	F	G	H	I	J	K	L
85	12	11.3977	10.8254	0.5723	0.3422	0.6168	0.1198	0.0173	
86	13	9.4047	7.0326	2.3721	0.3424	2.5570	0.1199	0.2969	R
87	14	9.3132	10.0180	-0.7049	0.3012	-0.7484	0.0928	0.0191	
88	15	8.3040	6.9252	1.3788	0.3288	1.4784	0.1105	0.0905	
89	16	6.7296	7.6586	-0.9290	0.4772	-1.0726	0.2328	0.1164	
90	17	5.7478	6.6290	-0.8812	0.3127	-0.9394	0.1000	0.0327	
91	18	6.1853	7.3339	-1.1486	0.3977	-1.2687	0.1618	0.1036	
92	19	5.7854	7.2418	-1.4564	0.3785	-1.5942	0.1465	0.1454	
93	20	6.5548	5.7309	0.8239	0.4475	0.9343	0.2048	0.0749	
94	21	5.8162	5.5458	0.2704	0.4963	0.3162	0.2519	0.0112	
95	22	4.2970	5.6815	-1.3844	0.4601	-1.5817	0.2165	0.2304	

SE = 0.9889, Mean = 8.6301, Coef. of Var. = 11.46% in Fit Space
R denotes an observation with an unusual Dependent variable value.

Partial Correlation Coefficients of Independent Stepwise Variables

Ind Variable	Partial Correls	T-Statistic	P-Value	Equation	R-Squared (Adj)	MAD (%Error)
Original Equation				\$/year = 0.06526 * Weight ^ 0.9716 * Mobility ^ 4	79.30%	88.84%
Speed	0.5378	2.7063	0.0144	\$/year = 0.006502 * Weight ^ 0.7249 * Mobility ^ 2.561 * Speed ^ 1.591	84.47%	75.58%
Reliability	0.4087	1.9000	0.0735	\$/year = 1.5604e-018 * Weight ^ 1.118 * Reliability ^ 8.179 * Mobility ^ 4.151	81.80%	82.37%
Availability	0.3000	1.3341	0.1988	\$/year = 0.02216 * Weight ^ 1.146 * Mobility ^ 3.953 * Availability ^ 11.7	80.12%	87.30%
Speed_XC	0.2561	1.1243	0.2756	\$/year = 0.06445 * Weight ^ 0.8051 * Mobility ^ 3.698 * Speed_XC ^ 0.6662	79.59%	83.93%
HP_to_Wt_Ratio	0.0906	0.3861	0.7040	\$/year = 0.01143 * Weight ^ 0.9819 * Mobility ^ 3.936 * HP_to_Wt_Ratio ^ 0.5808	78.33%	86.57%
Maintenance_hours	-0.1858	-0.8023	0.4328	\$/year = 0.1684 * Weight ^ 0.914 * Mobility ^ 3.593 * Maintenance_hours ^ (-0.1692)	78.91%	90.99%

III. Predictive Measures (in Unit Space)

Percentage Error Table

Navigation: 4. Final CER | 3. WT MOB and Speed (SPD) | 2. Wt and Mobility (MOB) | 1. Weight (WT) | Dataset | Sheet2



Example of CO\$TAT Stepwise Analysis



Excel interface showing a dataset with columns B, C, and D. The data is as follows:

	B	C	D
3	Observations	\$/year	Weight
4	System Name	\$/year	Weight
5	M121	39,674	60,
6	M122	16,787	31,
7	M123	1,903	14,
8	M124	39,674	60,
9	M125	16,787	31,
10	M126	1,903	14,
11	M127	8,237	12,
12	M128	87,499	71,
13	M129	39,152	33,
14	M130	17,977	36,
15	M131	14,731	33,
16	M132	72,313	77,
17	M133	25,085	28,
18	M134	47,817	33,
19	M135	89,117	66,

Log Linear Model Specifications Results

Case Name: Case 3

Dependent Variable: Name: \$/year

Weighting Variable:

Options:

- Ridge Parameter: []
- Maximum Iterations: []
- Method: []
- MUPE Intercept (Non Origin)
- Stepwise Analysis

Report Precision: Precision: 4 Digits

Variables:

Name	Not Used	Independe	Dummy	Stepwi
Weight	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Reliability	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Mobility	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Speed	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
HP_to_Wt_Ratio	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Maintainance_hours	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Availability	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Speed_XC	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
MR	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
MR_10	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
MR_20	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
MR_30	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
UA_15	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
\$/mile_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
GWV_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
NMC_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
WT_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
CT_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
ConRep_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Age_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
\$/mile_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
GWV_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
NMC_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
WT_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
CT_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
ConRep_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Age_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Var1	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>

Buttons: OK, Cancel, Help



Example of CO\$TAT Stepwise Analysis



AUW Example.xlsx - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Apprevlt Acrobat COSTAT

Criteria Prediction Intervals Report Styles Dataset

Pairwise Analysis Distribution Finder Cases

Linear Log Linear Non Linear Learning Beta Univariate Analysis

Update Report Clear Report Copy To Clipboard

Examples Help Rules of Thumb About Close Application

N6

Log Linear Analysis for Dataset Dataset, Case 3

I. Model Form and Equation Table

Model Form:	Unweighted Log-Linear model
Number of Observations Used:	22
Equation in Unit Space:	$\$/\text{year} = 0.006502 * \text{Weight}^{0.7249} * \text{Mobility}^{2.561} * \text{Speed}^{1.591}$

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	P-Value	Prob Not Zero
Intercept	-5.0357	3.0003		-1.6784	0.1104	0.8896
Weight	0.7249	0.3107	0.2603	2.3330	0.0314	0.9686
Mobility	2.5613	0.8447	0.4163	3.0323	0.0071	0.9929
Speed	1.5912	0.5880	0.3758	2.7063	0.0144	0.9856

Goodness-of-Fit Statistics

Std Error (SE)	R-Squared	R-Squared (Adj)	Pearson's Corr Coef
0.8565	86.69%	84.47%	0.9311

Analysis of Variance

4. Final CER 3. WT MOB and Speed (SPD) 2. Wt and Mobility (MOB) 1. Weight (WT) Report Dataset Dataset Sheet2

Ready 75%



Example of CO\$TAT Stepwise Analysis



- Multiple criteria for stopping
 - No further improvement possible
 - Degrees of freedom
 - Multicollinearity
 - Too many variables (variables lose meaning)
 - How will the results be used
- Stopping criteria is a best practices and peer review topic



CO\$TAT Stepwise Analysis (Note: Step 0)



Step 0: When no independent variables are selected, CO\$TAT will automatically “recommend” the best base variable.

Name	Not Used	Indepe	Dummy	Stepwi
Weight	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Reliability	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Mobility	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Speed	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
HP_to_Wt_Rat	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
Maintenance_h	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>

A1 Costat Report

Partial Correlation Coefficients of Independent Stepwise Variables

Ind Variable	Partial Correls	T-Statistic	P-Value	Equation	R-Squared (Adj)	MAD (%Error)
Mobility				This variable was the most highly correlated, and has been used for the regressor	72.03%	118.86%
Speed	0.5909	3.1929	0.0048	$\$/year = 2.157 * Mobility^3.063 * Speed^1.994$	80.84%	87.08%
Weight	0.5449	2.8330	0.0106	$\$/year = 0.06526 * Weight^0.9716 * Mobility^4$	79.30%	88.84%
Speed_XC	0.4142	1.9836	0.0619	$\$/year = 43.7 * Mobility^4.351 * Speed_XC^1.178$	75.61%	96.18%
Reliability	0.2067	0.9209	0.3686	$\$/year = 3.0889e-007 * Reliability^4.798 * Mobility^5.47$	71.82%	123.41%
Availability	0.0374	0.1631	0.8722	$\$/year = 829.1 * Mobility^5.295 * Availability^1.621$	70.60%	120.21%
HP_to_Wt_Ratio	0.0345	0.1506	0.8819	$\$/year = 362.6 * Mobility^5.247 * HP_to_Wt_Ratio^0.2632$	70.60%	118.82%
Maintenance_hours	-0.2632	-1.1890	0.2491	$\$/year = 1461 * Mobility^4.472 * Maintenance_hours^(-0.2798)$	72.60%	119.55%

4. Final CER | 3. WT MOB and Speed (SPD) | 2. Wt and Mobility (MOB) | 1. Weight (WT) | Report Dataset | Dataset | Sheet2



CO\$TAT Stepwise Analysis (Transforms)



Ability to compare variable transforms quickly

Observations	MR	MR+10	MR+20	MR+30
System Name	MR	MR_10	MR_20	MR_30
M121	18.00	28.00	38.00	48.00
M122	15.60	25.60	35.60	45.60
M123	7.20	17.20	27.20	37.20
M124	13.20	23.20	33.20	43.20
M125	12.00	22.00	32.00	42.00
M126	8.40	18.40	28.40	38.40
M127	7.20	17.20	27.20	37.20
M128	9.60	19.60	29.60	39.60
M129	8.40	18.40	28.40	38.40
M130	20.40	30.40	40.40	50.40
M131	4.80	14.80	24.80	34.80
M132	13.20	23.20	33.20	43.20
M133	2.40	12.40	22.40	32.40
M134	10.80	20.80	30.80	40.80
M135	9.60	19.60	29.60	39.60

Log Linear Model

Specifications | Results

Case Name: Case 5

Dependent Variable: Name: \$/year, Transform: [Dropdown]

Weighting Variable: [Dropdown]

Options: Ridge Parameter: [Input], Maximum Iterations: [Dropdown], Method: [Dropdown], MUPE, Intercept (Non Origin), Stepwise Analysis

Report Precision: Precision: 4 Digits

Variables:

Name	Not	Indepe	Dummy	Ste
Weight	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Reliability	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Mobility	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Speed	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
HP_to_Wt_Ratio	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Maintenance_hours	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Availability	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Speed_XC	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
MR	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
MR_10	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
MR_20	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
MR_30	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	<input checked="" type="checkbox"/>
UA_15	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
\$/mile_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
GVW_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
NMC_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
WT_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
CT_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
ConRep_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Age_T	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
\$/mile_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
GVW_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
NMC_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
WT_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
CT_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
ConRep_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Age_TL	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Var1	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>
Var2	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="checkbox"/>



CO\$TAT Stepwise Analysis (Transforms)



Ability to compare variable transforms quickly

Costat Report

SE = 0.8565, Mean = 8.6301, Coef. of Var. = 9.92% in Fit Space
 R denotes an observation with an unusual Dependent variable value.
 D denotes an observation with an unusual influence on the fitted regression equation.

Partial Correlation Coefficients of Independent Stepwise Variables

Ind Variable	Partial Correls	T-Statistic	P-Value	Equation	R-Squared (Adj)	MAD (%Error)
Original Equation				$\$/\text{year} = 0.006502 * \text{Weight} \wedge 0.7249 * \text{Mobility} \wedge 2.561 * \text{Speed} \wedge 1.591$	84.47%	75.58%
MR_30	-0.7365	-3.9257	0.0017	$\$/\text{year} = 610553 * \text{Weight} \wedge 0.8531 * \text{Mobility} \wedge 2.502 * \text{Speed} \wedge 1.642 * \text{MR}_30 \wedge (-5.32)$	87.99%	45.41%
MR_20	-0.7398	-3.9641	0.0016	$\$/\text{year} = 1588 * \text{Weight} \wedge 0.8553 * \text{Mobility} \wedge 2.526 * \text{Speed} \wedge 1.621 * \text{MR}_20 \wedge (-4.02)$	88.11%	44.98%
MR_10	-0.7456	-4.0341	0.0014	$\$/\text{year} = 6.137 * \text{Weight} \wedge 0.8591 * \text{Mobility} \wedge 2.579 * \text{Speed} \wedge 1.576 * \text{MR}_10 \wedge (-2.685)$	88.34%	44.15%
MR	-0.7588	-4.2008	0.0010	$\$/\text{year} = 0.0464 * \text{Weight} \wedge 0.866 * \text{Mobility} \wedge 2.808 * \text{Speed} \wedge 1.392 * \text{MR} \wedge (-1.205)$	88.86%	40.91%

III. Predictive Measures (in Unit Space)

Percentage Error Table



Advanced Topics



Advanced Topics



Developing best practices for advanced topics:

- Handling Multicollinearity
- Missing Data
- Other advanced topics



Handling Multicollinearity

- True multicollinearity can be avoided
- Residual multicollinearity can be explained
- Should be investigated even when there are no CO\$TAT flags

The screenshot shows the Microsoft Excel interface with the COSTAT ribbon active. The main content area displays two tables: a Pairwise Correlation Matrix and a Multicollinearity Analysis table. Below the tables is a legend explaining the 'X' flag.

Pairwise Correlation Matrix

Variables	cost	weight	age	reliability
cost	1.0000	0.9774	-0.9913	0.9946
weight	0.9774	1.0000	-0.9419	0.9505
age	-0.9913	-0.9419	1.0000	-0.9396
reliability	0.9946	0.9505	-0.9396	1.0000

Multicollinearity Analysis

Indep Variables	Indiv R-Sqr (%)	F-Stats	Prob Related to Other Vars	Indiv R-Sqr/Model R-Sqr	Flags
weight	99.49%	387.6945	1.0000	0.9949	X
age	100.00%	49580.2320	1.0000	1.0000	X
reliability	100.00%	58024.8576	1.0000	1.0000	X

X = The indicated independent variable could be harmfully correlated to the other independent variables, i.e., it has a nearly better fit using the remaining independent variables than the dependent variable.

Excel status bar: Average: 1.569204167 Count: 46 Sum: 37.6609 75%



Handling Multicollinearity

Consider:

- ACEIT recommendations
- ACEIT Flags
- Defense Acquisition University (DAU) guidance?
- What level of Multicollinearity is OK?
- Is the estimated system in the range of the data?
- Should the coefficients in the CER stand alone?
- What are the predicted values?



Handling Multicollinearity (Lessons Learned)



Techniques:

- Get more data points
- Use fewer variables
- Get better explanatory variables
- Consider meaningful transforms and combinations of the variables
- Ridge Regression
- Principal Component Analysis



Handling Multicollinearity (Explain away)

In this example, we can make a case that the exponent of GVW is meaningful.

Calc	Type	Name	Status	Equation	DF	F-Prob	T-Prob Intercept	T-Prob b1	T-Prob b2	T-Prob b3	T-Prob b4	R ² Adj(%)	SE (fit space) [^] 1
<input type="checkbox"/>	Log Linear	Case 125	Calculated	\$/mile = 1.0376e-005 * var1P ^ 3.82	15	1.0000	0.9668	1.0000	0.9475	0.9996	1.0000	93.5826	0.5506
<input type="checkbox"/>	Log Linear	Case 71	Calculated	\$/mile = 0.0001523 * GVW ^ 0.5604	12	1.0000	0.9888	0.9296	0.9912	0.9999		90.6174	0.5576
<input type="checkbox"/>	Log Linear	Case 63	Calculated	\$/mile = 1.7483e-005 * GVW ^ 0.765	17	1.0000	0.9999	0.9976	0.9966	1.0000		92.1461	0.5759
<input type="checkbox"/>	Log Linear	Case 64	Calculated	\$/mile = 1.7483e-005 * GVW ^ 0.765	17	1.0000	0.9999	0.9976	0.9966	1.0000		92.1461	0.5759
<input type="checkbox"/>	Log Linear	Case 61	Calculated	\$/mile = 1.7149e-006 * GVW ^ 0.965	17	1.0000	1.0000	0.9997	0.9930	1.0000		92.9730	0.5768
<input type="checkbox"/>	Log Linear	Case 94	Calculated	\$/mile = 1.3445e-005 * GVW ^ 0.647	17	1.0000	1.0000	0.9928	0.9814	0.9319	0.9998	92.8923	0.5795
<input type="checkbox"/>	Log Linear	Case 103	Calculated	\$/mile = 6.8967e-007 * GVW ^ 0.653	17	1.0000	1.0000	0.9930	0.9794	0.9229	0.9999	92.8147	0.5826
<input type="checkbox"/>	Log Linear	Case 77	Calculated	\$/mile = 4.5910e-005 * GVW ^ 0.613	10	1.0000	0.9920	0.9270	0.6047	0.9636	0.9994	89.1117	0.5831
<input type="checkbox"/>	Log Linear	Case 102	Calculated	\$/mile = 3.6206e-008 * GVW ^ 0.657	17	1.0000	1.0000	0.9932	0.9782	0.9216	0.9999	92.7728	0.5843
<input type="checkbox"/>	Log Linear	Case 35	Calculated	Cost_mile = 1.9510e-007 * GVW ^ 1.	13	1.0000	0.9996	1.0000				72.2508	0.5873
<input type="checkbox"/>	Log Linear	Case 39	Calculated	Cost_mile = 1.9510e-007 * GVW ^ 1.	13	1.0000	0.9996	1.0000				72.2508	0.5873
<input type="checkbox"/>	Log Linear	Case 46	Calculated	Cost_mile = 1.9510e-007 * GVW ^ 1.	13	1.0000	0.9996	1.0000				72.2508	0.5873
<input type="checkbox"/>	Log Linear	Case 105	Calculated	\$/mile = 1.9510e-007 * GVW ^ 1.81	13	1.0000	0.9996	1.0000				72.2508	0.5873
<input type="checkbox"/>	Log Linear	Case 106	Calculated	\$/mile = 1.9510e-007 * GVW ^ 1.81	13	1.0000	0.9996	1.0000				72.2508	0.5873
<input type="checkbox"/>	Log Linear	Case 122	Calculated	\$/mile = 6.8457e-008 * var1P ^ 3.88	17	1.0000	1.0000	1.0000	0.9997	1.0000	0.9817	92.5298	0.5941
<input type="checkbox"/>	Log Linear	Case 59	Calculated	\$/mile = 2.5306e-006 * GVW ^ 0.924	16	1.0000	1.0000	0.9993	0.8538	0.9115	0.9988	92.4363	0.5978
<input type="checkbox"/>	Log Linear	Case 51	Calculated	Cost_mile = 1.8585e-007 * GVW ^ 2.	12	0.9997	0.9994	0.9457	0.3597			70.5033	0.6055
<input type="checkbox"/>	Log Linear	Case 92	Calculated	\$/mile = 2.7054e-005 * CW ^ 0.7995	16	1.0000	1.0000	0.9977	0.8695	0.7509	0.9611	92.0897	0.6113
<input type="checkbox"/>	Log Linear	Case 24	Calculated	Cost_mile = 9.8518e-008 * GVW ^ 1.	12	0.9998	0.9995	0.9999	0.5230			71.3532	0.6173
<input type="checkbox"/>	Log Linear	Case 87	Calculated	\$/mile = 0.0001176 * GVW ^ 0.8602	17	1.0000	0.9990	0.9985	0.9994	0.9997	0.9993	91.9288	0.6175
<input type="checkbox"/>	Log Linear	Case 127	Calculated	\$/mile = 0.0002267 * GVW ^ 0.811 *	16	1.0000	0.9969	0.9966	0.9989	0.9965	0.6051	91.8152	0.6218
<input type="checkbox"/>	Log Linear	Case 128	Calculated	\$/mile = 4.2187e-005 * GVW ^ 0.811	16	1.0000	0.9995	0.9966	0.9989	0.6051	0.9992	91.8152	0.6218
<input type="checkbox"/>	Log Linear	Case 89	Calculated	\$/mile = 1.9720e-005 * GVW ^ 0.875	17	1.0000	0.9999	0.9984	0.8805	0.9971	0.9988	91.4183	0.6367
<input type="checkbox"/>	Log Linear	Case 99	Calculated	\$/mile = 1.9720e-005 * GVW ^ 0.875	17	1.0000	0.9999	0.9984	0.8805	0.9971	0.9988	91.4183	0.6367
<input type="checkbox"/>	Log Linear	Case 100	Calculated	\$/mile = 1.9720e-005 * GVW ^ 0.875	17	1.0000	0.9999	0.9984	0.8805	0.9971	0.9988	91.4183	0.6367
<input type="checkbox"/>	Log Linear	Case 101	Calculated	\$/mile = 1.9720e-005 * GVW ^ 0.875	17	1.0000	0.9999	0.9984	0.8805	0.9971	0.9988	91.4183	0.6367
<input type="checkbox"/>	Log Linear	Case 22	Calculated	Cost_mile = 2.1636e-007 * Annual FC	11	0.9996	0.9987	0.1876	0.9991	0.0804		67.3290	0.6379
<input type="checkbox"/>	Log Linear	Case 9	Calculated	Cost_mile = 1.5363e-009 * GVW ^ 3.	4	0.8838	0.9412	0.8981	0.4501	0.5023		54.2086	0.6479



Handling Multicollinearity (Ridge Regression)



- Already in CO\$TAT
- Applies a penalty to sum of squares based on the sum of squares of the coefficients
- Can be used to reduce the effects of multicollinearity, such as counterintuitive coefficients



Handling Multicollinearity (Ridge Regression)

The Excel spreadsheet displays a dataset with 7 observations. The columns are labeled 'Observations', 'cost', 'weight', and 'age'. The data points are as follows:

Observations	cost	weight	age
Vehicle 1	1	2	
Vehicle 2	2	5	1.58578
Vehicle 3	3	10	1.26794
Vehicle 4	4	17	
Vehicle 5	5	26	0.76393
Vehicle 6	6	37	0.5505
Vehicle 7	7	50	0.35424

The Linear Model dialog box is open, showing the Specifications tab. The Case Name is 'Case 32'. The Dependent Variable is 'cost'. The Weighting Variable is empty. The Ridge Parameter is set to 0.02, and the Maximum Iterations is 30. The Method is set to Modified Marquardt. The MUPE and Stepwise Analysis options are unchecked. The Report Precision is set to 4 Digits. The Variables table is as follows:

Name	Not	Indepen	Dummy
weight	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
age	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
reliability	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>



Handling Multicollinearity (Ridge Regression)

PCA 1.xlsx - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Approvalt Acrobat COSTAT

Criteria Prediction Intervals Report Styles
New Properties Dataset

Pairwise Distribution Analysis Cases
Finder

Linear Learning Update Report
Log Linear Beta Clear Report
Non Linear Univariate Copy To Clipboard
Analysis

Examples About
Help Close
Rules of Thumb Application

J182

Ridge Parameters	BETA (1)	BETA (2)	BETA (3)	SSE
0.00	0.1172	2.4853	3.3675	0.0000
0.02	0.3518	-0.3224	0.3313	0.0103
0.04	0.3459	-0.3235	0.3293	0.0143
0.06	0.3411	-0.3231	0.3277	0.0207
0.08	0.3370	-0.3222	0.3261	0.0293
0.10	0.3334	-0.3210	0.3244	0.0402
0.12	0.3301	-0.3196	0.3226	0.0532
0.14	0.3270	-0.3180	0.3209	0.0682
0.16	0.3242	-0.3164	0.3191	0.0851
0.18	0.3215	-0.3148	0.3172	0.1039
0.20	0.3189	-0.3131	0.3154	0.1245
0.22	0.3165	-0.3113	0.3136	0.1469
0.24	0.3141	-0.3096	0.3117	0.1708
0.26	0.3118	-0.3079	0.3099	0.1964
0.28	0.3095	-0.3061	0.3081	0.2235
0.30	0.3074	-0.3044	0.3063	0.2521

VII. Charts

Report New Dataset New Dataset Collinear dataset Sheet7

Select destination and press ENTER or choose Paste

75%



Handling Multicollinearity (Ridge Regression)



This example shows the somewhat dramatic effects on the regression with the application of a Ridge Parameter (Case 2). Notice the coefficients in the equation.

Ridge Regressions.xlsx - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Approvalt Acrobat COSTAT

Criteria Prediction Intervals Report Styles Dataset

Pairwise Analysis Distribution Finder Cases

Linear Learning Log Linear Beta Non Linear Univariate Analysis

Update Report Clear Report Copy To Clipboard

Examples Help Rules of Thumb Application

J39

Linear Analysis for Dataset New Dataset, Case 1

I. Model Form and Equation Table

Model Form:	Unweighted Linear model
Number of Observat	7
Equation in Unit Sp	$cost = (-54.3) + 0.01432 * weight + 9.184 * age + 9.226 * reliability$

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	P-Value	Prob Not Zero
Intercept	-54.2990	1.4395		-37.7202	0.0000	0.9999
weight	0.0143	0.0008	0.1172	17.6417	0.0004	0.9996
age	9.1835	0.2769	2.4853	33.1656	0.0000	0.9999
reliability	9.2260	0.2221	3.3675	41.5399	0.0000	1.0000

Linear Analysis for Dataset New Dataset, Case 2

I. Model Form and Equation Table

Model Form:	Unweighted Linear model
Number of Observat	7
Equation in Unit Sp	$cost = (-0.4188) + 0.0434 * weight + (-1.181) * age + 0.9155 * reliability$
Ridge Parameter Sp	0.0100

II. Fit Measures (in Fit Space)

Coefficient Statistics Summary

Variable	Coefficient	Std Dev of Coef	Beta Value	T-Statistic (Coef/SD)	P-Value	Prob Not Zero
Intercept	-0.4188	31.7533		-0.0132	0.9903	0.0097
weight	0.0434	0.0002	0.3553	275.8771	0.0000	1.0000
age	-1.1805	0.0027	-0.3195	-433.9550	0.0000	1.0000
reliability	0.9155	0.0016	0.3342	564.5355	0.0000	1.0000

F1_5x5_PCA F2_6x6_PCA G_PCA5VarOut H_Results Report New Dataset New Dataset Collinear dataset Sheet7

Ready 100%



Handling Multicollinearity – Principal Component Analysis (PCA)

ODASA
Cost &
Economics

- Transforms the variables into new orthogonal variables
- Computationally tricky (requires calculating eigenvalues of the pairwise correlation matrix)
- New variables lose meaning, but the cumulative loading of each variable can be used to recommend suitable variable bases



Handling Multicollinearity – Principal Component Analysis (PCA)

Transformed variables have no pairwise correlation.

Pairwise Correlation Matrix

Variables	\$/mile	Var1	var2	var3	var4	var5
\$/mile	1.0000	0.7374	-0.0288	0.5393	0.1262	0.1382
Var1	0.7374	1.0000	0.0000	0.0000	0.0000	-0.0002
var2	-0.0288	0.0000	1.0000	0.0000	0.0000	0.0000
var3	0.5393	0.0000	0.0000	1.0000	0.0000	0.0000
var4	0.1262	0.0000	0.0000	0.0000	1.0000	0.0000
var5	0.1382	-0.0002	0.0000	0.0000	0.0000	1.0000

Multicollinearity Analysis

Indep Variables	Indiv R-Sqr (%)	F-Stats	Prob Related to Other Vars	Indiv R-Sqr/Model R-Sqr	Flags
Var1	0.00%	0.0000	0.0000	0.0000	
var2	0.00%	0.0000	0.0000	0.0000	
var3	0.00%	0.0000	0.0000	0.0000	
var4	0.00%	0.0000	0.0000	0.0000	
var5	0.00%	0.0000	0.0000	0.0000	

Outlier Analysis Summary



Incomplete Data

- Most users believe in throwing out the observation or disregarding the variable
- As is, CO\$TAT treats the missing point as a 0
- Other techniques are available
- On a limited basis, the data can still be used acceptably
- Working on techniques and best practices for this case



Other Advanced Topics

- Sensitivity analysis
 - Input data
 - Omission of a data point
 - Communicating sensitivity results
- Pooled Regression and Dummy Variables (tomorrow)
- Using Minimum-Unbiased-Percentage Error (CO\$TAT feature)



Developing Best Practices



Developing Best Practices

ODASA
Cost &
Economics

- Best Practices in CO\$TAT is extremely challenging
- Rules are difficult to quantify
- Best practices should consist of a series of “world-class” well-documented illustrations
- Target experienced users
- Data sources and documentation of data sources is critical
- Use peer reviews



Summary

- CO\$TAT Stepwise Analysis tool benefits the analyst
 - Time and effectiveness
 - Documentation
 - Defending results
- Consider building best practices for Stepwise Analysis
 - General guidance
 - Advanced topics (expand the analyst toolset)
- Peer review is a critical component of best practices



Questions

ODASA
Cost &
Economics