RegressIt   # User Manual[1]

RegressIt_TM is a free Excel add-in for multivariate descriptive data analysis and linear and logistic regression, first released to the public in 2014 at **regressit.com** and significantly enhanced since then. It was originally developed at Duke University for teaching regression and time series analysis in MBA courses but is also intended for use at other levels of statistics education and in applications. This document[1] gives an overview of the program's features. More advice and examples, installation instructions, and tech support information can be found on the web site.

The latest version features a novel ribbon interface which controls every aspect of generating and navigating and presenting your data analysis, as well as providing access to internal documentation and teaching materials.   It is not necessary to use the native Excel ribbon or scrolling and paging tools nor do any typing during a RegressIt session apart from optional customization of analysis names.   The ribbon interface is designed so that the entire analysis can be driven efficiently not only with a mouse but also by tapping one's finger on a touchscreen, and the output files are easily viewable and navigable on smart phones. If you hover the mouse over a button you will see a pop-up description of how to use it. *These navigation tools are described in detail on pages 11-15.*
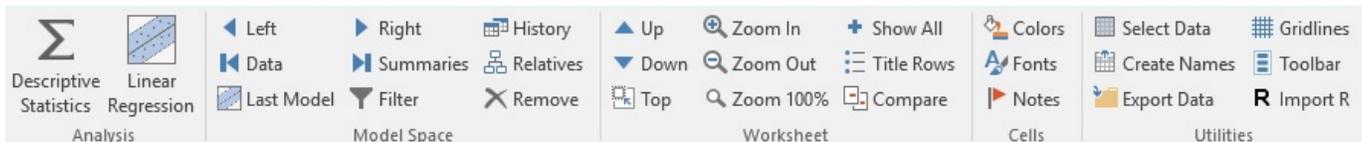
There are linear regression versions of the program for both PC's and Macs, and they have identical features and can exchange analysis files. A separate version that performs both linear and logistic regression, with highly interactive table and chart output for logistic models, runs on PC's. All versions of the program include an interface with R that allows both linear and logistic models to be fitted and to generate output in both RStudio and Excel, without any writing of R code. This document illustrates the use of the descriptive analysis and linear regression tools. The additional features of the logistic version are described in other documents and the **logistic regression pages** on the web site.

RegressIt can be launched either by opening the program file when needed[2] or by installing it on the Excel add-ins menu. (See the "Installation instructions" document for details of the latter method.) When it is running, it adds the ribbon pictured above to Excel's top menu. The two leftmost buttons launch the main analytic procedures:
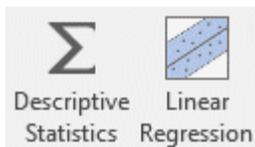

From the dialog boxes of both procedures, you can launch the **variable transformation** procedure.

DESCRIPTIVE STATISTICS PROCEDURE: This procedure always produces a descriptive stats table and correlation matrix. There are also options for autocorrelations, series plots, scatterplots, and histogram plots. There are several options for detail to be included on scatterplots, such as regression lines and center-of-mass points.   *See pages 6-10 for screen shots of RegressIt output worksheets.*

---

[1] February 27, 2023, Robert Nau. Visit **regressit.com** for the free software and complete documentation and sample data.
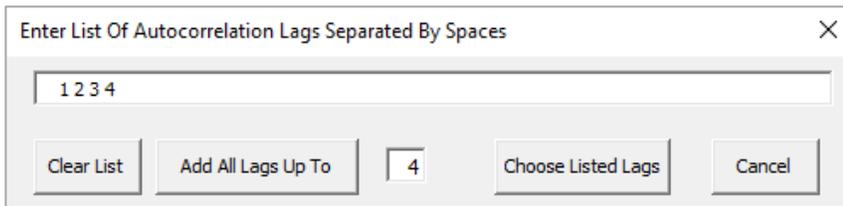[2] Excel's macro security level in the Trust Center Settings must be set to "Disable all macros with notification" and you must choose "Enable macros" at the security prompt when the program file is opened. On a PC, a newly downloaded program file must be "unblocked" before it will work for the first time. Right-click on the file and choose "Properties/Unblock/Apply/OK.
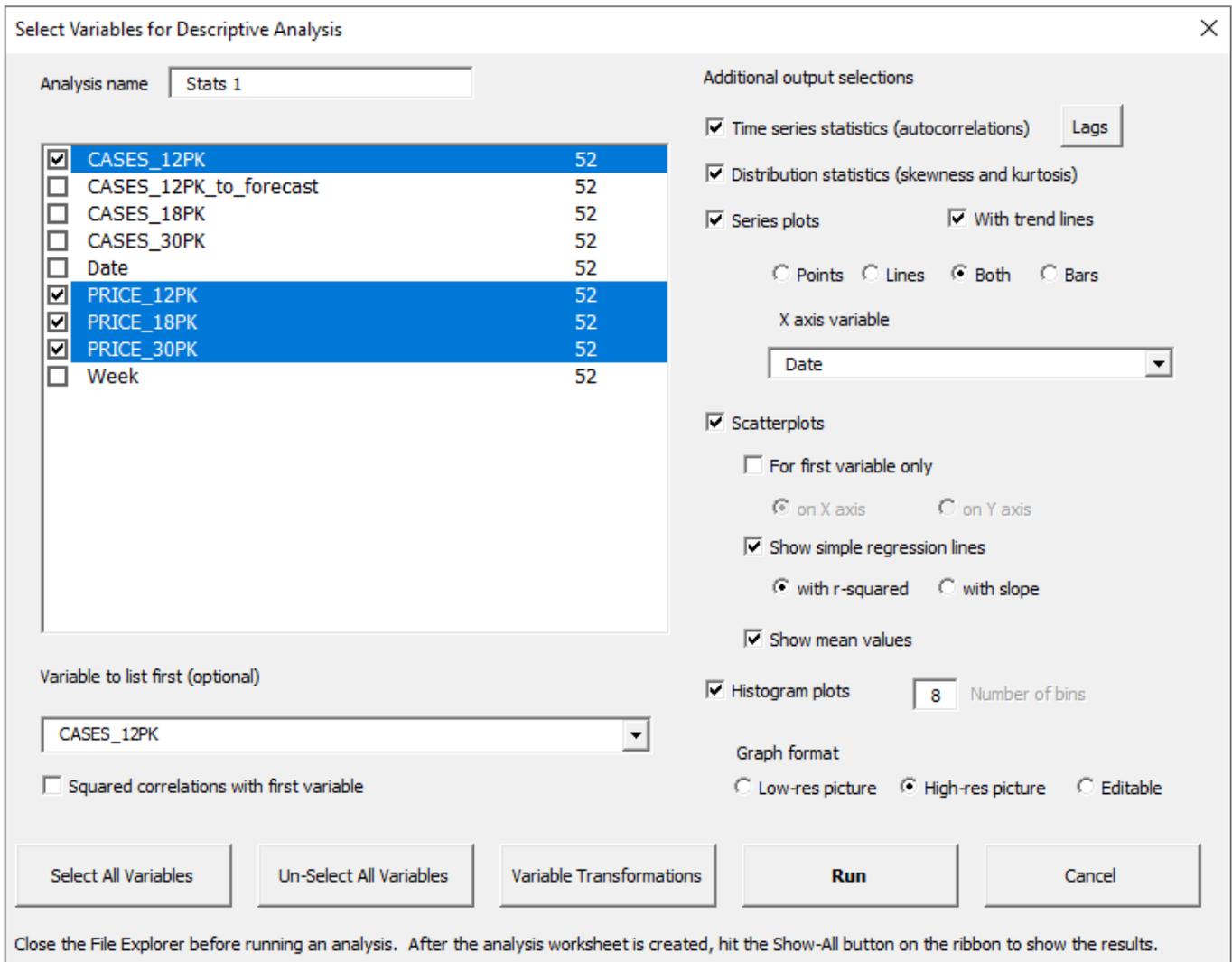
The "Variable-to-list-first" option is normally used when one of the variables is pre-identified as the dependent variable to be used in subsequent regression analysis. This will cause it to appear in the first row of the descriptive stats table and first row and column of the correlation matrix and scatterplot matrix, out of alphabetical order.
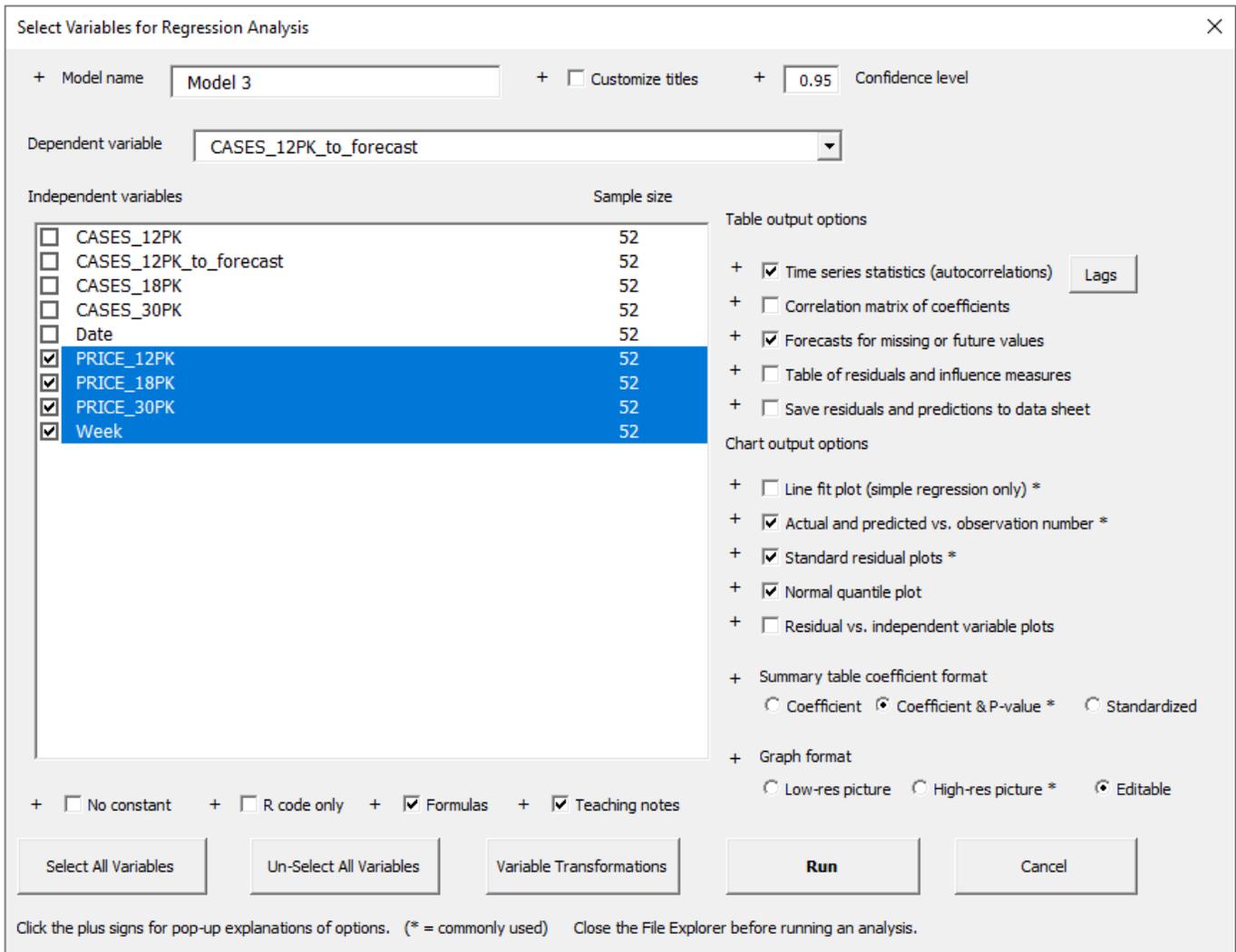
If the "Squared correlations with first variable" option is also chosen, the correlation matrix is replaced by a *table of correlations and squared correlations versus the first variable*, as a prelude to regression. This table can be sorted on the basis of squared correlations via the "Filter" tool on the ribbon, and regression variables can be preselected from it by using the "Remove" tool to deselect those having only tiny correlations with the prospective dependent variable. Clicking on a variable name in the table and hitting the Remove tool will cause that line to be grayed out, and a regression model launched from this stats sheet will have only the non-removed variables preselected. An example is shown at the bottom of page 6.

LAGS FOR AUTOCORRELATIONS: If the "Time series statistics" box is checked, an autocorrelation table will be included on the output worksheet. This option applies to both descriptive statistics (autocorrelations of variables) and regression (autocorrelations of residuals). Lags for which autocorrelations are to be computed can be consecutive or not, and they are chosen from a dialog box launched by the "Lags" button. For example, for monthly time series data, you might choose only lag 1 and lag 12. You can also generate a list of lags up to a given maximum.

2

**Enter List Of Autocorrelation Lags Separated By Spaces** ✕

| 1 2 3 4 |

| Clear List | Add All Lags Up To | 4 | Choose Listed Lags | Cancel |

LINEAR REGRESSION PROCEDURE:

The regression dialog box includes many options for presentation-ready table and chart output.

Select Variables for Regression Analysis ✕

+ Model name [Model 3]   + ☐ Customize titles   + [0.95] Confidence level

Dependent variable [CASES_12PK_to_forecast ▼]

Independent variables   Sample size

Table output options

☐ CASES_12PK	52	
☐ CASES_12PK_to_forecast	52	+ ☑ Time series statistics (autocorrelations) [Lags]
☐ CASES_18PK	52	+ ☐ Correlation matrix of coefficients
☐ CASES_30PK	52	+ ☑ Forecasts for missing or future values
☐ Date	52	+ ☐ Table of residuals and influence measures
☑ PRICE_12PK	52	+ ☐ Save residuals and predictions to data sheet
☑ PRICE_18PK	52	
☑ PRICE_30PK	52	Chart output options
☑ Week	52	

+ ☐ Line fit plot (simple regression only) *
+ ☑ Actual and predicted vs. observation number *
+ ☑ Standard residual plots *
+ ☑ Normal quantile plot
+ ☐ Residual vs. independent variable plots

+ Summary table coefficient format
   ○ Coefficient ◉ Coefficient & P-value * ○ Standardized

+ Graph format
   ○ Low-res picture ○ High-res picture * ◉ Editable

+ ☐ No constant + ☐ R code only + ☑ Formulas + ☑ Teaching notes

| Select All Variables | Un-Select All Variables | Variable Transformations | **Run** | Cancel |

Click the plus signs for pop-up explanations of options. (* = commonly used)   Close the File Explorer before running an analysis.

Standard text output (summary stats table, coefficient table, ANOVA table) is always produced, and there are options for additional tables and charts that are commonly used in regression analysis. Forecasts can be generated for rows where the dependent variable is missing and all independent variables are present. Teaching notes that explain the key output statistics can be optionally embedded in the output worksheet as cell comments. Numeric values in cells can be coded either as fixed numbers or as formulas. The "Formulas" and "Editable" options can be used to make the tables and charts interactive as well as to demonstrate the underlying math for students.
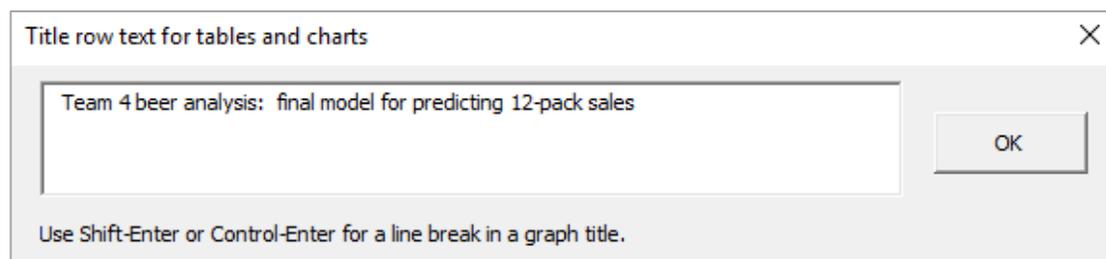
If the "R code only" box is checked, only the variable selections are read from this screen, and the R interface dialog box will be launched when the Run button is hit. See pages 14 and 17 of this document for details.

Notice that there is a plus sign to the left of each output option checkbox. If you click on a plus sign, it will open a message box with a description of that piece of output and how to use it. For example, here is the message box for the "Time series statistics" option:
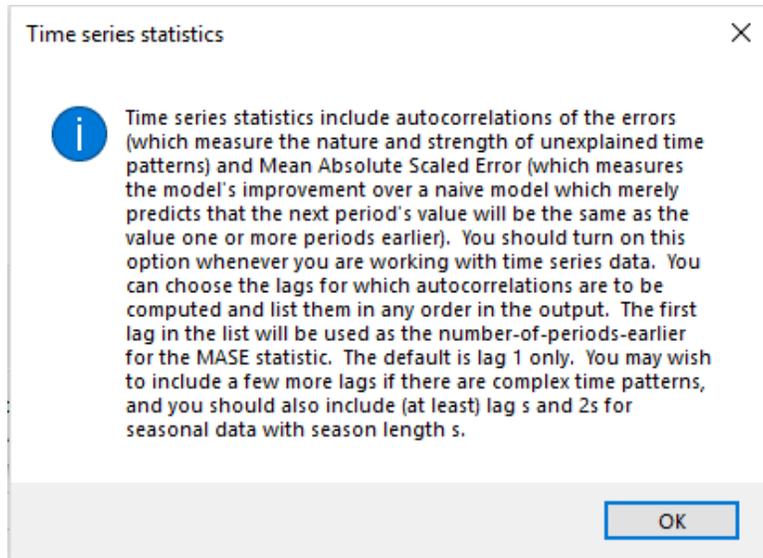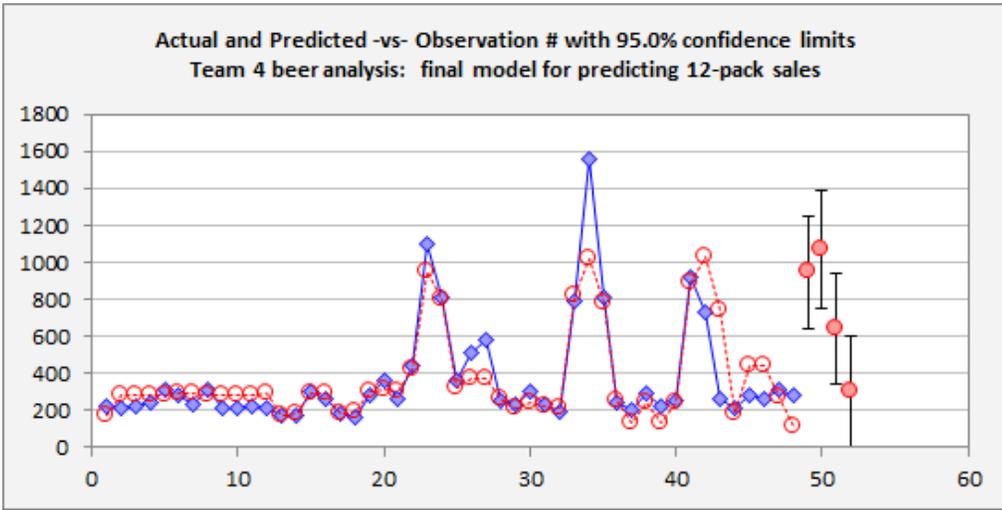

Model names can be customized in an automatically numbered sequence. The default name of the next model in a sequence is generated by adding 1 to the number that appears at the right end of the current model name if it is preceded by a space or period. Thus, if the name of the previous model was "12-pack model 3.2", the default that is generated for the next one will be "12-pack model 3.3," without any typing required on your part. However, an arbitrary new model name can always be typed in the name field in the upper left of the dialog box. The model name will also become the model's worksheet name by default.   The default name for the very first model fitted is "Model 1".

The model name shown in table and chart titles can also be customized independently of the worksheet name. **I**f the "Customize titles" box is checked, the dialog box below pops up when the "Run" button is hit. This feature is useful when preparing a final set of charts and tables for presentation or inclusion in a report, and it allows for long titles. You can go well beyond the 30-character limit that applies to worksheet names. Students can be also asked to include identifying information about themselves or their team. For example, when this custom title is entered...
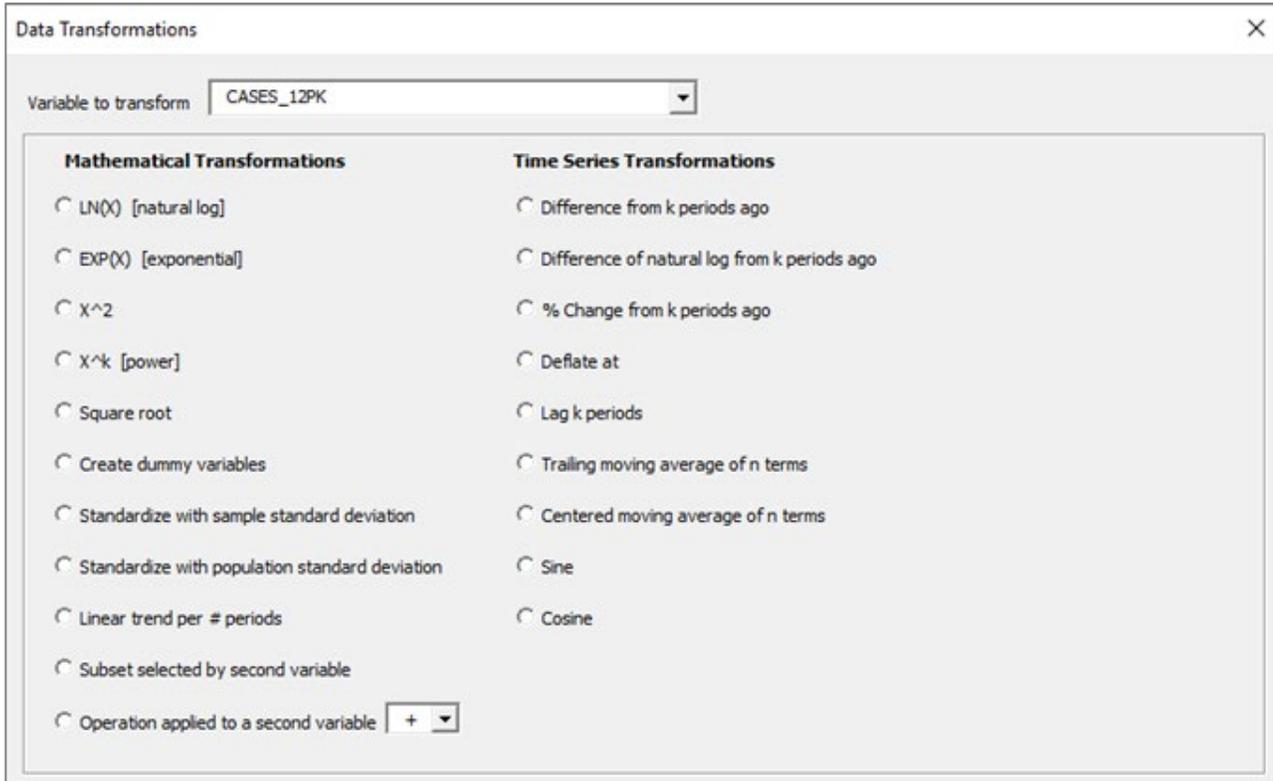

...the charts come out looking like those in the picture below.

Actual and Predicted -vs- Observation # with 95.0% confidence limits
Team 4 beer analysis: final model for predicting 12-pack sales

However, it is generally best to use the default chart titles for routine work, because they link back to the names of the worksheets in the file from which the charts were copied. It's usually enough to just be thoughtful in assigning model names, and RegressIt's default naming convention helps with this.

VARIABLE TRANSFORMATION TOOL: This can be launched from either the descriptive stats or regression dialog box by hitting its button at the bottom of the screen. It includes time transformations if the "Time Series Data" box is checked when it is launched. Transformed variables are automatically assigned descriptive names, e.g., the natural log of X is automatically named X.Ln. X lagged by 1 period is automatically named X.Lag1, and so on. The transformation name is tacked on at the end rather than the beginning so that the original and transformed variables appear adjacent to each other in alphabetically sorted variable lists. Also, the new column of data for the transformed variable is inserted next to the column of the parent variable on the data worksheet.


Data Transformations

Variable to transform   CASES_12PK

Mathematical Transformations

○ LN(X) [natural log]

○ EXP(X) [exponential]

○ X^2

○ X^k [power]

○ Square root

○ Create dummy variables

○ Standardize with sample standard deviation

○ Standardize with population standard deviation

○ Linear trend per # periods

○ Subset selected by second variable

○ Operation applied to a second variable [+ ▾]

Time Series Transformations

○ Difference from k periods ago

○ Difference of natural log from k periods ago

○ % Change from k periods ago

○ Deflate at

○ Lag k periods

○ Trailing moving average of n terms

○ Centered moving average of n terms

○ Sine

○ Cosine

Descriptive Statistics Stats 1

Variable	# Fitted	Mean	Median	Std.Dev.	Root.M.Sqr.	Std.Err.Mean	Minimum	Maximum	Skewness	Kurtosis
CASES_12PK	52	399.163	272.500	302.555	499.110	41.957	159.000	1,557	2.157	4.601
PRICE_12PK	52	19.088	19.980	2.088	19.199	0.290	14.330	21.280	-1.441	0.500
PRICE_18PK	52	16.725	18.650	2.411	16.894	0.334	13.260	19.500	-0.266	-1.901
PRICE_30PK	52	14.379	14.395	0.806	14.401	0.112	12.830	15.190	-0.334	-1.388

Autocorrelations

Variable	Lag 1	Lag 2	Lag 3	Lag 4
CASES_12PK	0.478	-0.035	-0.071	-0.087
PRICE_12PK	0.544	-0.007	-0.282	-0.254
PRICE_18PK	0.306	-0.141	0.173	0.430
PRICE_30PK	0.341	0.069	0.109	-0.089

Series Plots
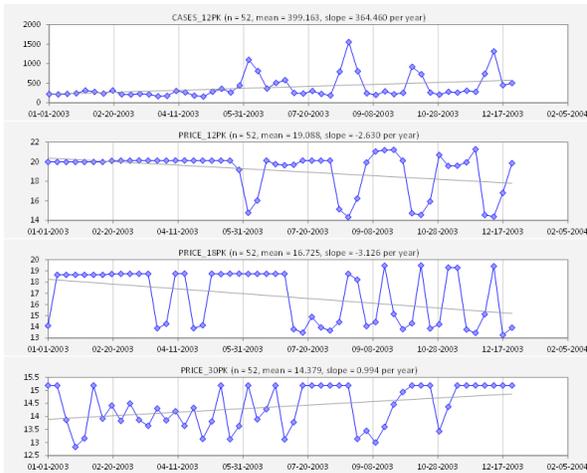
Histogram Plots
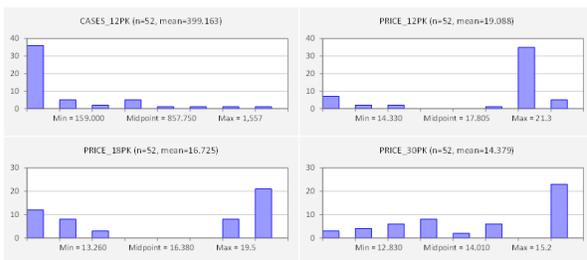
Correlation Matrix (n=52)

Variable	CASES_12PK	PRICE_12PK	PRICE_18PK	PRICE_30PK
CASES_12PK				
PRICE_12PK	-0.859	1.000		
PRICE_18PK	0.241	-0.084	1.000	
PRICE_30PK	0.300	-0.364	-0.251	1.000

Scatterplots
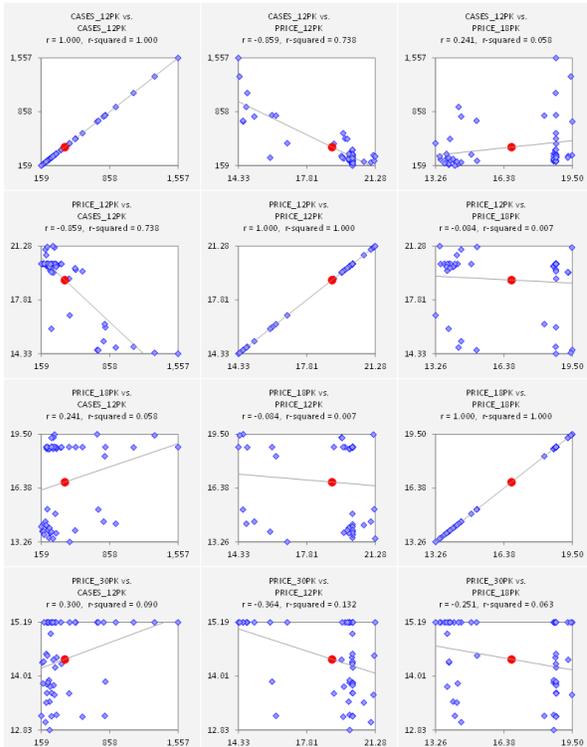
Sample output worksheet from the **DESCRIPTIVE STATISTICS** procedure, showing descriptive stats, autocorrelations, series plots, histogram plots, correlations, and a full scatterplot matrix.

Note: the tables and charts are hidden (by grouping of rows) when the worksheet is first produced. Hit the "Show all" button or click the plus signs in the left sidebar to unhide the output.

Here the series plots include optional trend lines and their slopes as well as an optional custom time axis scale.

Scatterplots include optional regression lines and center-of-mass points.

Color-coding has been turned on for the autocorrelations and correlations. This can be toggled on and off at any time if RegressIt is active by hitting the **Colors** button.

If the "Editable" option was used, all 24 charts are native Excel charts that are subject to editing, although their descriptive titles and high-quality default formatting generally make it unnecessary to edit them. There are also options for high-res and low-res picture format. See page 16 of this document for complete details.

Point sizes on graphs are automatically scaled between 4 and 8 points to suit the size of the data set.

The correlation matrix can be optionally replaced by a sortable table of correlations and *squared* correlations versus the first variable. The picture below is an example that uses the same variables as in the analysis shown to the left. The "Filter" tool can be used to sort the table on squared correlations to bring the most important variables to the top. If you click the "Remove" button while positioned in some row of the table (e.g., one that is at the bottom after sorting), that row will be grayed-out, and the variable will not be preselected in a regression model launched from this sheet. PRICE_18PK has been removed below:
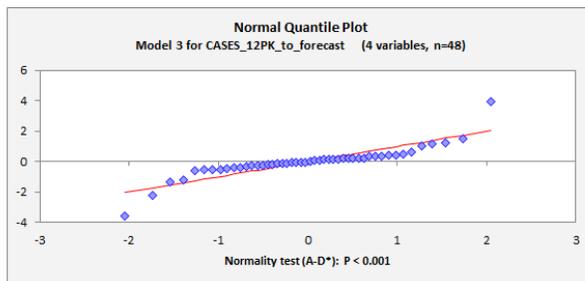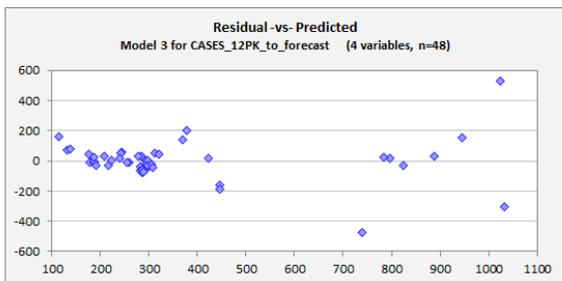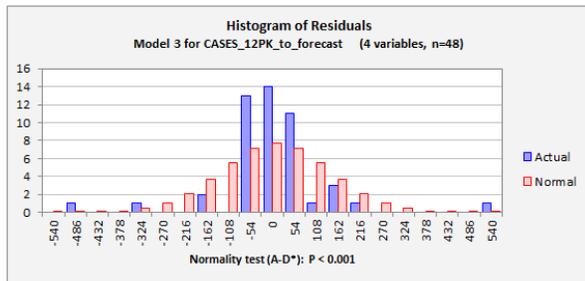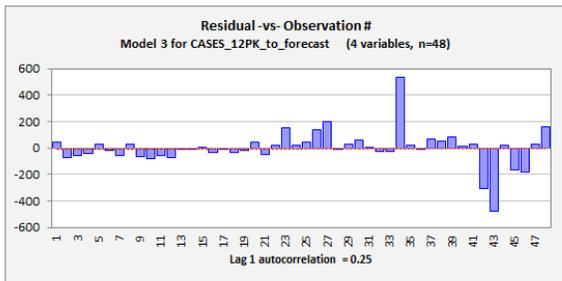
Correlations and Squared Correlations -vs- CASES_12PK (n=52)

Variable	Correlation	Squared
CASES_12PK	1.000	1.000
PRICE_12PK	-0.859	0.738
PRICE_30PK	0.300	0.090
PRICE_18PK	0.241	0.058

Model: Model 3
Dependent Variable: CASES_12PK_to_forecast
Independent Variables:
PRICE_12PK, PRICE_18PK, PRICE_30PK, Week
Equation:
Predicted CASES_12PK_to_forecast = 2,130 − 117.219*PRICE_12PK + 23.775*PRICE_18PK + 3.242*PRICE_30PK + 2.297*Week

	R-Squared	Adj.R-Sqr.	Std.Err.Reg.	Std.Dep.Var.	# Fitted	# Missing	Critical t	Confiden
	0.769	0.748	140.079	278.878	48	4	2.017	95.0%

Variable	Coefficient	Std.Err.	t-Statistic	P-value	Lower95%	Upper95%	VIF	Std. Coef
Constant	2,130	590.762	3.606	0.001	939.107	3,322		0.000
PRICE_12PK	−117.219	11.863	−9.881	0.000	−141.143	−93.296	1.231	−0.803
PRICE_18PK	23.775	9.656	2.462	0.018	4.301	43.249	1.265	0.203
PRICE_30PK	3.242	27.923	0.116	0.908	−53.070	59.553	1.203	0.009328
Week	2.297	1.658	1.386	0.173	−1.046	5.639	1.290	0.115

Source	Deg. Freedom	Sum Squares	Mean Square	F-Statistic	P-value
Regression	4	2,811,584	702,896	35.821	0.000
Residual	43	843,755	19,622		
Total	47	3,655,338			

	Mean Error	RMSE	MAE	Minimum	Maximum	MAPE	A-D* stat	MASE lag
Fitted (n=48)	−1.066E-12	132.583	77.613	−476.797	532.650	22.4%	2.95 (P=0.000)	0.468

Lag	1	2	3	4
Autocorrelation	0.245	0.028	0.213	0.109
StdErrorsFromZero	1.682	0.187	1.431	0.725
Durbin-Watson	1.475			

Obs#	Forecast	StErrFcst	Lower95%F	Upper95%F	StErrMean	Lower95%M	Upper95%M	PRICE_1
49	945.271	152.657	637.409	1,253	60.679	822.900	1,068	14
50	1,070	155.959	755.208	1,384	68.564	931.458	1,208	14
51	641.661	149.486	340.193	943.129	52.192	536.407	746.916	16
52	302.131	149.084	1.473	602.788	51.029	199.221	405.040	19

Forecasts and 95.0% confidence limits for means and forecasts
Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)

Actual and Predicted -vs- Observation # with 95.0% confidence limits
Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)

Residual -vs- Observation #
Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)
Lag 1 autocorrelation = 0.25

Histogram of Residuals
Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)
Normality test (A-D*): P < 0.001

Residual -vs- Predicted
Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)

Normal Quantile Plot
Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)
Normality test (A-D*): P < 0.001

Sample output worksheet from the **LINEAR REGRESSION** procedure, including forecasts that were automatically generated for 4 missing values of the dependent variable at the end of the series.

Color coding has been toggled on for the t-stats, standardized coefficients, and residual autocorrelations. It would also be applied to the correlation matrix of coefficients if it was included.

The coefficient table can be sorted on any statistic (such as P-value) by clicking on it and hitting the "Filter" button on the ribbon, which activates Excel's filter tool.

A variable that is insignificant can be flagged for removal from the next model launched from this sheet by clicking on any cell in its row and then hitting the **"Remove"** button, which will cause its line in the coefficient table to be grayed out. When the Run button is hit, it will not be preselected by default.

The residual table, if it had been included, would appear at the bottom and would include actual and predicted values, residuals, standardized residuals and their absolute values, leverage, and Cook's D.   The filter tool on the ribbon can be used to easily sort the residual table on any of these statistics.

Note: *this is actually a poor model for predicting sales*. A log transformation of both variables should be used for this kind of price-demand analysis, and an example is discussed on the RegressIt web site.

The list of independent variables and the model equation are stored as text strings in cells A4 and A6 as pictured below. These rows are hidden at first by default when a new model sheet is produced, but they can be displayed by hitting "Show all" or by clicking the plus sign next to them in the left sidebar.

Model: Model 3
Dependent Variable: CASES_12PK_to_forecast
Independent Variables:
PRICE_12PK, PRICE_18PK, PRICE_30PK, Week
Equation:
Predicted CASES_12PK_to_forecast = 2,130 – 117.219*PRICE_12PK + 23.775*PRICE_18PK + 3.242*PRICE_30PK + 2.297*Week

	R-Squared	Adj.R-Sqr.	Std.Err.Reg.	Std.Dep.Var.	# Fitted	# Missing	Critical t	Confidence
	0.769	0.748	140.079	278.878	48	4	2.017	95.0%

Variable	Coefficient	Std.Err.						oeff.
Constant	2,130	590.762						00
PRICE_12PK	-117.219	11.863						03
PRICE_18PK	23.775	9.656						03
PRICE_30PK	3.242	27.923						328
Week	2.297	1.658						5

> The standard error of a coefficient is the (estimated) standard deviation of the error that has been made in estimating it from the given sample of data in the context of the given model.
>
> In general it gets smaller in proportion to 1 divided by the square root of the sample size as the sample size increases. Thus, 4 times as much data should be expected to reduce the standard errors of all the coefficient estimates by a factor of 2 (approximately), assuming that the additional data is described by the same model.

	Mean Error	RMSE	MAE	Minimum	Maximum	MAPE	A-D stat	MASE lag 1
Fitted (n=48)	-1.066E-12	132.583	77.613	-476.797	532.650	22.4%	2.95 (P=0.000)	0.468

TEACHING NOTES can be optionally included as cell comments behind the key statistics in the output by checking the "Teaching notes" box in the regression dialog box at run time. The comment for standard errors of coefficients is shown above.

Red flags that mark cells with notes can be toggled on and off by hitting the "Notes" button on the ribbon. When the flags are displayed, the note in a cell will pop up when the cell is moused over.

At present there are around 6000 words of teaching notes in the linear regression version of RegressIt, and there are extensive additional notes in the logistic regression version. Altogether there are around 10,000 words of teaching notes and pop-up documentation of features.

Several other tables on a regression model worksheet are also hidden by default because they are not always of interest or are potentially huge: the analysis of variance (ANOVA) table, the correlation matrix of coefficients (if any), and the residual table (if any). Here too, their rows can be displayed all at once by hitting the "Show all" button or selectively by clicking plus signs in the left sidebar.

If you hit the "Titles" button on the ribbon, it will toggle the display of a 1-line descriptive title in the row immediately above each table and chart.   *In the worksheet pictured below, the ANOVA table has been unhidden and the title rows have been displayed.* It is useful to show the title rows when working with many models and flipping back and forth among worksheets, so that you know which model's tables you are seeing at any given moment. It is also good to include the title rows when copying and pasting tables to reports, because this will provide an audit trail leading back to the worksheet in the Excel file from which it was obtained*.

Model: Model 3
Dependent Variable: CASES_12PK_to_forecast
Independent Variables:
PRICE_12PK, PRICE_18PK, PRICE_30PK, Week
Equation:
Predicted CASES_12PK_to_forecast = 2,130 - 117.219*PRICE_12PK + 23.775*PRICE_18PK + 3.242*PRICE_30PK + 2.297*Week

Regression Statistics: Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)

	R-Squared	Adj.R-Sqr.	Std.Err.Reg.	Std.Dep.Var.	# Fitted	# Missing	Critical t	Confidence
	0.769	0.748	140.079	278.878	48	4	2.017	95.0%

Coefficient Estimates: Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)

Variable	Coefficient	Std.Err.	t-Statistic	P-value	Lower95%	Upper95%	VIF	Std. Coeff.
Constant	2,130	590.762	3.606	0.001	939.107	3,322	0.000	0.000
PRICE_12PK	-117.219	11.863	**-9.881**	0.000	-141.143	-93.296	1.231	**-0.803**
PRICE_18PK	23.775	9.656	2.462	0.018	4.301	43.249	1.265	0.203
PRICE_30PK	3.242	27.923	0.116	0.908	-53.070	59.553	1.203	0.009328
Week	2.297	1.658	1.386	0.173	-1.046	5.639	1.290	0.115

Analysis of Variance: Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)

Source	Deg. Freedom	Sum Squares	Mean Square	F-Statistic	P-value
Regression	4	2,811,584	702,896	35.821	0.000
Residual	43	843,755	19,622		
Total	47	3,655,338			

Error Distribution Statistics: Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)

	Mean Error	RMSE	MAE	Minimum	Maximum	MAPE	A-D* stat	MASE lag 1
Fitted (n=48)	-1.066E-12	132.583	77.613	-476.797	532.650	22.4%	2.95 (P=0.000)	0.468

If you toggle the "Show all" button to hide all the output, the title rows will remain displayed and will give you an outline view of the model worksheet as shown below. You can then click the plus signs in the left sidebar to open up individual tables and charts.

Model: Model 3
Dependent Variable: CASES_12PK_to_forecast

+

Regression Statistics: Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)

+

Coefficient Estimates: Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)

+

Analysis of Variance: Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)

+

Error Distribution Statistics: Model 3 for CASES_12PK_to_forecast   (4 variables, n=48)

+

MODEL SUMMARIES WORKSHEET

Linear Model For CASES_12PK_to_forecast	Model 1	Model 2	Model 3
Run Time	11/16/20 10:09 AM	11/16/20 10:10 AM	11/16/20 10:12 AM
# Fitted	48	48	48
Mean	369.688	369.688	369.688
Standard Deviation	278.878	278.878	278.878
Number Of Variables	1	3	4
Standard Error of Regression	**145.213**	**141.536**	**140.079**
R-squared	0.735	0.759	0.769
Adjusted R-squared	**0.729**	**0.742**	**0.748**
Mean Absolute Error	90.035	86.665	77.613
Mean Absolute Percentage Error	27.4%	26.6%	22.4%
Maximum VIF		1.190	1.290
Normality Test	***	***	***
Mean Absolute Scaled Error	0.542 (lag 1)	0.522 (lag 1)	0.468 (lag 1)
Residual Autocorrelation	0.24 (lag 1)	0.26 (lag 1)	0.25 (lag 1)
Coefficients:	Model 1	Model 2	Model 3
Constant	2,785 (0.000)	2,298 (0.000)	2,130 (0.001)
PRICE_12PK	**-125.087 (0.000)**	**-121.654 (0.000)**	**-117.219 (0.000)**
PRICE_18PK		18.849 (0.044)	23.775 (0.018)
PRICE_30PK		7.274 (0.797)	3.242 (0.908)
Week			2.297 (0.173)

Model 3 (#vars=4, n=48, AdjRsq=0.748)
Dependent variable = CASES_12PK_to_forecast
Run time = 11/16/2020 10:12:15 AM
File name = Beer_sales_2_models.xlsx
Data sheet name = Beer Data
Computer name = FACDS414
Program file name = RegressItPC
Version number = 2020.11.11
Execution time = 00h:00m:04s

Model = Model 3
Variable = PRICE_12PK
Coeff = -117.219
StdErr = 11.8628
t-stat = -9.881
P-value = 0
VIF = 1.231
StdCoeff = -0.8032

The "Model summaries" worksheet shows side-by-side side model stats in journal-article format. Optional color coding of coefficients can be toggled on here too, as shown above.

This sheet (which is located to the right of all analysis worksheets) is updated with every new regression model, and it provides an audit trail for all models that have been fitted in this file's lifetime, even those whose output worksheets may have been subsequently deleted. It is very easy to spin off large numbers of models with memory-intensive chart output, so the user will often want delete inferior ones in the course of a long analysis. However, it is good to maintain the entire audit trail, particularly to verify the efforts of students.

When the worksheet is first produced, the rows for the coefficient estimates are hidden (grouped). Click the plus sign in the left sidebar to open up the coefficient table.

The cell comment behind the model name includes much more audit trail information, and cell comments behind the coefficient estimates provide more detail of those, as shown above.

The user has the option at run time to store the coefficients in the table in coefficient-plus-P-value text format (which was done here), or else in numeric form, either standardized or unstandardized. Regardless of which option was chosen for display, the cell comments provide all the other information about each coefficient estimate.

Another feature is that any regression model whose summary output is stored here can be easily re-run. The model specifications are stored in the cell right above the model name and hidden in white font. If the regression button is hit while positioned on this cell, the model information will be read from it, and that model's worksheet will be moved to if it is present. If it is not present, the variable specifications will become the default selections for a new model, so the original one can be regenerated with a couple of clicks. The entire history of models in the workbook can be re-created from this single sheet in this way.

If models had also been fitted to some other dependent variable, their results would be stored in a different block of rows farther down. Thus, only models fitted to the same dependent variable are displayed side-by-side.

Some of the most unique features of RegressIt, which take full advantage of the interactivity and 4-dimensional structure of an Excel workbook, are the navigation and display tools on the ribbon menu. These tools are designed to support best practices of regression analysis and to facilitate exploration of the output and presentation of results to others. Virtually every aspect of a thoughtful analysis can be controlled by clicking the buttons here.

MODEL SPACE NAVIGATION
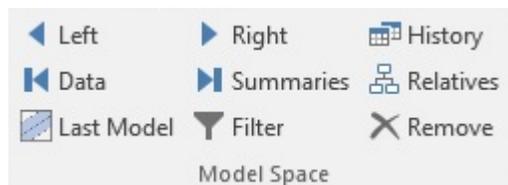
VCR buttons: move one worksheet to the right or left, or jump to data sheet (leftmost sheet) or model summary sheet (rightmost sheet)

Last model: toggle back and forth to the last-visited model worksheet from wherever you are, which is good for head-to-head model comparisons or for jumping back to a model worksheet after visiting a stats worksheet. *Note: this only works if you use RegressIt's navigation tools to move around in the workbook, not if you jump to other worksheets by clicking on their tabs at the bottom of the screen.*

History: This very powerful feature provides a detailed list of all worksheets in the file and allows you to jump directly to any of them: click on the line for the desired sheet and hit the Move-To-Sheet button. The history box stays open on the screen until "Cancel" is hit, so you can visit a sequence of worksheets without having to re-launch it. It can be kept on top of the worksheet while you work. The last sheet visited is remembered, so that you can easily toggle back and forth between any two sheets.
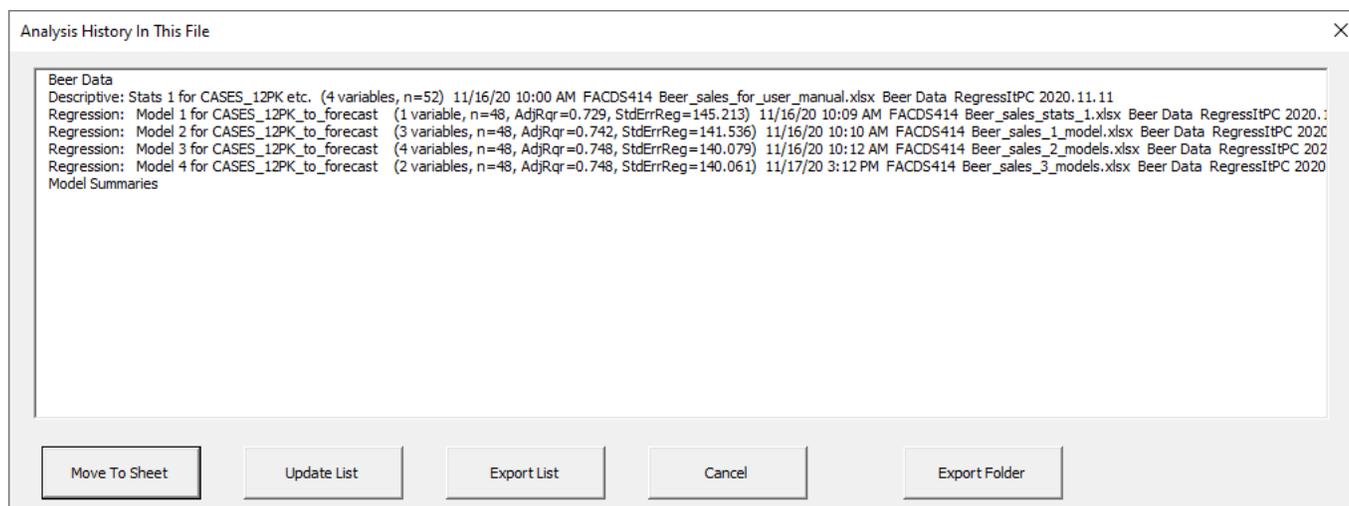
Key statistics of the models are shown here. Audit trail information for each analysis, such as run time and computer name, is also stored here and in several other places in the workbook, both in text strings in cells and in cell comments. This is useful for managing your own individual work or teamwork by keeping track of the temporal sequence as well as which of your many computers you may have used to store related files. It also serves to authenticate original work performed by students on homework assignments or laboratory exercises.

The "Export List" and "Export Folder" buttons allow the contents of the analysis history for a single file or many files to be exported to an Excel worksheet for more detailed analysis and auditing and for purposes of grading student files. See the **RegressIt Features for Instructors** document for more details.

Relatives: Allows you to jump to the "parent" or the "children" of the regression model whose sheet you are currently looking at, in order to easily retrace your steps or back up to an earlier model in order to go down another search path. (Each regression model worksheet contains information about its parents and children in cells R2, S2, etc.) In this example, models 3 and 4 were both launched from the model 2 worksheet:
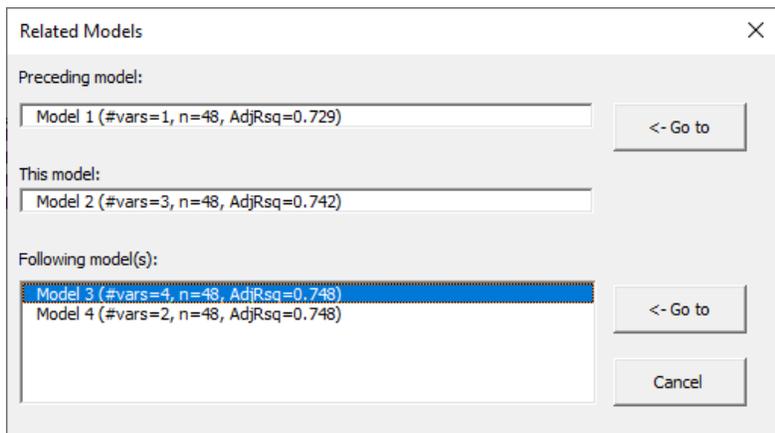
Filter: this executes the Excel filter tool. It can be used to filter and sort the coefficient table (say, on P-values) or to sort a table of squared correlations with the dependent variable (to preselect independent variables) or to sort the residual table on error size or other stats there.

Remove: clicking this button while selecting some range of cells in the coefficient table for an existing regression model will cause those rows to be grayed-out, which deselects them from the next regression model launched from that sheet. This is intended to be used in conjunction with the filter tool: *sort the coefficient table on P-values and deselect a bunch of insignificant variables all at once.* It can also be used on a descriptive statistics output sheet in which the "squared correlations with first variable" option was chosen, as illustrated on page 6. If you launch a regression model from that sheet, the predictor variables in the table will be selected by default unless some have been deselected with this tool. This is convenient for throwing out variables with insignificant correlations with the dependent variable when going from descriptive analysis to regression analysis, particularly when the set of candidate variables is very large.

WORKSHEET NAVIGATION
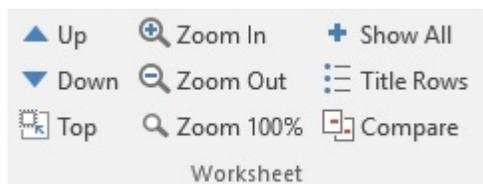
Up/down: jump up and down by whole table or whole chart, so that the uppermost one is always correctly aligned with the top left corner of the screen. This is important for efficient viewing because it doesn't require fine adjustment of the viewing area with the scroll bar. It also minimizes re-drawing speed, which is important for hi-res or editable charts. This feature is also very useful in live presentations.

Zoom in/out: zooms all worksheets at once in a series of predefined levels for a closer or more distant look.

Top: jump to top left of current worksheet

Show all: toggle the display of all charts and tables at once by grouping or ungrouping of their rows. You will usually hit this button right after running a descriptive analysis, because all the output is hidden by default. You may also want to toggle this button to *hide* all the output while at the same time using the "Title rows" button to show the title rows above the tables and charts. An example is shown on page 9.

Title rows: toggle title rows on and off. Above each table or chart is a title row that includes the same model details (name, dep. variable, sample size) that are in chart titles, as shown on page 9. For audit trail purposes it is useful to include this extra line in the copied area when tables are individually copied to other documents.

Compare: Hitting this button while positioned on the title row of any table or chart will reposition it at the top of the viewing area *and do the same for all other model sheets*, so that the same pieces of output will appear in the same relative positions on the screen when moving back and forth among model sheets to compare their results. *If you hit "Top" followed by "Compare", you will re-set the entire workbook to its original viewing state, with cell A1 in the top left corner of every window. You can also hit the "Zoom" button to choose a different zoom level for the whole workbook to suit a new computer (e.g., for a presentation).*

CELL DISPLAY
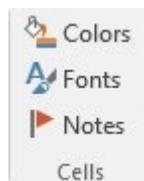
Colors: toggle fill colors on and off. This tool affects correlations and autocorrelations on the descriptive stats sheet, t-stats and standardized coefficients and correlations of coefficients on the model sheet, and coefficients in the model summary worksheet. There is a predefined scale of fill colors ranging from dark blue (very significantly positive) to dark red (very significantly negative).

Fonts: toggle font coding on and off. This tool affects the same cells as the coloring tool, but changes the color (shade of gray) and boldness of font rather than fill color as an indicator of significance (better for printing).

Notes: toggles red flags for cell comments on and off. When the flags are on, the comment will pop up next to the cell when you hover the mouse over it. There is always a cell comment behind the model name on the model sheet and behind the run time cell on the model summary sheet, containing more facts about the model (e.g., run time and name of user and file at run time, for audit trail purposes). The coefficient cells on the model summary worksheet also have comments with more details of each coefficient (t-stats, standardized values, etc.). On the model sheet, if the "teaching notes" box was checked at run time, there are teaching notes behind the column titles in the various statistics tables and the titles above the charts, i.e., there is a teaching note for virtually every statistic and chart on the model sheet. *Customization of the teaching notes to fit an instructor's own materials is something that could be arranged.*

UTILITIES
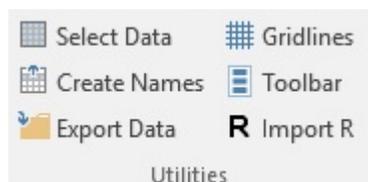
Select data: if you hit this button while positioned anywhere on the data worksheet, it will feel out and select the range containing the data (provided that no additional rows and columns have previously been occupied).

Create names: executes the create-name-from-selection command to assign variable names (= range names). Hit this immediately after hitting the "Select data" button when setting up a new file.

Export data: exports the contents of the data worksheet to a CSV file and writes R code to the clipboard for loading it into RStudio. There is an option to customize the file name and location when doing this. The default file name is the data worksheet name and default file location is the current file location. After hitting this button, go to the command prompt in RStudio and hit <Paste><Enter> to run the code, which loads the data into a frame with the given name.

Import R: A unique feature of RegressIt is that it can generate R code for running very sophisticated analyses in RStudio, and this code can also send output back to Excel where it is placed on a worksheet with exactly the same design as RegressIt model worksheets. Thus, R can be used as computational back end for work carried out in Excel that goes beyond RegressIt's own capabilities, and Excel can be used as a front end for running models in RStudio. More details are provided on the last page of this handout and in other documents on the web site. *After a model has been run in RStudio with code generated by RegressIt, using the option to export results back to Excel, hit this button to carry out the step of creating the new model worksheet.* It will show model statistics formatted in exactly the same way as RegressIt's own model worksheets, and its summary information will also be written to the model summaries worksheet in exactly the same way. The interactive features such as color-coding of coefficient rows will also work.
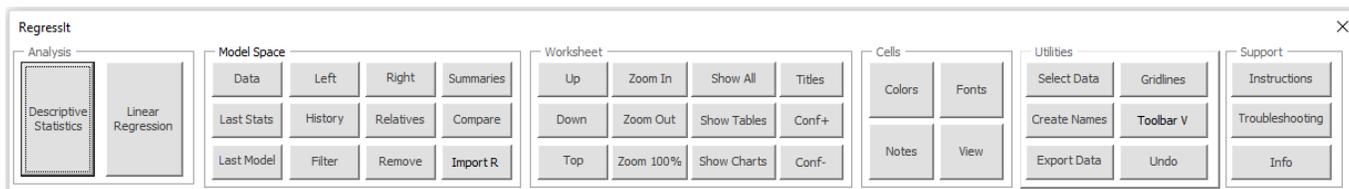
Gridlines: toggles gridlines and column headings on and off. This can be used for manually re-sizing or hiding rows and columns. (Normally this is not needed: the output is designed for easy reading by default.)

Toolbar: launches the vertical floating toolbar interface (below), which has the same buttons as the ribbon but can be positioned anywhere on the computer screen, something else that can be useful for presentations. It is better than the ribbon for use on a touchscreen or tablet computer. You can put it at the right of the screen as shown below so that your hand won't block your view of the spreadsheet, then hide the ribbon to maximize vertical space in the Excel window. The buttons are also a bit bigger than those on the ribbon, better for clumsy fingers, and there are two additional buttons,"Conf+" and "Conf-" which can be used to interactively adjust confidence levels on worksheets with formulas and editable charts.

There is also a horizontal version of the floating toolbar which is launched from the "Toolbar H" button on the vertical one and can be placed below the Excel window if that location is preferred.

RegressIt																×
Analysis		**Model Space**				**Worksheet**				**Cells**		**Utilities**		**Support**	
		Data	Left	Right	Summaries	Up	Zoom In	Show All	Titles	Colors	Fonts	Select Data	Gridlines	Instructions	
Descriptive Statistics	Linear Regression	Last Stats	History	Relatives	Compare	Down	Zoom Out	Show Tables	Conf+			Create Names	Toolbar V	Troubleshooting	
		Last Model	Filter	Remove	Import R	Top	Zoom 100%	Show Charts	Conf-	Notes	View	Export Data	Undo	Info	

SUPPORT: These buttons at the right of the ribbon display message boxes with general comments on how to use the program, as well as program version information. You should check the website (**regressit.com**) periodically for updates, particularly if you have encountered technical issues.

? Instructions
⚒ Troubleshooting
ⓘ Info
Support
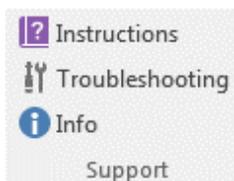
Instructions:

Start your project in a new workbook with clean data on the first worksheet arranged columnwise with variable names up to 30 characters in the first row. (Only use letters, numbers, spaces, periods or underscores, NO currency or math symbols.) Hit the Select Data and Create Names buttons to assign the names. Each time you run the descriptive statistics or regression procedure the results will be produced on a new worksheet. Regression statistics and coefficient estimates will also be written to the Model Summaries worksheet for side-by-side comparison.

CLOSE THE FILE EXPLORER when running analyses. Do not edit the model name cell or the title rows above the tables and charts nor insert or delete or hide rows or columns. Also do not hide worksheets or rename them. If you need to modify a RegressIt worksheet, first make a copy of it, move the copy to the right of all RegressIt worksheets, and edit there. Save your work frequently under new file names. Enjoy!

Troubleshooting:

See regressit.com for the latest version and complete tech support information. Do not rename, hide, or modify output worksheets. (Deletion is OK.) Non-English language, number, and date formats may cause problems. Numbers must use a period as the decimal separator and a comma as the grouping separator. Euro Currency Tools must not be used, and RegressIt will turn it off. CLOSE THE FILE EXPLORER while working, as it may cause crashes, particularly when producing non-editable graphs. This is the most common cause of crashes. Other apps could also cause conflicts.

Save files frequently under new names as you work. If an error occurs and/or an unfinished output worksheet is produced and/or a worksheet called _running_ is visible at the far left, you must delete those sheets before continuing. Also go to Formulas/Name Manager and delete any variables whose names start with_adj_. It may be better to back up to an earlier working file or clean copy of the original data in a new Excel session.

CHART FORMAT OPTIONS

There are 3 chart format options on a PC: low-res pictures, high-res pictures, and editable. On a Mac only the latter two are available. The choice of which one to use generally depends on sample size and presentation quality that is needed and whether or not it is desired for the charts to be interactive. On a PC it is possible to have a very large number of separate charts on a single output worksheet (many hundreds—no kidding), although drawing speed and file size are limiting factors, particularly with large sample sizes.

Editable charts and high-res pictures require less storage space than low-res pictures for small sample sizes, and vice versa for big ones. Editable charts are also faster to produce than either kind of picture, because the operation of converting an editable chart to a picture requires some time. (Also—remember!—creation of picture files may cause a crash on a PC if the File Explorer is open.) On both machines, editable charts are more space-efficient than high-res pictures for all sample size, but they must be treated with care and their elements do not scale up proportionally when they are stretched.

Caution: if you copy and paste an editable chart to a Word or Powerpoint document, you should use one of the paste-special-picture formats to avoid broken links. If you want to edit the chart, do it in Excel before copying. Make a second copy of the chart on its original worksheet, edit the copy, and then copy and paste it from there.

The following table illustrates how the increase in file size (in kB) due to the addition of analysis worksheets depends on the sample size and type of chart output:

Additional kB of storage for a simple regression worksheet with 6 charts						
n	PC low-res	PC high-res	PC editable		Mac picture	Mac editable
100	120	88	39		294	77
500	132	197	103		439	143
1000	161	331	189		615	233
10000	200	2254	1620		3229	1715
Additional kB of storage for a 5x5 scatterplot matrix						
n	PC low-res	PC high-res	PC editable		Mac picture	Mac editable
100	250	190	116		423	125
1000	368	1221	769		1972	795
10000	516	11612	7021		17559	7212

Some stylized facts are that (i) low-res charts on a PC use very little space in general and it doesn't increase much with sample size, (ii) editable charts take up similar amounts of memory on PCs and Macs , (iii) editable charts on a PC are around 60% the size of high-res charts, and (iv) editable charts on a Mac are around 40% of the size of pictures, so picture-format charts on a Mac require considerably more space than high-res charts on a PC.

Bottom line: if your sample sizes are small (less than a few hundred), you should use high-res format on a PC or picture format on a Mac for routine work. For large samples on a PC, use low-res format. For large samples on a Mac, editable charts use less than half as much space as pictures and they are produced more quickly because there is not an extra step for converting editable to picture format. For logistic models you should stick with editable format (the default) because most logistic regression charts are interactive. When copying and pasting editable charts to Word and Powerpoint documents, use paste-special-picture or bitmap format to preserve chart appearance and avoid broken links.

R CODE INTERFACE

A very novel and powerful feature of RegressIt is that it includes a menu interface for generating code to run models in RStudio with selections of variables and analysis options that have been chosen in Excel. The dialog box is pictured below. It pops up after you hit "Run" in the regression box with the "R code only" option checked. With just a few keystrokes the code can be run in RStudio, where it produces nicely formatted output, including some customized tables and charts very similar to those which RegressIt produces in Excel. It also offers many additional modeling options such as automatic stepwise variable selection and several different kinds of out-of-sample testing.

If the "Export model summary for Excel" box is checked here, the R code will also send output back to Excel via the clipboard. Return to Excel, and hit the "Import R" button on the ribbon. The output on the clipboard will be placed on a new model worksheet which looks exactly like those produced when running the analysis in Excel, including features such as color-coding of coefficients and creation of an entry on the model summaries worksheet. The entire sequence of events—selecting variables and choosing model options from menus in Excel, running the code in RStudio, and bringing the results back to Excel—requires only some mouse clicks and does not require the user to know the R programming language, although curious Excel users will get a chance to see R in action.

One other important advantage of the R interface is that it allows very large models to run, with data sets that are bigger than those which will fit in Excel (>1M rows) or which would take a long time to run there. The dataset does not actually have to reside in Excel in order to use R as a computational engine for fitting the models. All that is needed in Excel is to have variable names that match the ones in the data frame used in RStudio.

Details of the R interface can be found in the **RegressIt-slide-show** document and the **RegressIt-interface-with-R** document on the web site. The same information that is in the latter document also appears in html form in the **R users pages** on the web site.