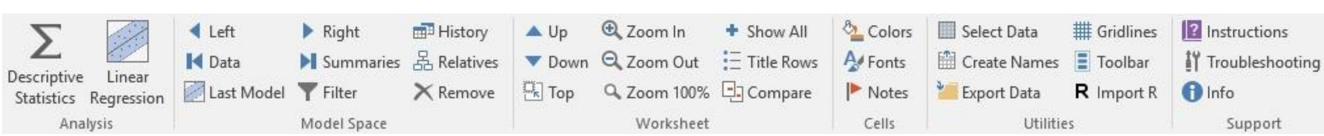
RegressIt Features for instructors[1]

RegressIt has been designed to fill a need for a free and largely self-instructive add-in that makes optimal use of the Excel environment and can be used to teach regression analysis at both elementary and advanced levels. It has many novel features that are designed to serve the needs of instructors and their students, which are described below. It supports and helps to teach best practices of data analysis (data visualization, model selection and comparison, testing of assumptions, effective presentation of results), analyses are easy to navigate and evaluate and modify, and it is even possible for an instructor to automatically generate a single worksheet with a gradeable summary of the entire contents of hundreds of student files. Most students find it much more fun to use than the usual suspects. RegressIt is also intended to be a useful addition to a practitioner's toolkit. If one of the full-featured commercial add-ins is also in use, RegressIt will probably peacefully coexist with it and effectively add another set of tools to its menu.

RegressIt is an excellent tool for online teaching with platforms such as Zoom and for interactive exercises and production of videos. The students can follow a very detailed analysis in real time by just clicking the mouse or tapping a touchscreen. Recorded videos and Excel files containing many models look good and are easily navigated on cell phones and tablets as well as desktop computers. It also has a number of features to support systematic grading and auditing of student files submitted through platforms such as Canvas and Sakai. These are described in more detail on pages 6-9 of this document.

RegressIt also provides **a bridge between Excel and R** for linear and logistic regression, opening a path for Excel users to exploit R and for R users to exploit Excel, thus making it relevant in an R-based curriculum. The R interface is described in a separate document and at this link on the web site: https://regressit.com/r-interface.html. It includes options such as stepwise regression and a number of different forms of out-of-sample testing that are not available within Excel, and it produces customized tables and charts in both RStudio and Excel that are more detailed than those of the commonly used regression packages. This tool can be used by students without any experience in R programming. It will give them a good look at R code and the RStudio environment without requiring them to write any code of their own. For those who *are* R users, RegressIt effectively provides a new package for linear and logistic regression.



The **menu interface**, whose top level is shown above**,** allows a very detailed and thoughtful analysis to be performed by using a mouse or a touchscreen. No typing is needed except for optional customization of model names and chart titles. (These have unique names anyway by default.) The navigation buttons on the Excel ribbon allow the model space to be traversed along multiple dimensions (more about this below) and they also allow the output to be interactively manipulated in many ways: browsing tables and charts, comparing them between models, zooming all worksheets at once, selectively showing or hiding outputs, highlighting numbers with shades of color, and displaying cell comments with teaching notes. These interactive features of RegressIt are very good for demonstrations and exercises in the classroom and in video sessions. A menu interface for **variable transformations**—a vital but often underemphasized tool of regression—is also provided.

----

[1] Sept. 2, 2020, Robert Nau. Visit **regressit.com** for the free software and complete documentation and sample data

There are **three layers of internal documentation** embedded in the interface, containing around 10,000 words:

- The buttons on the ribbon (top menu) have pop-up messages that explain their functions.
- The dialog boxes for regression and R code have detailed instructions for each analysis option and output option, which pop up when the user clicks the plus signs adjacent to check-boxes or entry fields.
- Very detailed **teaching notes** can be embedded in regression output worksheets by checking the teaching notes option when running the model. The notes are stored in the worksheet in the form of cell comments, which often consist of several paragraphs of text. They are activated by clicking the Notes button on the ribbon, which turns on red flags in cells that contain comments. A comment pops up when the mouse is moved over a cell with a visible red flag.

Examples are shown in the RegressIt user manual: https://regressit.com/RegressIt2020UserManual.pdf. This internal documentation allows users to educate themselves about program features and reinforce their knowledge of regression as they go along. Most will have very little need to consult external documentation or to bother their instructor because they are stuck.

The teaching note for t-stats is shown below. If you would prefer to substitute your own notes for a class, this could be arranged.

Model: Model 1
Dependent Variable: GallonsPer100Miles

Regression Statistics: Model 1 for GallonsPer100Miles (8 variables, n=392)

	R-Squared	Adj.R-Sqr.	Std.Err.Reg.	Std. Dev.	# Fitted	# Missing	Critical t	Conf. level
	0.887	0.884	0.566	1.664	392	0	1.966	95.0%

Coefficient Estimates: Model 1 for GallonsPer100Miles (8 va

Variable	Coefficient	Std.Err.	t-statistic
Constant	9.392	0.800	11.739
Cylinders	0.150	0.055	2.734
Displacement100ci	-0.323	0.131	-2.467
Horsepower100	1.275	0.235	5.437
Origin.Eq.2	-0.296	0.097	-3.053
Origin.Eq.3	-0.203	0.095	-2.142
Seconds0to60	0.034	0.017	2.051
Weight1000lb	1.125	0.112	10.035
Year	-0.131	0.008858	-14.750

Analysis of Variance: Model 1 for GallonsPer100Miles (8 vari

Source	Deg.Freedom	Sum Squares	Mean Square
Regression	8	960.035	120.004
Residual	383	122.531	0.320
Total	391	1,083	

Error Distribution Statistics: Model 1 for GallonsPer100Miles

	Mean Error	RMSE	MAE
Fitted (n=392)	0.000	0.559	0.417

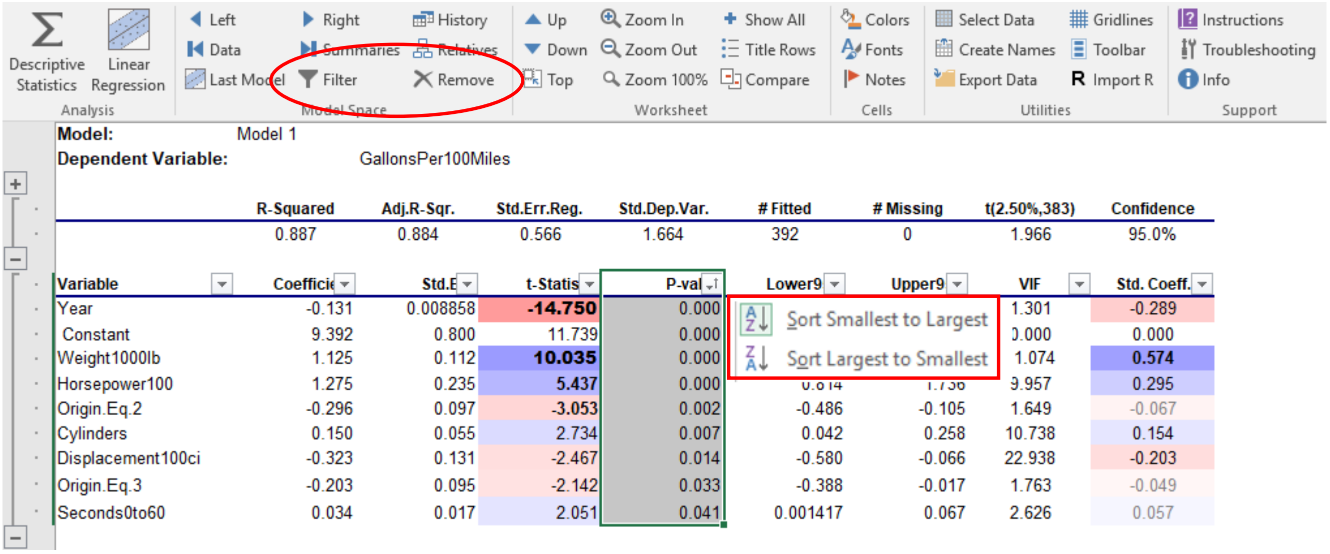
The t-statistic of an independent variable is its estimated coefficient divided by the coefficient's own standard error, i.e., its number of standard errors away from zero. The t-stat's value (which has the same sign as the coefficient) is an indicator of whether that variable has been found to have a measurably non-zero effect in explaining or predicting variations in the dependent variable, in the context of other variables included in the same model.

The t-statistic associated with any one variable is model-dependent. Its value may change, sometimes significantly, if other related variables are added or removed. Also, the t-stat of a variable whose true coefficient is non-zero tends to grow larger in magnitude as the sample size increases, because standard errors of coefficients grow smaller as the sample size increases.

A commonly used rule of thumb is that a variable's contribution to a model is not statistically significant if its t-stat is less than 2 in absolute value, i.e., if its estimated value is less than 2 standard errors away from zero, which is the approximate standard for significance at the 0.05 level. This is not a hard-and-fast rule, but as a practical matter the removal of a variable whose t-stat is much less than 2 in magnitude will probably not increase the standard error of the regression by very much.

Whether a variable should be removed also depends on other considerations, such as the objectives of the analysis and whether there are other supporting arguments for its use in the presence of the other variables. If the model specification is a priori unknown and the data have been collected in an ad hoc fashion, simpler is generally better.

A novel model-selection feature of the program that is enabled by Excel is the ability to **sort the variables in the coefficient table on P-values** (or any other statistic) by using the **Filter** tool on the ribbon. Variables that are insignificant or otherwise problematic can be interactively de-selected from the table by using the **Remove** tool (which grays them out), and the simplified model can be immediately launched from there in a couple of clicks. Zapping of insignificant variables directly from the coefficient table is another nice demonstration tool by itself.



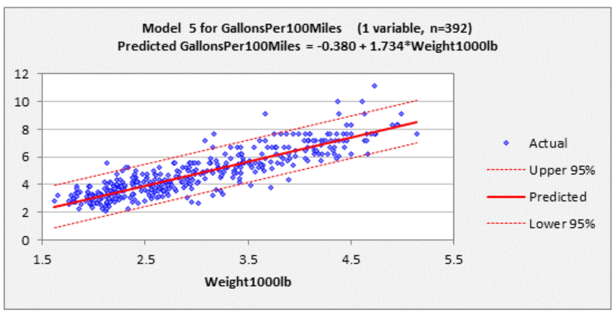
Model: Model 1
Dependent Variable: GallonsPer100Miles

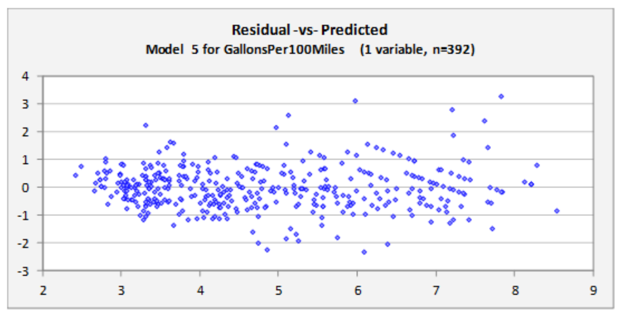
R-Squared	Adj.R-Sqr.	Std.Err.Reg.	Std.Dep.Var.	# Fitted	# Missing	t(2.50%,383)	Confidence
0.887	0.884	0.566	1.664	392	0	1.966	95.0%

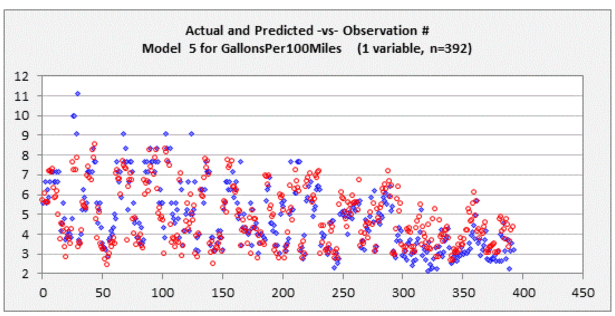
Variable	Coefficie	Std.E	t-Statis	P-val	Lower9	Upper9	VIF	Std. Coeff.
Year	-0.131	0.008858	**-14.750**	0.000			1.301	-0.289
Constant	9.392	0.800	11.739	0.000			0.000	0.000
Weight1000lb	1.125	0.112	**10.035**	0.000			1.074	**0.574**
Horsepower100	1.275	0.235	5.437	0.000	0.814	1.736	9.957	0.295
Origin.Eq.2	-0.296	0.097	-3.053	0.002	-0.486	-0.105	1.649	-0.067
Cylinders	0.150	0.055	2.734	0.007	0.042	0.258	10.738	0.154
Displacement100ci	-0.323	0.131	-2.467	0.014	-0.580	-0.066	22.938	-0.203
Origin.Eq.3	-0.203	0.095	-2.142	0.033	-0.388	-0.017	1.763	-0.049
Seconds0to60	0.034	0.017	2.051	0.041	0.001417	0.067	2.626	0.057

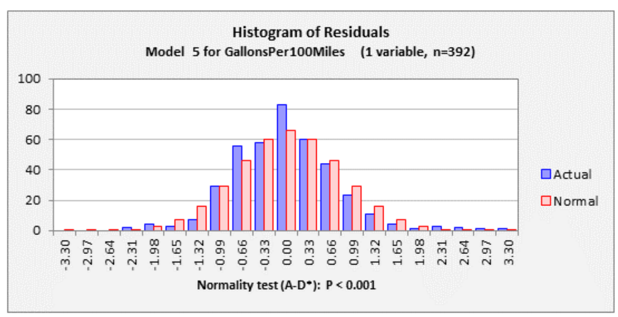
Residual tables can be similarly sorted in a few seconds to highlight the largest or most influential ones.

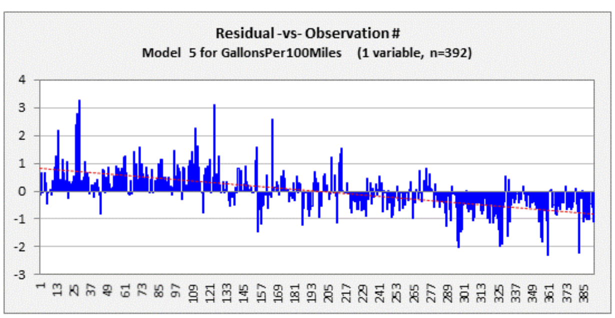
The linear regression procedure produces an array of presentation-ready charts, each of which has a descriptive title that includes the model name, sample size, and number of variables. Confidence bands on the line fit plot for a simple regression model can be interactively adjusted via the Conf+ and Conf- buttons on the ribbon.

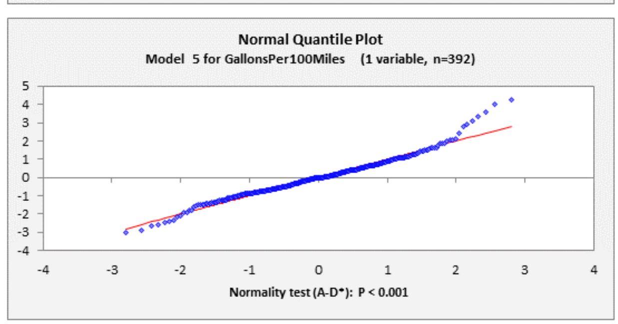












The descriptive statistics procedure can generate a large amount of well-formatted table and chart output on a single worksheet. It includes the option to produce either a full correlation matrix or or a **table of correlations and squared correlations versus a selected variable** (e.g., the dependent variable to be used). The table can be sorted on squared correlations by using the Filter tool, insignificant variables can be de-selected using the Remove tool, and a regression model can be launched directly from there without the need to select variables in the regression dialog box. This tool is useful for prescreening independent variables in situations with large numbers of candidates and is also good for live demonstrations in the classroom.

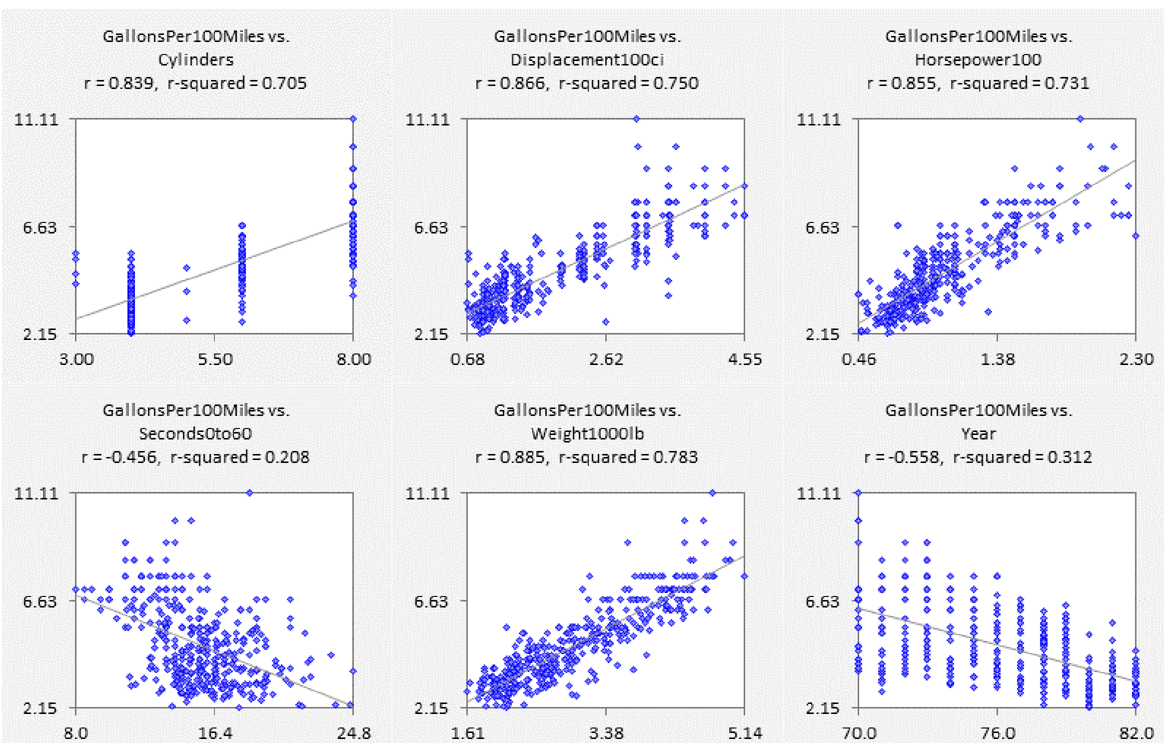
Descriptive Statistics Stats 1

Variable	# Fitted	Mean	Median	Std.Dev.	Root.M.Sqr.	Std.Err.Mean	Minimum	Maximum	Skewness	Kurtosis
GallonsPer100Miles	392	4.782	4.396	1.664	5.063	0.084	2.146	11.111	0.758	0.083
Cylinders	392	5.472	4.000	1.706	5.731	0.086	3.000	8.000	0.508	-1.398
Displacement100ci	392	1.944	1.510	1.046	2.207	0.053	0.680	4.550	0.702	-0.778
Horsepower100	392	1.045	0.935	0.385	1.113	0.019	0.460	2.300	1.087	0.697
Seconds0to60	392	15.541	15.500	2.759	15.784	0.139	8.000	24.800	0.292	0.444
Weight1000lb	392	2.978	2.804	0.849	3.096	0.043	1.613	5.140	0.520	-0.809
Year	392	75.980	76.000	3.684	76.069	0.186	70.000	82.000	0.020	-1.167

Correlations and Squared Correlations -vs- GallonsPer100Miles (n=392)

Variable	Correlat	Squa
GallonsPer100Miles	1.000	1.000
Weight1000lb	0.885	0.783
Displacement100ci	0.866	0.750
Horsepower100	0.855	0.731
Cylinders	0.839	0.705
Year	-0.558	0.312
Seconds0to60	-0.456	0.208

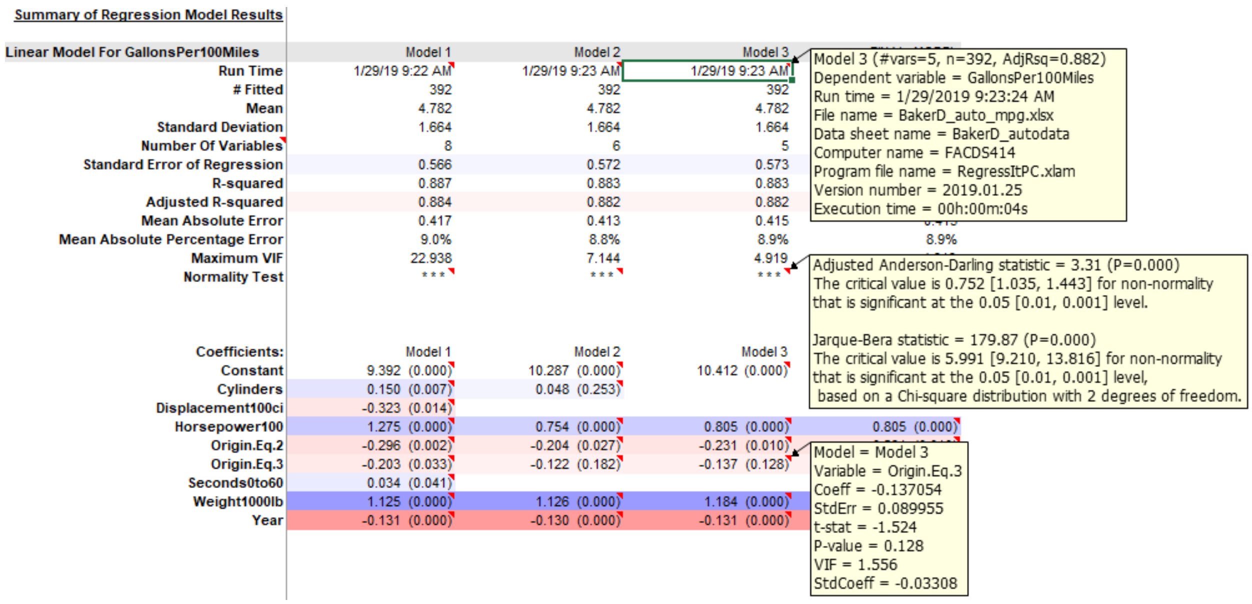
In the scatterplot matrix, all of the plots have informative axis scales that show the minimum, maximum, and midpoints of the ranges, and their titles include the variable names and correlations and either squared correlations or slope coefficients. Regression lines and center of mass points can be optionally included. The set of plots can be restricted to those involving the first variable only, either on the X or Y axis as shown here:



Series plots and histogram plots of all variables can also be produced on the same worksheet.

The program keeps a very detailed **audit trail** that can be viewed along multiple dimensions to help the user to do logically sequenced work as well as to help instructors and others who might want to browse the contents of files and evaluate the quality and originality of the analyses they contain.

The **Model Summaries** worksheet provides a side-by-side comparison of all key model statistics (error measures, diagnostics, and coefficient estimates), formatted as in a table for a journal article. This is a good tool for reviewing the entire modeling process that was followed, and it makes life easier for everyone in preparing writeups and presentations of analyses. Additional detail about some of the statistics and the model coefficients is contained in cell notes that can be activated via the Notes button on the ribbon. Examples of the cell notes behind some of the outputs are shown below



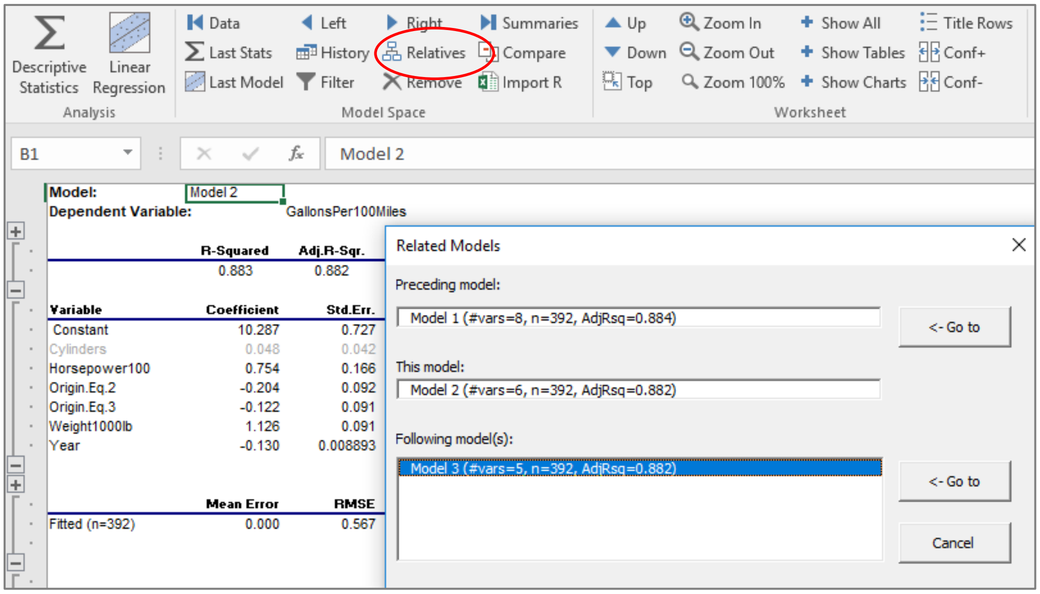
Summary of Regression Model Results

Linear Model For GallonsPer100Miles	Model 1	Model 2	Model 3	
Run Time	1/29/19 9:22 AM	1/29/19 9:23 AM	1/29/19 9:23 AM	
# Fitted	392	392	392	
Mean	4.782	4.782	4.782	
Standard Deviation	1.664	1.664	1.664	
Number Of Variables	8	6	5	
Standard Error of Regression	0.566	0.572	0.573	
R-squared	0.887	0.883	0.883	
Adjusted R-squared	0.884	0.882	0.882	
Mean Absolute Error	0.417	0.413	0.415	
Mean Absolute Percentage Error	9.0%	8.8%	8.9%	8.9%
Maximum VIF	22.938	7.144	4.919	
Normality Test	***	***	***	

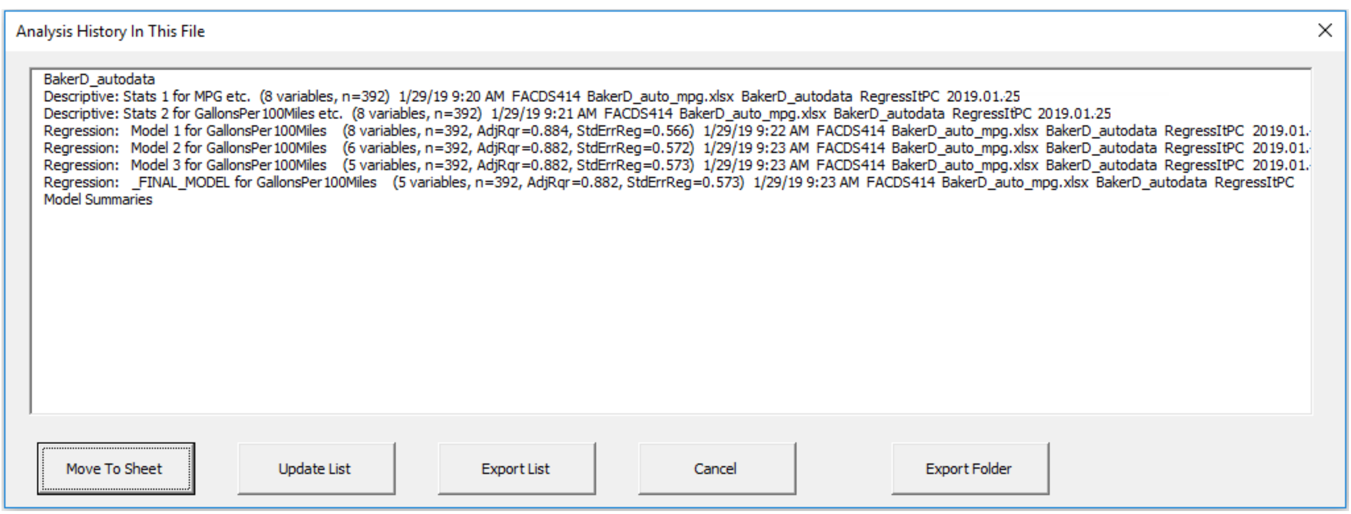
Coefficients:	Model 1	Model 2	Model 3	
Constant	9.392 (0.000)	10.287 (0.000)	10.412 (0.000)	
Cylinders	0.150 (0.007)	0.048 (0.253)		
Displacement100ci	-0.323 (0.014)			
Horsepower100	1.275 (0.000)	0.754 (0.000)	0.805 (0.000)	0.805 (0.000)
Origin.Eq.2	-0.296 (0.002)	-0.204 (0.027)	-0.231 (0.010)	
Origin.Eq.3	-0.203 (0.033)	-0.122 (0.182)	-0.137 (0.128)	
Seconds0to60	0.034 (0.041)			
Weight1000lb	1.125 (0.000)	1.126 (0.000)	1.184 (0.000)	
Year	-0.131 (0.000)	-0.130 (0.000)	-0.131 (0.000)	

The model summaries worksheet provides a complete audit trail of **all models that have ever been fitted in the workbook**, even those whose worksheets were later deleted. The seemingly blank cell above the model name actually contains the model equation in white font and can be used to jump to a model sheet if it exists or to re-run a model whose worksheet was deleted: place the cursor on that cell and hit the Linear Regression button. Optional color coding of coefficients by sign and significance can be turned on as shown above.

When a sequence of regression models is fitted within RegressIt by successively adding or removing variables, they can be navigated by parent-child relationships via the **Relatives** button on the ribbon, as shown below. This feature allows the user to jump to the parent of the currently visible model or to any of its descendants. An instructor can use this tool to trace the paths that students followed in exercises that may have required a lot of searching in the model space. It's also good for re-navigating your own work.



The **History** procedure provides a scrollable pop-up list of all worksheets currently in the file, with some summary information about what each one contains. This is not only a display tool but also a navigation tool. The user can jump directly to any worksheet in the file by clicking on its entry in the list and hitting Move-to-Sheet.



There are also a couple of export options included in the History procedure. The **Export-List** tool will export a one-row-per-worksheet table to the clipboard, containing all the information in the history window and more, from which it can be pasted to a new Excel worksheet. (File Explorer must be closed when doing this.) The two screen shots below show some (not all) of the fields. On the spreadsheet these are side by side in one large table.

Sheet#	Type	Analysis Name	Dep. [1st] Variable	# Variables	# Values	Adj.R-sq.[Mean]	Std.Err.Reg.[Std.Dev.]
1	BakerD_autodata						
2	Descriptive	Stats 1	MPG	9	392	23.446	7.805
3	Descriptive	Stats 2	GallonsPer100Miles	9	392	4.782	1.664
4	Regression	Model 1	GallonsPer100Miles	8	392	0.884	0.566
5	Regression	Model 2	GallonsPer100Miles	6	392	0.882	0.572
6	Regression	Model 3	GallonsPer100Miles	5	392	0.882	0.573
7	Regression	_FINAL_MODEL	GallonsPer100Miles	5	392	0.882	0.573
8	Model Summaries						

Date/Time	Computer	Model Equation
1/29/2019 9:20	FACDS414	
1/29/2019 9:21	FACDS414	
1/29/2019 9:22	FACDS414	Model 1 (#vars=8, n=392, AdjRsq=0.884): GallonsPer100Miles << Cylinders, Displacement100ci, Horsepower100, Origin.Eq.2, Origin.Eq.3, Seconds0to60, Weight1000lb, Y
1/29/2019 9:23	FACDS414	Model 2 (#vars=6, n=392, AdjRsq=0.882): GallonsPer100Miles << Cylinders, Horsepower100, Origin.Eq.2, Origin.Eq.3, Weight1000lb, Year
1/29/2019 9:23	FACDS414	Model 3 (#vars=5, n=392, AdjRsq=0.882): GallonsPer100Miles << Horsepower100, Origin.Eq.2, Origin.Eq.3, Weight1000lb, Year
1/29/2019 9:23	FACDS414	_FINAL_MODEL (#vars=5, n=392, AdjRsq=0.882): GallonsPer100Miles << Horsepower100, Origin.Eq.2, Origin.Eq.3, Weight1000lb, Year

The table includes complete specifications of every regression model, such as its key parameters and statistics, the names of the computer and file in which it was created, the run time, the name of the computer on which the table was downloaded, and the model equation (list of variables). If this worksheet is located in the original analysis file or in another file with the same defined variables, the cell with the list of variables can be used to re-launch the model: just select the cell and hit the Linear Regression button on the ribbon. This will jump to the model worksheet if it exists in the file, otherwise it will pre-select the listed variables for a new model.

The regression models in the file could have been generated in Excel or they could have been generated in RStudio with their output brought back to Excel via the import-R feature on RegressIt's top menu**. Model output that is imported from RStudio is formatted exactly like native RegressIt output, including all the navigation and model refinement options and an entry on the model summaries worksheet.** A user who has already developed some models in RStudio can easily export them all to a well-formatted and easily navigable Excel file, which can make a nice repository and a delivery vehicle for presenting the results to others. (The R code interface in RegressIt includes an option for retrieving the specs of an existing named model in the RStudio workspace. RegressIt can also parse the model specification out of the character string of an "lm" or "glm" command if it is pasted into a cell in the workbook.)

The **Export-Folder** tool (available on PC's only) can be used to draw the same information from all files in a selected folder, such as a folder into which all submitted files for a homework assignment have been downloaded. (Teaching platforms such as Sakai and Canvas offer this download option for grading purposes.) The tool exports a single large table to the clipboard, which contains one row for every worksheet in *every* file, and that row contains all of the same fields as in the Export-List table. This table is extremely useful for grading the work of an entire class. When using this tool, all files in the target folder should be closed, File Explorer should be closed, and a file in a different folder with a blank worksheet for pasting the table should be open.

Although the export-folder tool only runs on PCs, an instructor who is otherwise a Mac user just needs to a use a PC temporarily (or have a TA do so) in order to dump the contents of the homework folder to an Excel file. It makes no difference whether the students have used Macs or PC's to produce their files.

Here's an example of a portion of the exported table produced from a folder containing three user files. Each one contains one or more descriptive analyses and a sequence of regression models:

User File	Sheet#	Type [Name]	Analysis Name	Dep. [1st] Variable	# Variables	# Values	Adj.R-sq.[Mean]	Std.Err.Reg.[Std.Dev.]	Date/Time Created	Computer
AdamsJ_auto_mpg.xlsx	1	AdamsJ_autodata								
AdamsJ_auto_mpg.xlsx	2	Descriptive	Stats 1	MPG	9	392	23.446	7.805	1/29/2019 9:04	JohnLaptop
AdamsJ_auto_mpg.xlsx	3	Regression	Model 1	MPG	1	392	0.692	4.333	1/29/2019 9:05	JohnLaptop
AdamsJ_auto_mpg.xlsx	4	Regression	Model 2	MPG	7	392	0.82	3.31	1/29/2019 9:05	JohnLaptop
AdamsJ_auto_mpg.xlsx	5	Regression	Model 3	MPG	4	392	0.819	3.322	1/29/2019 9:06	JohnLaptop
AdamsJ_auto_mpg.xlsx	6	Regression	Model 4	MPG	3	392	0.818	3.334	1/29/2019 9:06	JohnLaptop
AdamsJ_auto_mpg.xlsx	7	Regression	_FINAL_MODEL	MPG	3	392	0.818	3.334	1/29/2019 9:10	JohnLaptop
AdamsJ_auto_mpg.xlsx	8	Model Summaries								
BakerD_auto_mpg.xlsx	1	BakerD_autodata								
BakerD_auto_mpg.xlsx	2	Descriptive	Stats 1	MPG	9	392	23.446	7.805	1/29/2019 9:20	BakerPC
BakerD_auto_mpg.xlsx	3	Descriptive	Stats 2	GallonsPer100Miles	9	392	4.782	1.664	1/29/2019 9:21	BakerPC
BakerD_auto_mpg.xlsx	4	Regression	Model 1	GallonsPer100Miles	8	392	0.884	0.566	1/29/2019 9:22	BakerPC
BakerD_auto_mpg.xlsx	5	Regression	Model 2	GallonsPer100Miles	6	392	0.882	0.572	1/29/2019 9:23	BakerPC
BakerD_auto_mpg.xlsx	6	Regression	Model 3	GallonsPer100Miles	5	392	0.882	0.573	1/29/2019 9:23	BakerPC
BakerD_auto_mpg.xlsx	7	Regression	_FINAL_MODEL	GallonsPer100Miles	5	392	0.882	0.573	1/29/2019 9:23	BakerPC
BakerD_auto_mpg.xlsx	8	Model Summaries								
ChenX_auto_mpg.xlsx	1	ChenX_autodata								
ChenX_auto_mpg.xlsx	2	Descriptive	Stats 1	GallonsPer100Miles	10	392	4.782	1.664	1/29/2019 9:26	LAB023
ChenX_auto_mpg.xlsx	3	Regression	Model 1	GallonsPer100Miles	1	392	0.783	0.776	1/29/2019 9:27	LAB023
ChenX_auto_mpg.xlsx	4	Regression	Model 2	GallonsPer100Miles	2	392	0.872	0.595	1/29/2019 9:28	LAB023
ChenX_auto_mpg.xlsx	5	Regression	Model 3	GallonsPer100Miles	3	392	0.873	0.592	1/29/2019 9:28	LAB023
ChenX_auto_mpg.xlsx	6	Regression	_FINAL_MODEL	GallonsPer100Miles	2	392	0.872	0.595	1/29/2019 9:29	LAB023
ChenX_auto_mpg.xlsx	7	Model Summaries								

For purposes of grading the selection of a regression model, students can be told to use a standard name such as _FINAL_MODEL for their solution, and the table can then be sorted on this field to group all of the solutions together. If an underscore is the first character in the name, it will sort first in the list, as shown here (other columns are hidden):

User File	Date/Time Created	Computer	Equation
AdamsJ_auto_mpg.xlsx	1/29/2019 9:10	JohnLaptop	_FINAL_MODEL (#vars=3, n=392, AdjRsq=0.818): MPG << Origin.Eq.1, Weight1000lb, Year
BakerD_auto_mpg.xlsx	1/29/2019 9:23	BakerPC	_FINAL_MODEL (#vars=5, n=392, AdjRsq=0.882): GallonsPer100Miles << Horsepower100, Origin.Eq.2, Origin.Eq.3, Weight1000lb, Year
ChenX_auto_mpg.xlsx	1/29/2019 9:29	LAB023	_FINAL_MODEL (#vars=2, n=392, AdjRsq=0.872): GallonsPer100Miles << Weight1000lb, Year

In addition to statistics and variable selections for each model, the table includes the name of the computer on which the model was run, the date and time at which it was run, and the names of the analysis file and data worksheet at the time the model was run. Students should also be told to personalize the names of their analysis files and data worksheets before beginning their work, as in the first column in the table above. (It good in general for users to frequently save files under new names and/or on different computers for backup purposes, which will get tracked here.) With this amount of detail, it is virtually impossible for one student to submit another's work, and the instructor can also get an overview of the modeling process followed by each student, for example, the sequence of models and the time spent on the whole analysis. If this table is copied to a worksheet in a file that contains the data, the text in the model equation cell can be used to launch any of the models.

With this tool it is feasible to take a helicopter tour of the histories of regression analyses submitted by hundreds of students, as well as to assign grades in a systematic way.

It may also be interesting to sort the analysis history table on other fields (e.g., model equation, number of variables, computer name, date/time, etc.) to get other views of how a class pursued the assignment and to look at issues such as collaboration. On assignments that involve fitting a sequence of multiple regression models starting from a large candidate set of variables, it can be quite interesting to see the distribution of search paths as well as final model selections. Again, the models in the file could have been generated in Excel or in RStudio or in both.

One more feature of the Export-List and Export-Folder tools is the ability for the user to include **comments** in the analysis file on one or more worksheets. If a worksheet has the word "Comments" in cell A1, the program will look for text strings in cells A2, A3, A4, etc.. These text string strings will be entered in cells in columns D, E, F, etc., in the row in the table for that worksheet (i.e. a column of text strings on a comments worksheet is mapped into a row of text strings in the exported table of workbook contents). While model-building is still going on, comments worksheets ought to be moved to the right of the Model Summaries worksheet as soon as it is created, in order to avoid interfering with the placement of new model worksheets in sequence.

The comments feature could be used for annotating the models in the file or for entering answers to homework assignment questions that accompany the analysis. For the latter purpose, a comments worksheet could be included in the original data file with questions and spaces for answers in the cells in column A. If the answers to questions are in multiple-choice format, coded as numbers or letters, it would be straightforward for the instructor to insert a column of formulas to automatically compute the total score by student, thus grading a large number of files almost instantly. (It might be good in this case to include an extra line for a short explanation of the answer to a multiple-choice question, to ensure its originality and thoughtfulness.) The file contents could be sorted on the name of this worksheet to group all the scores together. Another useful sorting strategy would be to sort the worksheet on the names of individual questions (i.e., sort on a column that contains the text of a question), which would group all the answers to a given question, allowing the instructor to get an overview of the distribution of answers to it.

An alternative format for entering the text of answers to be graded would be to include one comment worksheet for each question, with the following cell contents: A1 = "Comments", A2 = text of question to answer, A3 = space for grade points (score), A4 = text of student answer. In the instructor view of the history list, the student's answer to each question would extend all the way across the screen and a score could be entered in the cell to the left of it. The worksheet could be temporarily re-sorted on individual questions for ease and consistency of grading each one.

With its combination of model statistics and opportunities for question answering, the export-folder worksheet can serve as a data file on which to perform many kinds of statistical analysis on the performance of a class on any given assignment.

For those who teach **logistic regression**, there is a separate version of the program (RegessItLogistic) that performs logistic regression within Excel and produces detailed highly interactive table and chart output. Examples of table and chart output for a logistic model are shown below. Many of the tables and charts have parameters controlled by "spinners" (up and down arrow buttons) that are next to them and are linked to data on the worksheet. In particular, **the cutoff level can be dialed up and down** and you can observe the changes in classification tables and movement along the ROC curve in real time. Weighting of type I and type II errors can also be interactively varied on sensitivity charts. Confidence levels can also be varied as on linear regression worksheets. And a variable to be used for plotting a logistic curve (with others held fixed at their means) can also be interactively chosen. Customized labels for 0-1 outcomes (e.g. "Alive" and "Dead" in the example below) can be entered either before or after run time, and they interactively show up on tables and charts.

These interactive features are entirely self-contained within the worksheet, linked to tables of data outside the viewing area (columns AL+), and they can be used even if RegressIt is not running and no data worksheet is included. **A file containing a single logistic model worksheet makes a good interactive demonstration tool for a lecture on logistic regression or a toy for students to play with, even if other software is being used.**

Model: Titanic logistic model
Binary Dependent Variable: Survived1st800 **0-1 value labels:** Dead | Alive

Logistic Regression Statistics: Titanic logistic model for Survived1st800 (7 variables, n=800)

R-squared (McFadden)	Adj.R-Sqr.	RMSE	Mean	# Fitted	ROC area	Critical z	Conf. level
0.389	0.374	0.355	0.385	800	0.87	1.960	95.0%

Logistic Regression Coefficient Estimates: Titanic logistic model for Survived1st800 (7 variables, n=800)

Variable	Coefficient	Std.Err.	z-statistic	P-value	Lower95%	Upper95%	VIF	Std. coeff.
Constant	0.476	0.407	1.168	0.243	-0.322	1.274		
Age	-0.025	0.009283	-2.664	0.008	-0.043	-0.006538	1.364	-0.177
Class3	-1.857	0.256	**-7.262**	0.000	-2.358	-1.356	1.840	-0.510
Female	3.084	0.271	**11.385**	0.000	2.553	3.615	1.793	0.813
FemaleClass3S	-1.214	0.356	-3.412	0.001	-1.911	-0.516	1.666	-0.201
MaleAge_0_9	2.907	0.627	4.634	0.000	1.678	4.137	1.337	0.284
MaleClass2	-1.822	0.390	-4.668	0.000	-2.587	-1.057	1.500	-0.327
SiblingSpouseGT2	-2.357	0.568	-4.150	0.000	-3.470	-1.244	1.253	-0.283

Exponentiated Coefficients (Odds Ratios): Titanic logistic model for Survived1st800 (7 variables, n=800)

Variable	Exp(Coeff)	Exp(z*SE)	Lower95.0%	Upper95.0%	Exp(Std.coeff.)
Age	0.976	1.018	0.958	0.993	0.838
Class3	0.156	1.651	0.095	0.258	0.601
Female	21.846	1.701	12.847	37.150	2.256
FemaleClass3S	0.297	2.008	0.148	0.597	0.818
MaleAge_0_9	18.307	3.420	5.353	62.608	1.329
MaleClass2	0.162	2.149	0.075	0.347	0.721
SiblingSpouseGT2	0.095	3.043	0.031	0.288	0.753

Analysis of Deviance: Titanic logistic model for Survived1st800 (7 variables, n=800)

Source	Deg.Freedom	Deviance		P-value	AIC	ROC area		R-squared
Regression	7	415.003	= Chi-square	0.000	667.331	0.87	McFadden	0.389
Residual	792	651.331	= -2 * log likelihood				Cox-Snell	0.405
Null	799	1,066.334	= -2 * null model log likelihood				Nagelkerke	0.550

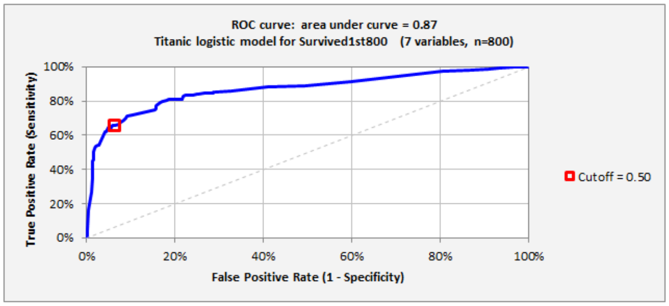
Classification Table: Titanic logistic model for Survived1st800 (7 variables, n=800)

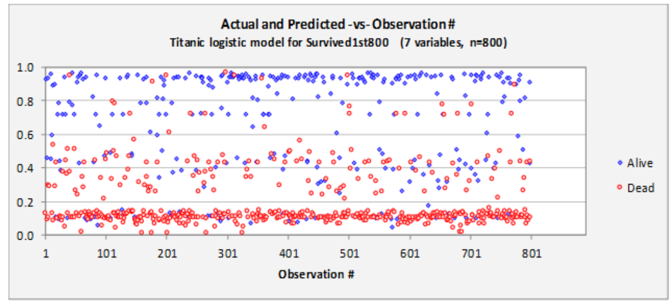
Cutoff value for prediction of Alive: 0.50 RMSE = 0.355

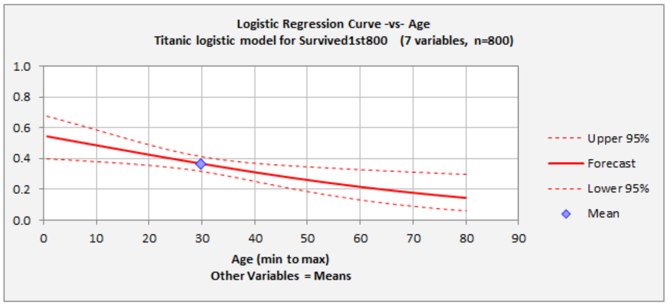
	Predicted:				Predicted:		
Actual:	# Dead	# Alive	Total	Actual:	% Dead	% Alive	Total
# Dead	461	31	492	% Dead	58%	4%	62%
# Alive	105	203	308	% Alive	13%	25%	39%
Total	566	234	800	Total	71%	29%	100%
Percent correct =	83.0%	True positive rate =	65.9%	True negative rate =	93.7%		

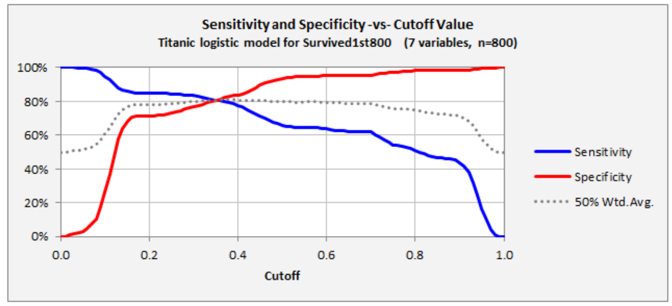
Spinners for adjusting confidence and cutoff levels.

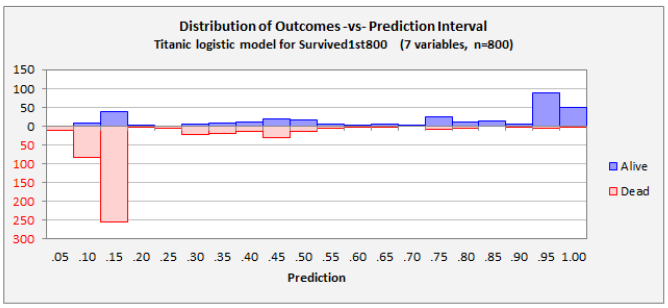
Six different charts are available, and four have interactive features controlled by spinners next to them.

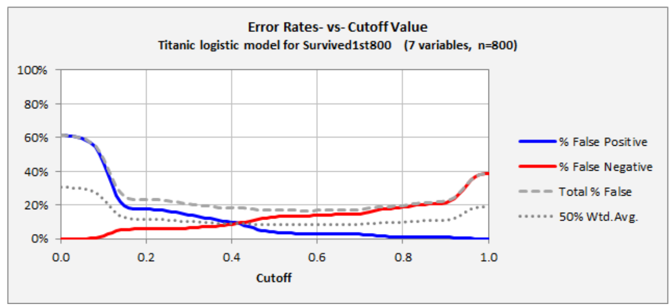












The logistic version of the program runs only in the Windows version of Excel but can be used on a Mac with virtualization software. Logistic models are also limited to 16K rows of data. However, the R interface allows much larger logistic models to be fitted on either PC's or Macs.

* * *

The volume and detail of the output that can be produced by RegressIt, and the number of options on its navigation bar and its menus, might look a little overwhelming for teaching students who are at an early stage of learning how to model. However, the features can be used selectively to show as much or as little detail as is needed, the interface is largely self-documenting, and the output can include embedded teaching notes that provide interpretations of nearly all of the statistics and charts. (Your suggestions for personalization or enhancement of this feature are welcome.) The software itself does not have much of a learning curve, and it can be driven almost entirely with a mouse or touchpad or even by tapping on the touchscreen of a tablet computer. It's a good tool for distributing examples of models to be studied as well as for creating new models. Students (and colleagues) can easily browse the contents of a well-organized Excel workbook, whether or not they have the software installed. It makes a good companion to other Excel add-ins as well as other software packages and programming languages (especially R) by giving students another lens through which to view the model space and another vehicle for presentation. It allows them to devote more time and attention to modeling issues rather than software issues per se and to have a little more fun while doing so. The instructor might also have a little more fun with its features to support grading of large numbers of files.