

ODS Statistical Graphics: How You SHOULD Be Developing Your SAS® Apps

David L. Cassell, Design Pathways, Corvallis, OR

ABSTRACT

SAS® has always been proclaimed by its competitors as trailing other applications in areas such as graphics. In particular, graphical analysis of data has always been a problem area, and manipulation or adjustment of graphs has always been more limited than some other statistical analysis and graphics packages. That is changing with ODS Statistical Graphics in SAS 9.1.3 and SAS 9.2, not to mention the other new graphical features introduced in SAS 9.2. As a result, any application development process has new tools to facilitate the development of analyses, plots, and charts. This will be illustrated with a basic cost modeling example. We will look at building a standard statistical plot with an information box at the side. We will show how to use ODS Statistical Graphics and ODS statements to build such charts for multiple facilities. We will also look at how to work with multiple destinations, since the exact same graphs that are needed in PDF formats are also likely to be needed for company webpages or in acceptable formats for documents and presentations.

INTRODUCTION

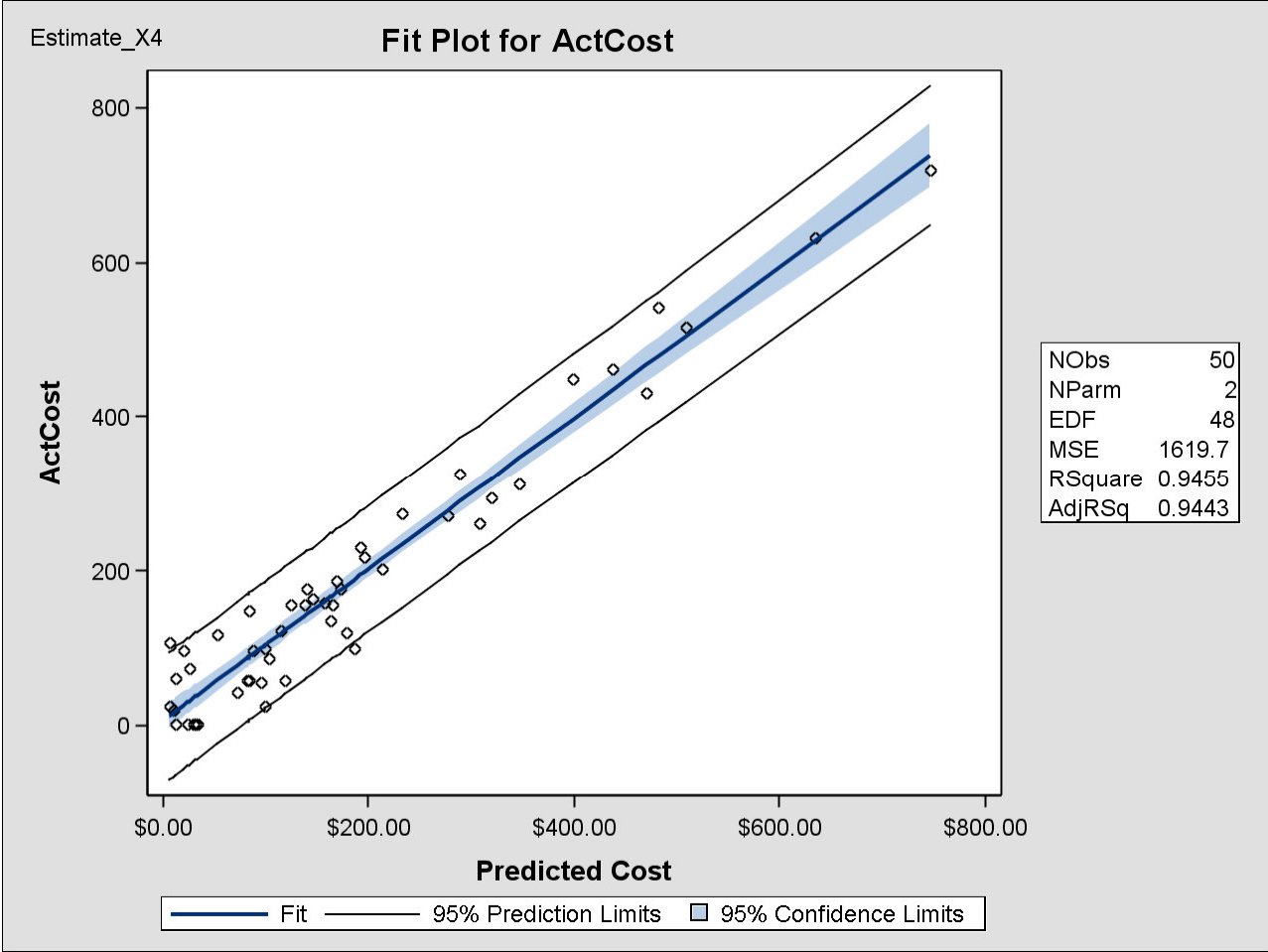
The most useful statistical plots have always been difficult to construct in SAS. And yet, the most useful statistical plots convey the most information to the viewer. For example, a plot of a dependent variable Y against its predicted estimate might not be all that helpful, if you can't tell whether the fit is halfway decent, or the confidence bounds actually cover the majority of the points, or if the points farthest from the 1-to-1 line are really outliers. The statistical 'fit plot' provides enough extra detail to show these features, but it is not easily programmed.

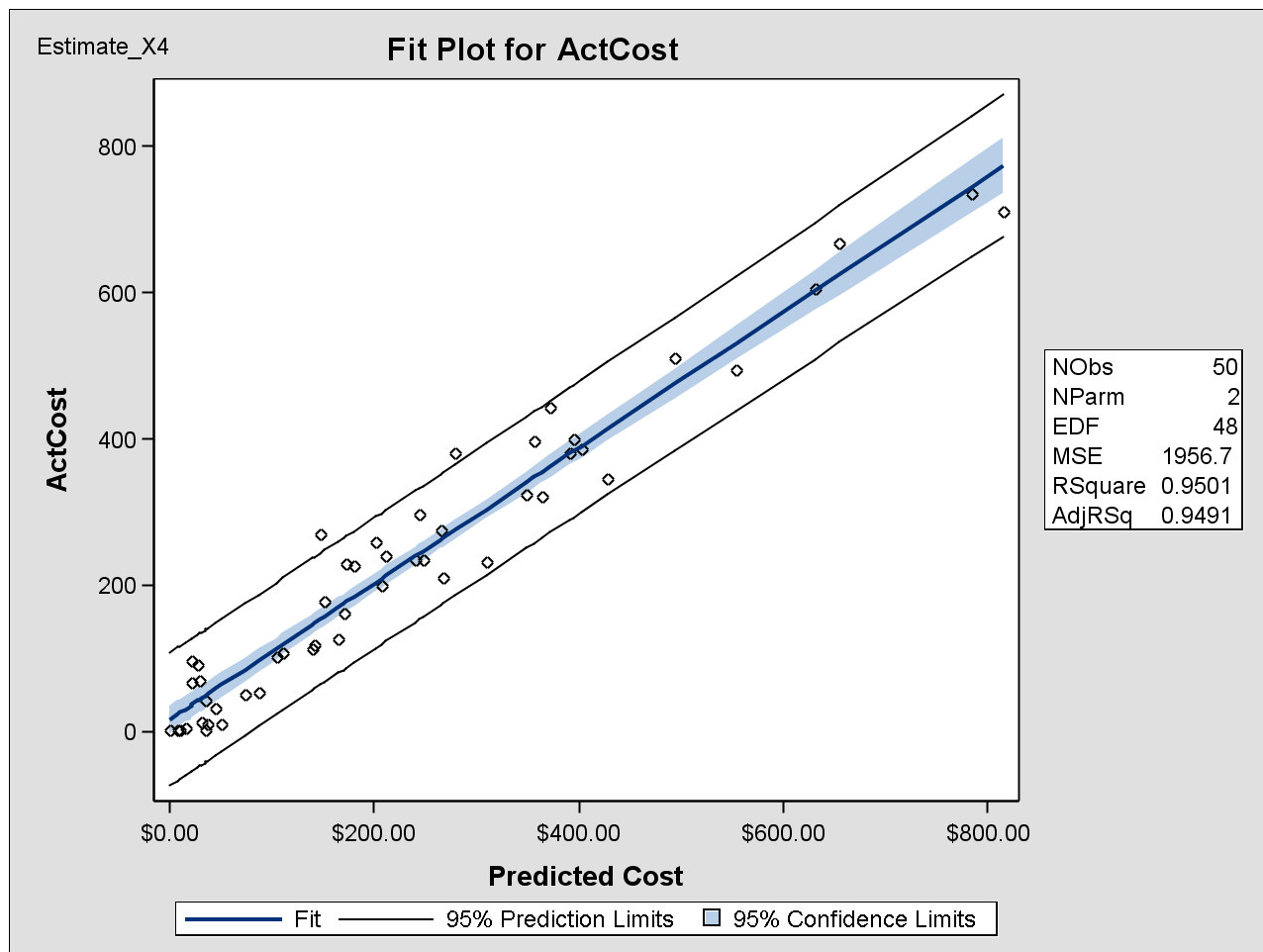
ODS Statistical Graphics comprise two parts: special graphical plots created for each relevant statistical procedure, and (as of SAS 9.2) special graphical procedures which can also use the SAS Graphical Template Language (GTL). This presentation is designed to focus on the former, since these are much easier to use. In fact, they require little more than basic knowledge of ODS statements, plus the knowledge you already had about the SAS statistical procedure you wanted to use.

Let's frame this discussion with an example. Suppose you are given a typical assignment. You have a complicated cost model for the actual cost of a hospital stay. You need to produce appropriate statistical plots for every facility involved. Each plot is a complicated fit plot with auxiliary information in the image. Every image needs to be of publication quality. HTML graphics are required for the company intranet, while PDFs are required for archiving and embedded graphics in RTF are needed for documentation. In addition, different styles are being requested for each output type. This can be done without ODS, but it's an extremely complicated process.

On the other hand, if there is an ODS Statistical Graphics template that meets your specifications, then everything becomes much simpler. (In fact, if there is an ODS Statistical Graphics template that is relatively close to your specifications, everything becomes much simpler, but we'll get to that later.)

So let's start out by assuming that we need, for each hospital facility, and for every different type of graphic format requested, the following type of statistical chart.





Clearly, this would be a lot of work to create by hand. This gets more complex with the more facilities we have, and the more types of output we are required to create for each facility. Fortunately, ODS Statistical Graphics is here to help us out.

CODING - AND NOT CODING

Before we start talking about how all of this can work for you, let's help you out. Assuming that you are reading this paper to learn more about the topic, you'll want a data set to play with, so you can see how this works. The sample data set used in this presentation for the above graphics (and graphics for another 45 facilities) is as follows:

```
data tempcost;
  label PredCost='Predicted Cost' ActCost='Actual Cost';
  format PredCost dollar9.2 ActCost dollar9.2;
  do facility = 1001 to 1093 by 2;
    do n = 1 to 50;
      PredCost = ranexp(38474626) * 200;
      ActCost = max(3.50, PredCost + rannor(38474626) * 40);
      output;
    end;
  end;
run;
```

This small snippet of SAS code will give you a data set to play with, so you can follow along at home.

Next, we need enough SAS code to create all the graphics, for all the facilities, for all the different output types. Most people immediately leap to the SAS Macro Language for an iterative structure like this. But the natural ability of SAS to work with normalized data sets leads us to a much simpler solution. The entire code needed to do all the work we have discussed is no more than this:

```
options nodate nonumber;
ods noproctitle;

title 'Actual Cost vs. Cost Estimate X4';
ods html style=Statistical path='C:\SAS\temp' gpath='C:\sas\temp\graphics'
(url='graphics/') file='file.htm';
ods rtf style=Default file='C:\sas\temp\ex1.rtf';
ods pdf style=Journal file='C:\sas\temp\appendix04.pdf';
ods graphics;
ods select Fit;

proc reg data= tempcost;
  Estimate_X4: model ActCost = PredCost;
  by facility;
run;

ods graphics off;
ods _all_ close;
ods listing;
```

There is no macro code here. And there are no special features apparent in the PROC REG code that we use here. Why is this? Because all the hard work is done by ODS under the hood. Now let's talk about how this actually works.

DISCUSSION

In order to see how all this works, we need to start by talking about the SAS Output Delivery System (ODS). ODS was created to separate the analytical process of a procedure from the mechanics of the resulting display. The output is handled separately, allowing SAS to create statistical graphics as automatically as it creates tables. ODS handles the output *destinations* slightly differently, since these destinations have different features and different structures. For example, HTML allows the easy assimilation of several different image types as separate charts or plots which are referenced from a master webpage. On the other hand, PDF and RTF require the images to be embedded within the file. So SAS ODS handles these details for us, although we may at some point need to know that we cannot treat our images exactly the same in HTML and PDF.

ODS Statistical Graphics is 'experimental' in SAS 9.1, and production in SAS 9.2. This does not mean that it is unstable or unusable in SAS 9.1.3. On the contrary, it is extremely stable. However, the 'experimental' monicker does mean that you do not need to own a copy of SAS/GRAPH to run it, as you will in SAS 9.2 and all subsequent SAS versions.

The SAS ODS destinations include nearly any format and destination you might want to use for SAS output. The most popular destinations include HTML (and a host of variants), RTF, PDF, LaTeX, and a variety of XML types. The DOCUMENT destination is also available. In addition, the LISTING destination for ODS is the SAS Display Manager 'Output' window. Unfortunately, the OUTPUT destination for ODS is not the output window, but output tables: the output destination creates SAS data sets.

The usual format for an ODS statement looks like one of the two following forms:

```
ODS [destination] <CLOSE | options>;

ODS [keyword];
```

Examples of both of these forms are in the source code above.

ODS Statistical Graphics gives SAS several needed capabilities. It allows the creation of important, useful statistical graphics and diagnostics without extensive coding. It provides presentation quality graphics without resorting to external programs. It provides customization and a variety of options, as well. This gives it the flexibility to provide the sort of application development that we so often require, without months of coding and debugging.

The second line of code, ODS NOPROCTITLE, addresses a common problem we have had for years. The output has an annoying line in the header that states the name of the procedure used in the analysis. ODS NOPROCTITLE (for NO PROCedure TITLE) removes this from the headers.

The ODS HTML statement simultaneously handles a variety of tasks for us. The option STYLE=STATISTICAL defines the graphical style to be used in the output. As of SAS 9.1.3, there were only 4 official output styles for ODS Statistical Graphics: DEFAULT, STATISTICAL, JOURNAL, and ANALYSIS. They work well for most purposes, so these are example styles in our code. The ODS HTML statement allows us to define a directory for our HTML file with the PATH= option, as well as defining a directory for the graphics files to be accessed from that file, with the GPATH= option. The URL= sub-option of the GPATH= option is phenomenally useful, since it forces SAS to encode all the paths as relative paths, instead of hard-coded paths; this means that moving the entire directory to another location will not break your encoded paths/filenames. The FILE= option lets us define the file name (of course).

But the ODS HTML has further flexibility. One option lets us specify - not a single file - but a trio of files to hold an HTML frame, an HTML table of contents, and an HTML body. The form for this looks like:

```
ods html [other options] body='file1' contents='file2' frame='file3';
```

The ODS RTF statement has different options. Among other issues, you have to embed the images in the file, and you can't have a separate frame/contents/body structure as in a webpage. The ODS RTF and ODS PDF statements can follow one another, or the ODS HTML statement, in any order, or can be used individually. You do not need to use only a single ODS destination at a time.

The ODS PDF statement, like the ODS RTF statement, allows you to define a single file to hold the entire generated graphical output. Both let you define the STYLE= option and the FILE= option too. Note that you can use different styles in different statements. You can make sure the HTML output meets the official company graphical style for its intranet, while the other destinations can meet other stylistic requirements. In fact, you can create your own style template and have that used by the ODS statements.

The ODS GRAPHICS statement turns on the ODS Statistical Graphics package, now that the destinations have all been defined. It is recommended that you turn this off again once you are done, as it is very CPU-intensive.

The ODS SELECT statement lets you choose a single graph or a list of graphs from the complete set that would normally be created by the SAS procedure. In this case, we select only the FIT graphic for our output. ODS also has an ODS EXCLUDE statement that lets you select only the graphics to be omitted. I recommend that you use whichever is easier: if the list of graphs to be excluded is shorter, then use that.

At this point, you have to be able to *find* the ODS Table Name for the graphs of interest. Fortunately, there are now a lot of ways to do this. The SAS Online Docs have a list of these Table Names for every SAS procedure which uses ODS Statistical Graphics. This list is always toward the end of the documentation, just before the list of references. The ODS TRACE statement can also be used to show the ODS Table Names, as well as details such as the template used to create the graph. Simply inserting the statement

```
ODS TRACE ON;
```

before the code executes, and inserting the statement

```
ODS TRACE OFF;
```

afterward provides a list of all the ODS Table Names in the log. I personally prefer the alternative form

```
ODS TRACE ON / LISTING;
```

as it inserts the table information interleaved with the graphics themselves, better enabling your ability to tell which table goes with which Table Name.

But these are the only options. You can also use the Display Manager. If you click on the Results window and right-click on the graph, one of the Properties will be the ODS Table Name. You can also use the Display Manager to access the SAS Templates from the menu.

The PROC REG code uses no additional features in order to access the ODS Statistical Graphics. The BY statement allows for efficient BY processing, but that is aimed at simplifying the process rather than affecting the ODS graphical process. In SAS 9.2, some ODS Statistical Graphics will be available in the PROC statement of each procedure: the details are in the SAS Online Doc. In SAS 9.1, this is not the case, although there are procedures in which ODS keywords are needed in order to get specific graphics.

Once the processing is done, you should use the ODS GRAPHICS OFF statement. In fact, you will probably also want to turn off real-time graphics processing while the computations are being done, as this can slow down the process. The Tools menu in the Display Manager has an Options sub-heading, which in turn has a Preferences sub-heading. That pops up a window which has a Results tab, with a check box allowing you to turn off the real-time processing of the display of the plots, while the plots are being generated for your files.

Then it is time to close the ODS destinations, so the destination files can be completed and saved. The statement ODS _ALL_ CLOSE does this, closing every one of the open destinations. Unfortunately, the output window counts as one of these destinations, so you need to re-open it afterward. The ODS LISTING statement does this for us.

Unfortunately, we know that nothing is perfect. We may need to customize, and re-organize. Fortunately, all of this is feasible now.

The style templates are customizable. But the *graphics* templates are customizable too. The ODSTEMPLATE command (usable in the Command window of the Display Manager) gives us instant access to the ODS graphics templates. The new Graphics Template Language (GTL) gives us functional manipulation of the templates, as well as tools for the new SG⁺ procedures of SAS 9.2. If you need to use PROC DOCUMENT to manipulate your output, you can use the DOCUMENT destination, too. This gives us enormous flexibility in re-structuring the order of the graphics and tables in our output without needing to re-arrange the procedures or remove the power of the by-processing. And - as of SAS 9.2 - there is a new ODS Graphics Editor. It will let us manipulate and change the style and appearance of an ODS graph without impacting the results of the graph. You can highlight a point or change the style of the presentation, but you *cannot* alter the data points. This is a good thing.

CONCLUSIONS

The new ODS Statistical Graphics features provide more flexibility, more convenience, and more graphical power than we have had in SAS before. The presentation quality of the output alone makes this worthwhile. The functionality of it gives us powerful application development tools. And the added power of the style templates, the graph templates, GTL, and PROC DOCUMENT make the process far easier than it used to be.

REFERENCES

SAS 9.1.3 Online Doc, Copyright 2007, SAS Institute Inc., Cary NC, USA.

SAS 9.2 Online Doc, Copyright 2008, SAS Institute Inc., Cary NC, USA

ACKNOWLEDGMENTS

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.

CONTACT INFORMATION

The author welcomes questions and comments. He can be reached at his private consulting company, Design Pathways:

David L. Cassell
Design Pathways
3115 NW Norwood Pl.
Corvallis, OR 97330
DavidLCassell@msn.com
541-754-1304