

Plotting Data in SAS

Programming Workshop

Presented by the Applied Statistics Lab

Ya Qi

February 15, 2018

1 Introduction

Today we will go over some basic plotting in SAS. Let's start by first importing our data with the import procedure and then look at it with the proc contents statement.

```
/* Import the data */
proc import out = practicedata file = 'C:\Users\AG ASI\Documents\Workshop\SAS\practicedata.xlsx'
  dbms = xls replace;
  getnames = yes;
  sheet = 'practicedata';
run;

/* Look at the data */
proc contents data = practicedata;
run;
```

Figure 1: Importing and looking at the data.

2 Plotting One Quantitative Variable

Plotting one quantitative variable is often useful when looking at a response variable. The two methods we will look at today include a

- histogram
- boxplot

Creating a histogram

In SAS, you can use either the sgplot or the univariate procedures to create a histogram. The code for plotting a histogram with proc sgplot is:

```

/* Creating a histogram */
proc sgplot data = practicedata;
  title 'Histogram of Response Variable'; /* change the main title */
  histogram respvar /
    /* scale = count /* histogram of counts */
    /* fillattrs = (color = red) /* change the color */;
  xaxis label = 'Response Variable'; /* change the x- axis label */
run;

```

Figure 2: Creating a histogram with the proc sgplot statement.

And here is the resulting output.

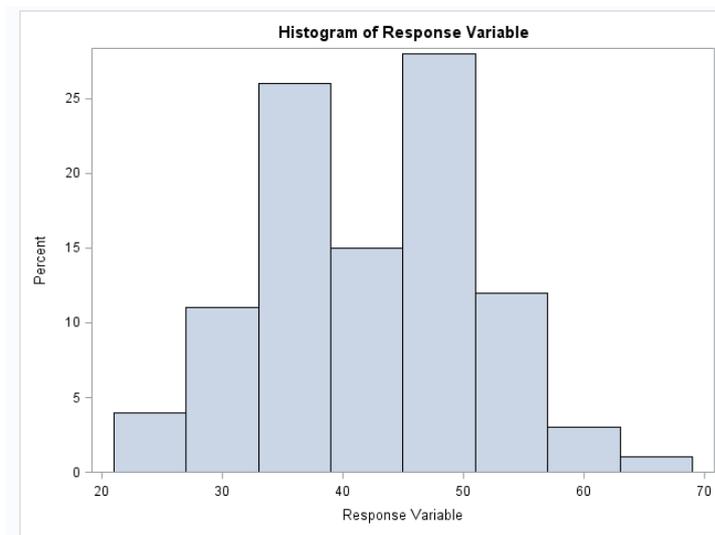


Figure 3: Histogram of the response variable.

If instead you would like to use proc univariate to create your histogram, you can do so with:

```

proc univariate data = practicedata noprint;
  title 'Histogram of Response Variable';
  histogram respvar;
run;

```

Figure 4: Creating a histogram with the proc univariate statement.

Creating a boxplot

To create a boxplot for a single quantitative variable in SAS, you will want to use proc sgplot. Here is the resulting boxplot created from the code.

```
/* Creating a boxplot */  
proc sgplot data = practicedata;  
  title 'Boxplot of Response Variable';  
  /* yaxis label = 'Response Variable'; /* change the y axis label */  
  vbox respvar /* use hbox instead if want horizontal boxplot*/  
  /* fillattrs = (color = red) /* change the color */;  
run;
```

Figure 5: Creating a boxplot with the proc sgplot statement.

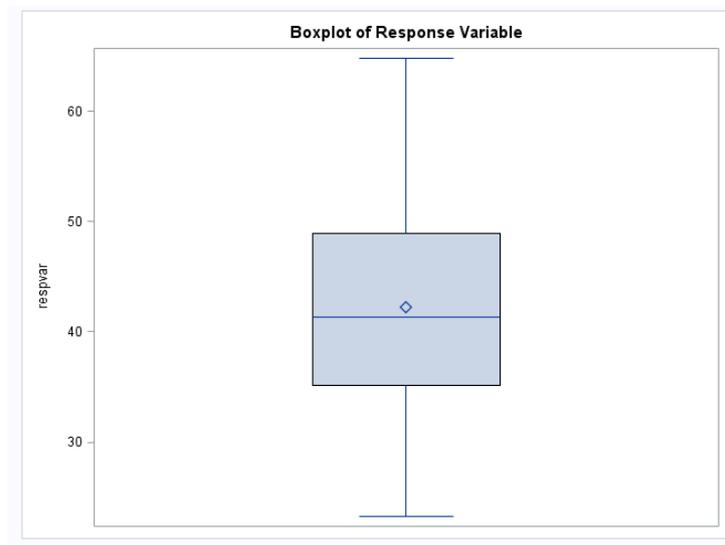


Figure 6: Boxplot of the response variable.

3 Plotting One Categorical Variable

Plotting one categorical variable is often useful when looking at a grouping variable. The two methods we will look at today include a

- bar chart
- pie chart

Creating a bar chart

To make a bar chart, we will once again use proc sgplot.

```
proc sgplot data = practicedata;  
  title 'Bar Chart of Groups';  
  vbar groupvar; /* use hbar for horizontal barplots instead of vertical */  
run;
```

Figure 7: Creating a bar chart with proc sgplot.

This code results in the following bar chart.

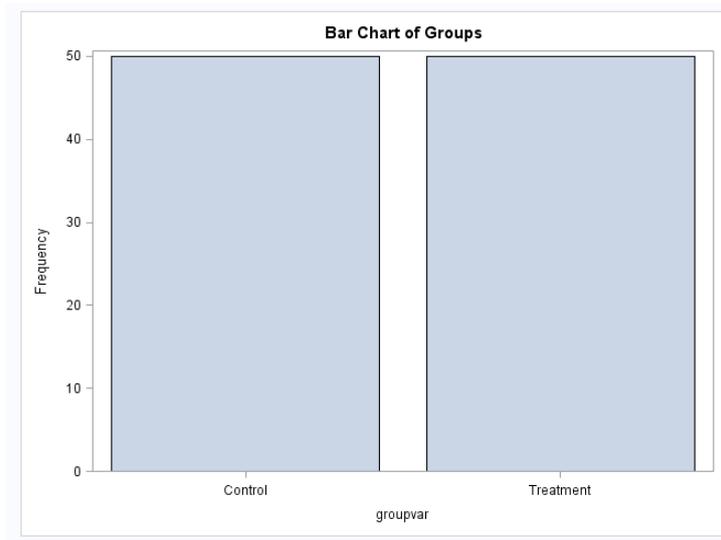


Figure 8: Bar chart of the grouping variable.

Creating a pie chart

To make a pie chart, we will use proc gchart.

```
proc gchart data = practicedata;  
  title 'Pie Chart of Grouping Variable';  
  pie groupvar;  
run;  
quit;
```

Figure 9: Creating a pie chart with proc gchart.

This code results in the following pie chart.

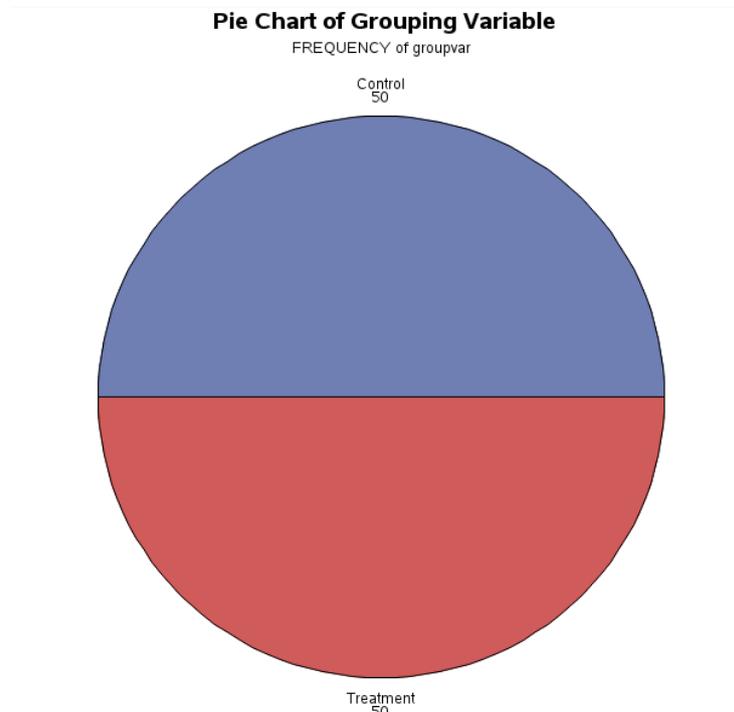


Figure 10: Pie chart of the grouping variable.

4 Plotting Two Variables

There are often times when you will want to look at graphs with two variables.

Plotting One Quantitative and One Categorical Variable

Today we will look at

- side-by-side boxplots
- two histograms

Side-by-side Boxplots

Side-by-side boxplots can be useful when looking at a response by two groups. We will look at how to do this using

- `proc sgplot`
- `proc boxplot`

Here is the code to make side-by-side boxplots using `proc sgplot`.

```

❏ proc sgplot data = practicedata;
    title 'Boxplots';
    vbox respvar / group = groupvar;
    yaxis label = 'Response Variable';
run;

```

Figure 11: Creating side-by-side boxplots with proc sgplot.

This code results in the following boxplots.

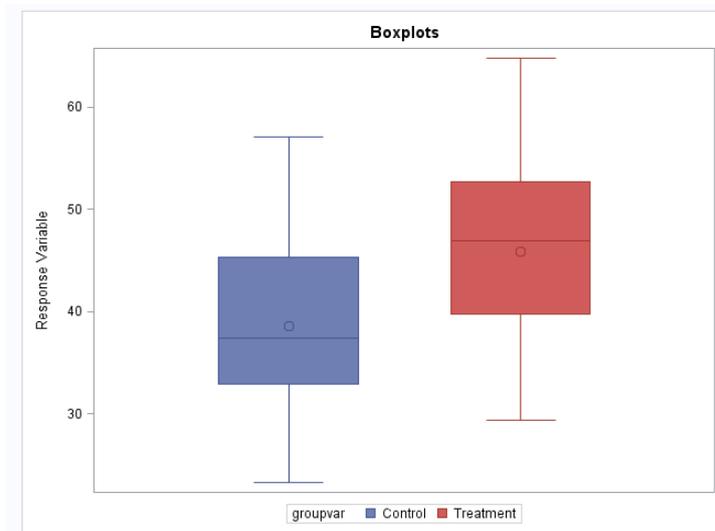


Figure 12: Side-by-side boxplots of the response by the grouping variables.

To make side-by-side boxplots using proc boxplot, you can use the following code. Note that you will want to first sort the data by the grouping variable. This code results in the following boxplots.

```

❏ proc sort data = practicedata;
    by groupvar;
run;
❏ proc boxplot data = practicedata;
    plot respvar*groupvar; /* continuous variable*categorical variable */
    label respvar = 'Response Variable';
run;

```

Figure 13: Creating side-by-side boxplots with proc sgplot.

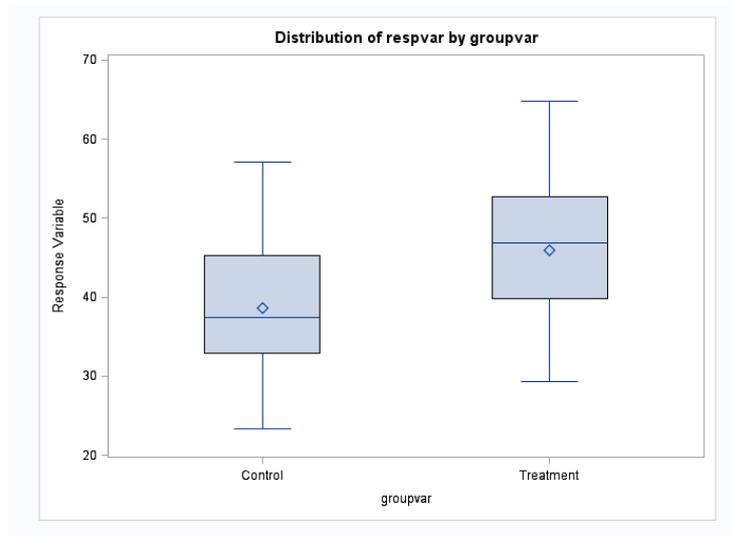


Figure 14: Side-by-side boxplots of the response by the grouping variables.

Two Histograms

Creating two histograms can also be useful when looking at a response variable by two different groups. Here is the code to make two histograms using proc sgplot. Note that first, we need to sort our data by the grouping variable.

```

proc sort data = practicedata;
  by groupvar;
run;
proc sgplot data = practicedata;
  title 'Histogram of Response Variable';
  by groupvar; /* grouping variable used to break into two histograms */
  histogram respvar;
run;

```

Figure 15: Creating two histograms with proc sgplot.

This code results in the following histograms.

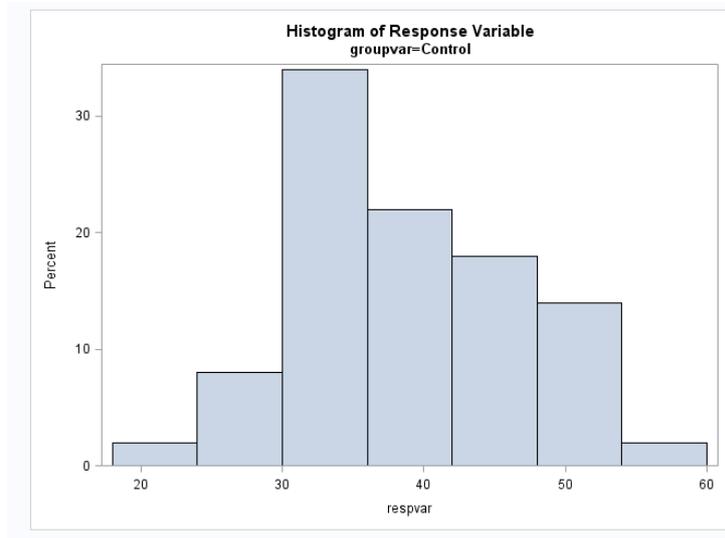


Figure 16: Histogram of the response for the control group.

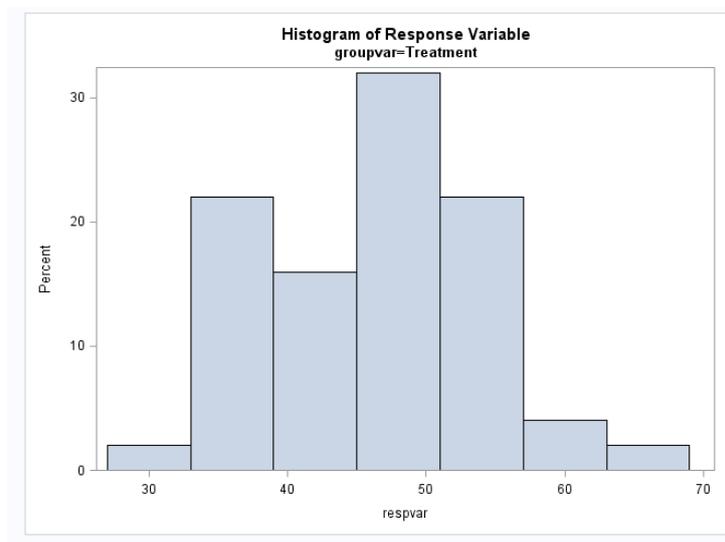


Figure 17: Histogram of the response for the treatment group.

Plotting Two Quantitative Variables Using a Scatter plot

A scatter plot is useful when you have two quantitative variable to look at. To create a scatter plot in SAS, you will use sgplot procedure.

```

proc sgplot data = practicedata;
  title 'Plot of Response Variable vs. Toy Variable';
  scatter x = toyvar y = respvar /
    markerattrs = (color = blue symbol = circlefilled); /* changes color and symbol */
  xaxis label = 'Toy Variable' display = (noticks novalues);
  yaxis label = 'Response Variable';
run;

```

Figure 18: Creating two histograms with proc sgplot.

This code results in the following scatter plot.

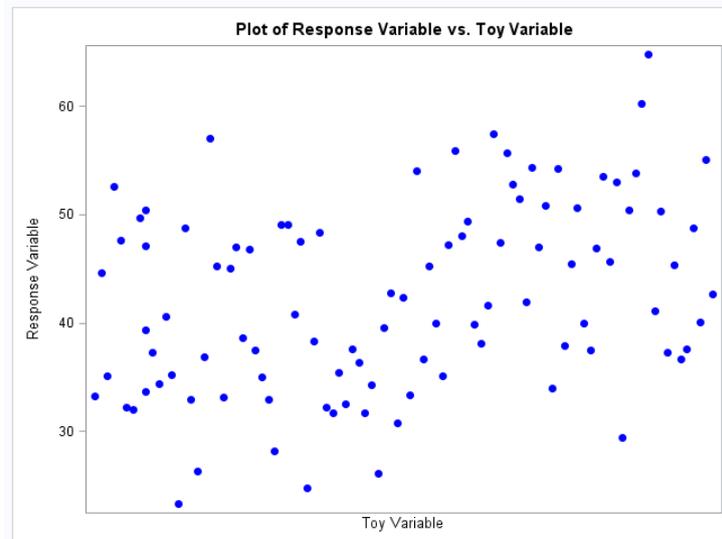


Figure 19: Scatter plot of the response against the explanatory variable.

5 Plots Showing More Than Two Variables

Sometimes you may want to include information from more than two variables on a plot. In SAS, often you can incorporate this within your plotting procedure (proc sgplot), but you can also do this using proc sgpanel. First, we will look at Creating a scatter plot in sgplot with three variables.

Creating a scatter plot with all of the data

Here, we want to create a scatter plot of the response variable against the toy variable as before, but now we will include the group variable by coloring the dots according to the grouping variable.

```

proc sgplot data = practicedata;
  title 'Plot of Response Variable vs. Toy Variable';
  scatter x = toyvar y = respvar /
    group = groupvar
    markerattrs = (symbol = circlefilled); /* changes symbol */
  xaxis label = 'Toy Variable' display = (noticks novalues);
  yaxis label = 'Response Variable';
  /* keylegend / noborder position = top title = 'Group Variable' location = inside; /* options to change the legend */
run;

```

Figure 20: Creating a scatter plot with three variables.

This code results in the following scatter plot.

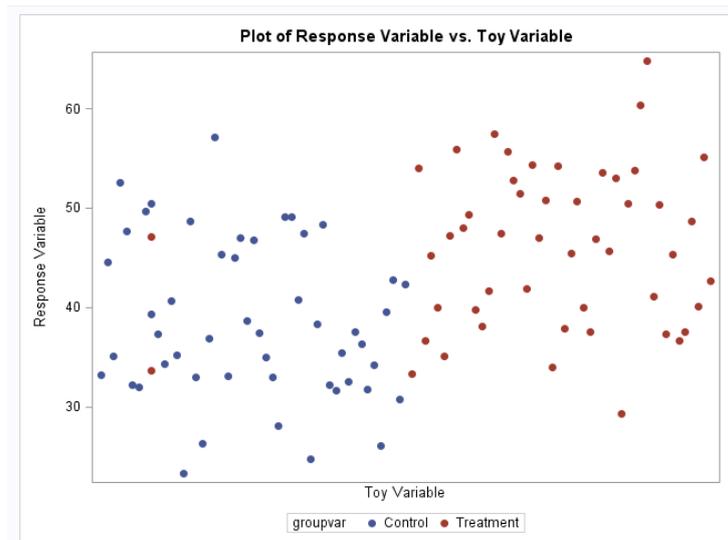


Figure 21: Scatter plot of the response against the toy variable with groups designated by color.

Using sgpanel to create side-by-side scatter plots

The sgpanel procedure is also really nice for creating side-by-side plots or for incorporating multiple plots into one big graph. In the following code, we are splitting the above scatter plot into two side-by-side scatter plots by group.

```

proc sgpanel data = practicedata;
  title 'Scatterplots of Response Variable against Toy Variable by Group';
  panelby groupvar; /* Variable you want to separate by */
  scatter x = toyvar y = respvar /
    markerattrs = (symbol = circlefilled);
  rowaxis label = 'Response Variable';
  colaxis label = 'Toy Variable' display = (noticks novalues);
run;

```

Figure 22: Creating side-by-side scatter plots of groups for response variable against toy variable.

This code results in the following set of scatter plots.

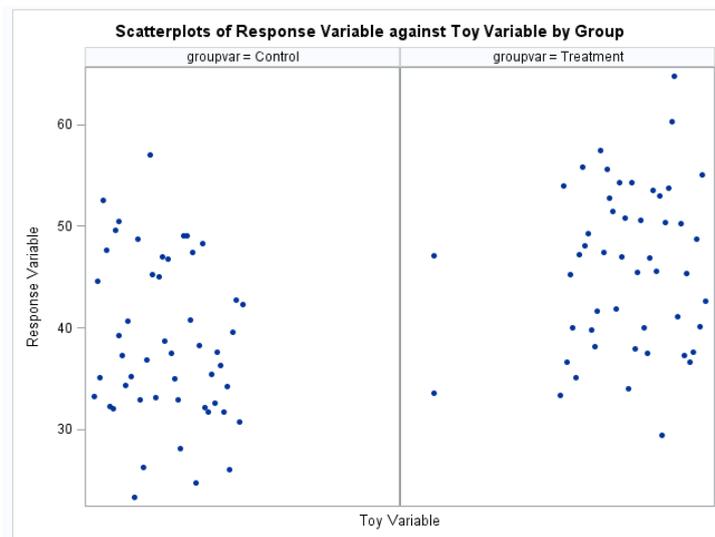


Figure 23: Scatter plots of the response against the toy variable by group.

6 Need statistical assistance?

If you are in need of statistical assistance feel free to contact us! To do so you can,

- Submit a request: <https://stat.as.uky.edu/asl>
- Email: asl@uky.edu