CHAPTER 5 Appendix

Two-Way Tables

While some of these technologies can help with the material in this chapter, a simple calculator (or even a hand calculation) is really all that is needed. The biggest concern for most students is working with conditional distributions. Care must be taken to determine the correct conditioning, which depends on how the question is worded. Key words and phrases such as "if/when/given x, what proportion of y…" mean we are restricting the focus to only the column (or row) of the table where variable x has a specific value. It is important to understand what the "whole" of interest (x) represents in these situations.

To graph a *marginal distribution* from a two-way table, follow the methods described for bar graphs in the Chapter 1 appendix, because these are distributions for a single variable. In this appendix, we describe some methods (where the technology can create them) to create bar graphs for *conditional distributions*. There are two methods: side-by-side (or clustered) bar graphs and stacked bar graphs (essentially pie charts, but with columns instead of pies).

As the "wholes" are typically of different sizes, it is important to scale these graphs in percents instead of counts.

Graphs for Two-Way Tables

Excel

1. Enter the data as a two-way table, and use the cursor to highlight the entire table, but not any column or row sums—just the data to be displayed.
2. **Insert ➔ Column or Bar Chart** 
3. From the 2-D Columns menu, select either clustered  or stacked (with the bars of equal heights)  .

   Note: The clustered bar option can display only counts. To use this option, you will need data entered as conditional percents, or else you will need equal numbers of observations for each value of the conditioning variable.
4. Excel uses the row variable as the default conditioning variable (the variable displayed on the x axis). If that is not how your data were entered, click **Design ➔ Switch Row/column**.
5. Double-click the "Chart Title" placeholder and give the graph a meaningful title.

To create a clustered bar graph (or stacked bar graph) with the y axis in counts, do the following:

1. Ensure that your data are entered as a table, with two columns for the categorical variables and one column for the counts.
2. **Graph ➔ Graph Builder**
3. Select the conditioning variable and drag it to the x axis.
4. Select the bar graph  from the chart styles row.
5. Select the "response" variable and drag it to the **Overlay** box.
6. Select and drag the count variable to the **Freq** box.

At this point, you have a side-by-side (clustered) bar graph with the *y* axis in counts.

7. If you want a stacked bar chart, use the drop-down for **Bar Style ➔ Stacked**.

For a *y* axis in percent—as in a stacked bar graph, where each bar has total height 100% (for a conditional distribution)—do the following:

1. **Analyze ➔ Tabulate**
2. Drag the conditioning variable to the **Drop zone for rows** box.
3. Drag the response variable to the **Drop zone for columns** box.
4. Drag the count variable to the **Freq** box.

At this point, JMP will display what it understands the two-way table to be.

5. Drag the **Row %** statistic into the table area.
6. Click the red triangle next to Tabulate. Then select **Make Into Data Table**.

You will be in a new data window with the data in percents.

7. **Graph ➔ Graph Builder**
8. Highlight and drag all the percent variables to the *y* axis.
9. Drag the conditioning variable to the *x* axis.
10. Click on the bar graph style. At this point, you have the clustered bar graph.
11. If you want a stacked bar graph, use the drop-down for **Bar Style ➔ Stacked**.
12. Double-click in the title placeholder and give the graph a meaningful title.
13. **Done**

Minitab

1. Your data should be entered as a table, with one column for each value of the categories and one column for the counts.
2. **Graph ➔ Bar Chart**
3. Use the drop-down to change "Bars represent" to **Values from a table**.
4. Under "One column of values" select either **Cluster** or **Stack**.
5. Highlight and **Select** the count variable column into the **Graph variables** box.
6. Highlight and **Select** the two categorical variables into the **Categorical variables for grouping** box. Make sure the variable you are using for conditioning (*x* as we have used it before) is specified first.
7. Use **Labels** to give the bar graph a meaningful title.
8. **Continue**
9. Click **Chart Options**. Click to check the box next to "Show Y as Percent." Click to change the radio button to "Within categories at level 1."
10. **Continue ➔ OK**

1. Your data should be entered as a table, with one column for each value of the categories, and one column for the counts.
2. **Graphs ➔ Chart Builder ➔ Bar**
3. Select and drag either the clustered bar or stacked bar icon into the graph area.
4. Select and drag the conditioning variable to **X-axis?**.
5. Select and drag the second categorical variable to **Cluster (or Stack) on X: set color**.
6. Select and drag the count variable to **Count** on the *y* axis.
7. Give the bar graph a meaningful title using **Titles/Footnotes**.
8. **OK**

9. If you are creating a stacked bar chart, you need to adjust from counts to percents. Double-click in the graph for the Graph Editor. On the menu ribbon, click **Scale to 100%**  at the far right. The default label for the Y axis for these graphs is "Mean count." To change this, within the Chart Editor, double-click on the label, and enter a new label (usually "Count" or "Percent."
10. Click **X** to close the graph editor.

Note: Clustered bar charts can be created only for counts; stacked bars can use either counts or percents.

Note: When using the Chart builder, be sure to **RESET** between graphs within a session. Otherwise, you may end up with a meaningless chart!

CRUNCH

CrunchIt! cannot create bar graphs for conditional distributions.

TI-83/-84

TI calculators cannot create bar graphs for conditional distributions.

There may be easier ways to graph conditional distributions with packages, but the following procedure uses only basic R commands.

Enter the data in order, with the outermost (conditioning) variable repeated the same number of times as the number of categories for the response variable. Note that the values of the conditioning variable must be in the columns of the matrix that is constructed.

Here we look at the results of a study of complications from different types of gastric surgery intended for weight loss. "NLT" means non-life threatening.

```
> surgery<-c('banding','banding','banding','bypass',
  'bypass','bypass','gastrectomy','gastrectomy',
  'gastrectomy')
> comp<-c('NLT','none','serious','NLT','none','serious',
  'NLT','none','serious')
> count<-c(81,606,31,5253,8110,604,46,325,19)
> gastric=data.frame(surgery,comp,count)
> levels(gastric$surgery)
> levels(gastric$comp)
> data=count
> data=matrix(data,ncol=3,byrow=T)
> colnames(data)=levels(gastric$surgery)
> rownames(data)=levels(gastric$comp)
```

At this point, you have created a matrix data structure. To see the matrix, simply enter the command **data**. The results are shown here.

```
> data
          banding   bypass gastrectomy
NLT            81      606          31
none         5253     8110         604
serious        46      325          19
```

To build the bar graph, we use the columns of this table (type of surgery) as the conditioning variable. We need column percents for our stacked bar graph.

```
> prop=prop.table(data,margin=2)
> par(mar=c(5.1, 4.1, 4.1, 7.1), xpd=TRUE)
> barplot(prop, col=heat.colors(length(rownames(prop))),
  width=2), legend("topright",inset=c(-0.25,0),
  fill=heat.colors(length(rownames(prop))),
  legend=rownames(data))
```

If you add the parameter **beside=T** to the barplot command, you will have a clustered bar graph.