



Using SPSS for Customer Analysis

Your Assignment is to replicate the Information Contained here in SPSS. Copy and paste the SPSS output for each of the statistical techniques in parts II to VI and write one to two sentences (short interpretations/explanations) of the results.

After completing all the analysis, please answer the following three questions regarding the overall analysis:

- A. What would the results you see mean to you as a marketing manager?
- B. Are there opportunities for growth?
- C. Should there be more analysis?

Please label your analyses beginning with Numeral II.

The Very Basics

I. Simple Analyses with Categorical Data

- Frequencies
- Bar Charts
- Cross-tabs

II. Simple Analyses with Continuous Data

- Descriptives
- Histograms
- Correlations

III. Simple Analyses with both Categorical and Continuous Data

- Histograms for subcategories
- Descriptives statistics for subcategories
- Graphing the relationship between two variables

IV. Data Manipulation – creating new variables

- Computing a new variable
- Recoding a variable into a new variable
- Generating a decile or quintile variable

VI. Selecting Records from a Dataset

All of the SPSS output shown in the following pages is based on a sample of 1000 customer records and variables drawn from the PCsUnlimited file of customer information. The variables used are:

- Respond: Did the customer respond to the latest promotional offer? (1=Yes, 0=No)
- Gender: 0 = female, 1 = male
- Age (in years)
- NumPurch: Total number of purchases made by the customer
- TotDol: Total dollar purchases made by the customer
- Region: region of the country where the customer resides

The sample dataset is found in PCsUnlimited.sav if you wish to reproduce the results shown below, or experiment with additional analyses.

I. The Very Basics

Help

Select anything you are curious about in an SPSS window by clicking it with the *right* mouse button.

There is an online tutorial that can be run by selecting Tutorial from the SPSS Help menu.

Data Files (*.sav files)

The SPSS Data Editor window displays your working data file in a format that looks very similar to a spreadsheet.

SPSS can read in (and export out) various types of files (including Excel files), but we will be working with SPSS formatted data files – which are designated as *.sav files.

Note: SPSS allows you to have only one data file open at a time. If you open a second file, the first one will be closed.

Output Files (*.spo files)

All the outputs from SPSS analyses are displayed in the Viewer Window and can be saved to a file. SPSS automatically designates output files with a *.spo extension. Once you have saved an output file, you can reopen it and continue with additional analyses or you can copy and paste parts of the output to Word or PowerPoint files.

The Data Editor Window

At the bottom left of this window you see two tabs (similar to Excel): one labeled 'Data View' and the other labeled 'Variable View'. Clicking on the Data View tab will display the data set. Clicking on the Variable View tab will display a list of the variables – including names, type (numeric vs. string), format and descriptive labels.

In the Data View window, each row is a case (or observation) and each column is a variable. The cells hold the value of a particular case for a particular variable.

At the top of every column in the data file is the **variable name** – which can be no more than eight characters long. Fortunately, SPSS allows us to assign more descriptive **labels** to variables. If you place the cursor over the variable name, the label (if one has been entered) will be displayed. To change or add descriptive labels, click on the Variable View tab and type in a new label in the label column for that variable. In the Variable View window you can also change formatting options such as how many decimals are displayed. To change formats click on the appropriate cell in the Variable View window and a menu or dialog box will appear.

Missing data are common in marketing datasets. A single period or decimal point (that is, ".") indicates missing data for that cell. It is also possible that missing data have been assigned some out-of-range value such as '9' or '99' – if this is the case, then you must verify that SPSS 'knows' that these values indicate missing data. To do this, click on the Variable View tab to see what are specified as missing values for that variable.

The Viewer Window

The Viewer window is where you will see the statistics and graphs -- i.e., the **output** from SPSS. The Viewer window is split into two parts, or panes. The left side (the **outline pane**) contains an outline view of all the different pieces in the Viewer, whether they are currently visible or not. The right side (the **display pane**) contains the output itself.

The Outline Pane

Individual portions of output are associated with “book” icons in the outline pane. Each icon represents a particular piece of output, such as a table of statistics or a chart. A “closed book” icon represents a hidden piece of output.

- If you click one of these icons in the outline pane, the associated piece of output appears instantly in the display pane. Hidden output doesn’t appear in the display pane but can be recovered when you want to look at it by double-clicking the closed book icon to open it.
- It is easy to rearrange or delete output using the outline pane. To rearrange the output, press the left mouse button on a book icon and drag it to a different place in the outline, then release the mouse button. To delete part of the output, click on the icon and press the Delete key.

The Display Pane

The display pane shows as much of the SPSS output as will fit in it. To see more, use the scroll bars or use the outline pane to jump around.

- To change something about an object (such as a table, a title, or a chart) in the display pane, double-click it in the display pane. This opens an editor with the editing options.
- Most of SPSS’s tabular and statistical output appears in the form of **pivot tables**. Double-clicking a pivot table lets you edit it ‘in place’. You can change almost anything about a pivot table. For example,
 - To change the label, double-click on it and enter a new one.
 - To change a column width, drag the column divider one way or the other.
 - To change the format, click on the text and change the font, etc. using the format menu.
 - To switch the rows and columns, activate a pivot table and select ‘Pivot ... Transpose Rows and Columns’ from the menu.
 - Charts can be resized by pointing the mouse at one of the corner or side ‘handles’, pressing the mouse button and dragging the handle.
 - To change anything inside a chart, activate it by double-clicking and then use the menu options.

To switch between the Data Editor and Viewer windows, use the 'window' menu on the menu bar.

SPSS allows you to have more than one output window open at a time. To open a second (or third, or fourth, ...) output window, from the 'File' pulldown menu, select 'New' and then 'Output'. If you have more than one output window open, SPSS will use whichever one is currently *designated*. To tell which window is designated, look at the center bottom of the screen – if there is an exclamation point there, then the window is designated. To change which window is designated, click the icon with the exclamation point at the right end of the upper tool bar.

Copying SPSS output to a Word document

- Click on the output that you want to copy or select it using the outline pane (or right-click on it)
- From the *Edit* menu, select *Copy Objects*. (It won't format correctly if you choose *Copy*).
- *Paste* into your Word document.

Copying SPSS output to a PowerPoint document

- Click on the output that you want to copy or select it using the outline pane (or right-click on it)
- From the *Edit* menu, select *Copy*. (It won't format correctly if you choose *Copy Objects*).
- *Paste* into your Powerpoint document.

Exporting Data from an SPSS Output Table to Excel

It is sometimes useful to export data from an SPSS output table to Excel. Here are instructions to do this:

- Double click on the table to 'activate' it
- Right click and choose *Select Table*
- With the table selected, right click again and select *Copy*
- Open a new worksheet in Excel and *Paste*

II. Simple Analyses with Categorical Data

Frequencies

The “Frequencies” procedure is useful for summarizing the number and percent of observations that have each value of a categorical variable. For example, the PCsUnlimited’s dataset includes a variable called REGION indicating which REGION each customer lives in.

To determine out how many customers are from each REGION:

1. From the pull-down menus, select **Analyze / Descriptive Statistics / Frequencies**.
2. In the dialog box that appears, select REGION for the variable list (click on REGION, then click on the arrow to move it to the Variable List).
3. Make sure that the “Display Frequency Tables” box contains a check.
4. Click on the **Charts** button. Select “None.”
5. Click “Continue.”
6. The **Statistics** and **Format** buttons allow you to change the output produced. For this example, no changes from the default values are required.
7. Click “OK”. The Output window will display both the number and percentage of customers with different values for the REGION variable.
8. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.

		Region			
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	NorthEast	227	22.7	22.7	22.7
	MidAtlantic	167	16.7	16.7	39.4
	SouthEast	81	8.1	8.1	47.5
	MidWest	263	26.3	26.3	73.8
	SouthCentral	98	9.8	9.8	83.6
	West	164	16.4	16.4	100.0
	Total	1000	100.0	100.0	

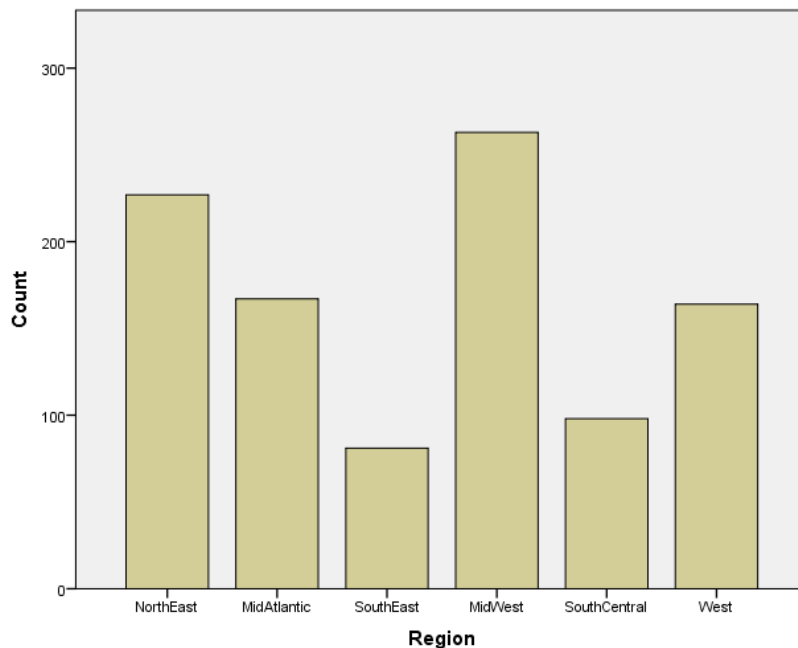
Hints and Options:

1. If your dataset contains missing values – the percentages in the ‘Percent’ and ‘Valid Percent’ columns will differ. The numbers in the ‘Percent’ column are based on all observations – including observations with a missing value for that variable. In contrast, the numbers in the ‘Valid Percent’ column are based only on those observations that contain valid data (i.e. missing observations are excluded).

Bar Charts

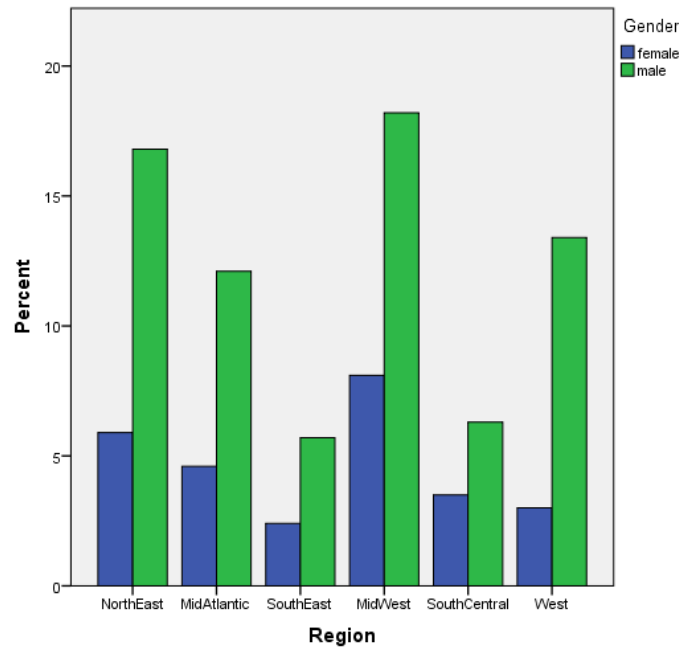
Bar charts are commonly used to provide a visual summary for categorical data. For example, to create a bar chart showing the number of customers in each region:

1. From the pull-down menus, select **Graphs / Chart Builder**.
2. In the dialog box that appears, select OK.
3. Gallery should be selected. If not select Gallery
4. In dialog box "Choose from" Bar should be selected. If not, select Bar.
5. Hover over icon of bar chart in "Choose from" dialog box. You will see "Simple Bar". Double Click.
6. This will activate a bar chart in "Chart preview" window.
7. Select REGION from the variable list and drag and place in X-axis? in "Chart preview" window.
8. (Optional) Click on the Titles/Footnotes tab and enter a title.
9. Click "OK" to create this chart.
10. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.



Hints and Options:

1. You can use the SPSS Graphs menu to create a wide variety of graphs. Below is another bar chart – this time a 'Clustered' bar chart showing the percent (instead of the frequency count on the vertical axis):



Crosstabs

By forming two-way (or multi-way) tables, the “Crosstabs” procedure counts records for two (or more) categorical variables. In addition, statistical tests (including the commonly used chi-square test) can be used to test for significant associations between categorical variables.

For example, if you want to know what percentage of females vs. males responded to the latest offering from PCsUnlimited, we need to calculate four percentages: (1) % of females who responded; (2) % of females who did not respond; (3) % of males who responded; and (4) % of males who did not respond.

1. From the pull-down menus, select **Analyze / Descriptive Statistics / Crosstabs**
2. In the dialog box that appears, select GENDER for the Row variable and REPOND for the column variable.
3. Click “Cells.” Click on “Row percentages.” This will compute the percentage of males and females who responded (as long as Gender is the row variable).
4. Click “Continue” and then click “OK” to produce the crosstabs listing:

Gender * Response Crosstabulation

		Response		Total
		no	yes	
Gender	female	Count	232	43
		% within Gender	84.4%	15.6%
	male	Count	632	93
		% within Gender	87.2%	12.8%
Total	Count	864	136	1000
	% within Gender	86.4%	13.6%	100.0%

If we want to know whether the higher percentage of females who responded (15.6%) compared with males (12.8%) is statistically significant, we can add the following to the analysis:

1. Click on the “Statistics” box at the bottom of the Cross-tabs dialog box.
2. In the dialog box that appears, click on Chi-square to produce the following.
3. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)	Exact Sig. (2- sided)	Exact Sig. (1- sided)
Pearson Chi-Square	1.339 ^a	1	.247		
Continuity Correction ^b	1.110	1	.292		
Likelihood Ratio	1.307	1	.253		
Fisher's Exact Test				.256	.146
Linear-by-Linear Association	1.337	1	.248		
N of Valid Cases	1000				

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 37.40.

b. Computed only for a 2x2 table

The p-value (i.e. Asymp. Sig. (2-sided)) for the Pearson Chi-square is .247 – well above the typical .05 cutoff, indicating that the difference in response rate between males and females is *not* statistically significant.

Finally, suppose that instead of finding the % of females vs. males who responded, we want to know the % of responders and non-responders who were male or female. In other words, we want to know: (1) % of responders who are female; (2) % of responders who are male; (3) % of non-responders who are female; and (4) % of non-responders who are male. To calculate these values, use the above procedure, with one small change: Instead of clicking on “Row percentages,” click on “Column percentages”. You can also create a crosstabs listing with **both** row and column percentages, though some find this confusing!

Gender * Response Crosstabulation

			Response		Total
			no	yes	
Gender	female	Count	232	43	275
		% within Response	26.9%	31.6%	27.5%
	male	Count	632	93	725
		% within Response	73.1%	68.4%	72.5%
Total	Count		864	136	1000
	% within Response		100.0%	100.0%	100.0%

Hints and Options:

1. It doesn't matter which variable is the row variable and which is the column variable – although you will need to think through whether you want to request row or column percentages depending on which is your row and which is your column variable.

Hint: if you are cross-tabbing two variables and one has many categories – it is best to make that one the row variable. For example, consider a cross-tabs of gender by profession (and you have 25 professions). That will either produce a table with 2 columns and 25 rows or one with 25 columns and 2 rows – and the first (2 columns and 25 rows) will fit on a sheet of paper much better!

2. The **Statistics** and **Format** buttons allow you to change the output produced. You may also request a bar chart and, if you want only a bar chart, you can suppress creation of the table, by clicking on the appropriate boxes.

III. Simple Analyses with Continuous Data

Descriptives

The “Descriptives” procedure produces summary statistics such as mean, standard deviation, min and max values, sum and range. These statistics are useful for variables measured on a continuous scale – such as dollars, years, units, pounds, or meters.

Suppose you want to know the *total* dollars spent by all customers and the *average* dollars spent per customer as well as the *largest* and *smallest* amounts spent by any customer. You also want the same summary statistics for number of purchases

1. From the pull-down menus, select **Analyze / Descriptive Statistics / Descriptives**.
2. In the dialog box that will appear, select TOTDOL and NumpURCH for the variable list.
3. Click on the **Options** button. Check the “Mean,” “Sum,” “Minimum,” and “Maximum” boxes, and remove the checks from the remaining boxes.
4. Click “Continue.”
5. Click “OK”. The Output window will display following result.
6. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.

Descriptive Statistics

	N	Minimum	Maximum	Sum	Mean
Total Dollars Spent	1000	50	8763	335229	335.23
Total Number of Purchases	1000	1	15	1822	1.82
Valid N (listwise)	1000				

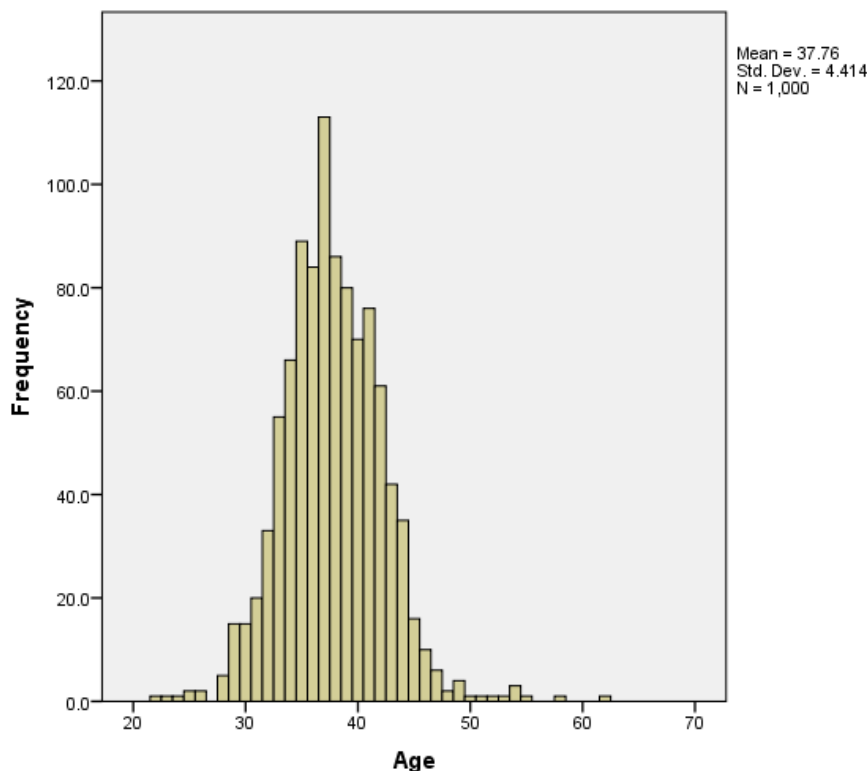
Histograms

The “Descriptives” procedure just used gives us some information about the distribution of the values of a variable. As we saw, we can find the minimum and maximum values and the mean; we can also display the variance and standard deviation, to give us some idea of how tightly values are clustered around the mean.

It is often more useful, though, to see a histogram showing the distribution of the values. There are two ways to create histograms in SPSS: (1) using the histogram option under the Charts menu, or (2) with the “Frequencies” procedure under the Analyze menu. The following examples show how to create a histogram showing the distribution of customer ages.

Using Histogram:

1. From the pull-down menus, select **Graphs / Chart Builder**.
2. In the dialog box that appears, select OK.
3. Gallery should be selected. If not select Gallery
4. In dialog box “Choose from” select Histogram.
5. Hover over icons of histograms in “Choose from” dialog box until you see the one labeled simple histogram. Double Click.
6. This will activate a histogram chart in “Chart preview” window.
7. Select AGE from the variable list and drag and place in X-axis? in “Chart preview” window.
8. (Optional) Click on the Titles/Footnotes tab and enter a title.
9. Click “OK”. A histogram showing the distribution of AGE will appear in the Output window. The mean and standard deviation for AGE are also printed.
10. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.



Using Frequencies: (this option allows you to get frequency tables as well as a histogram)

1. From the pull-down menus, select **Analyze / Descriptive Statistics / Frequencies**.
2. In the dialog box that appears, select AGE for the variable list.
3. Click on the **Charts** button. Select "Histograms."
4. Click "Continue."
5. The "Display Frequency Tables" box is probably checked. Click this box to "un-select" this option (if you forget to do this you will get a long – and probably useless – frequency table)
6. The **Statistics** and **Format** buttons allow you to change the output produced.
7. Click "OK". A frequency histogram showing the distribution of AGE will appear in the Output window. As before, the mean and standard deviation for AGE will also be printed.

Correlations

One measure of the relationship between continuous variables is their correlation.
To compute the correlation between total dollar purchases and number of purchases:

1. From the pull-down menus, select **Analyze / Correlate / Bivariate**.
2. Select NUMPURCH and TOTDOL for the Variables list.
3. Under “Correlation Coefficients,” select “Pearson.”
4. If you want statistically significant correlations marked with an asterisk, select “Flag significant correlations.”
5. Click “OK” to produce the output.
6. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.

Correlations		Total Number of Purchases	Total Dollars Spent
Total Number of Purchases	Pearson Correlation	1	.857**
	Sig. (2-tailed)		.000
	N	1000	1000
Total Dollars Spent	Pearson Correlation	.857**	1
	Sig. (2-tailed)	.000	
	N	1000	1000

** . Correlation is significant at the 0.01 level (2-tailed).

Hints and Options

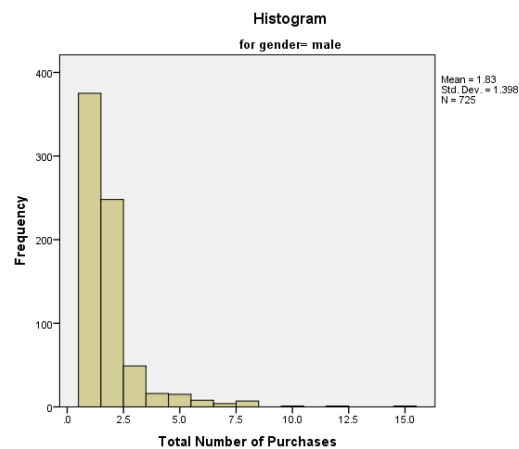
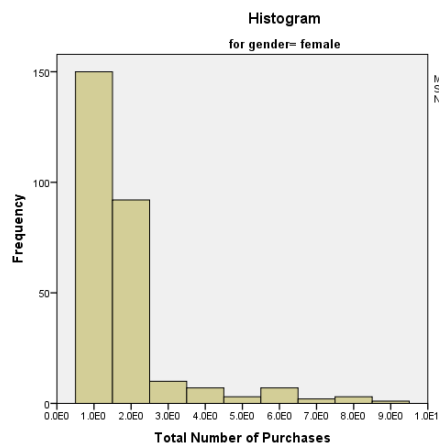
1. Bear in mind that customer datasets are often extremely large. Due to the large sample size, relatively small correlation coefficients may be statistically significant. You should look at the value of the coefficient as well as its significance level. In this example, the correlation is both statistically significant and high.

IV. Simple Analyses with both Continuous and Categorical Data

Histograms for subcategories

The “Explore” procedure can create histograms for a variable subdivided by categories. For example, to see how male versus female (a categorical variable) customers compare in terms of total number of purchases (a continuous variable):

1. From the pull-down menus, select **Analyze / Descriptive Statistics / Explore**.
2. In the dialog box that appears, select NUMPURCH for the Dependent List and GENDER for the Factor List.
3. You have three choices of Display options: Both, Statistics, and Plots. Select “Plots.”
4. Click on the **Plots** button. In the Boxplots section, select “None.” In the Descriptive section, select “Histogram.” If “Stem and Leaf” is selected, click on the box to “un-select” it.
5. Click “Continue” and then “OK” to produce the following:.
6. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.



Hints and Options

1. Select continuous variables for the Dependent List and categorical variables for the Factor List.
2. The Statistics options in the Explore procedure can identify extreme observations:

Extreme Values

		Case Number		Value
Total Dollars Spent	Highest	1	438	8763
		2	927	8430
		3	639	7076
		4	527	6405
		5	21	6218
	Lowest	1	548	50
		2	338	50
		3	180	50
		4	8	50
		5	993	51 ^a

a. Only a partial list of cases with the value 51 are shown in the table of lower extremes.

Descriptive statistics for subcategories

You may want more than a count of the number of records belonging to a subcategory. If you also want to see summary statistics for categories or groups, you can use the “Case Summaries” procedure.

Suppose you want to know both the number and percent of customers who live in each region. Earlier, we got that information using the “Frequencies.” In addition, you want to know the total dollars spent, average dollars spent, and percent of overall total dollars placed by customers in each region.

1. From the pull-down menus, select **Analyze / Reports / Case Summaries**.
2. In the dialog box, select REGION as the grouping variable. Select TOTDOL for the variable list.
3. If the “Display Cases” box is selected, click on it to “un-select” this option.
4. Click on the **Statistics** button. Select “Number of Cases” (counts the number of customers per region), “Percent of Total N” (gives the percentage of total customers from each region), “Mean” (average \$ per customer), “Sum” (total \$ spent by all customer) and “Percent of Total Sum” (gives each region’s customers spending as a percent of the overall total \$).
5. Click “Continue.”
6. If you wish, Click “Options,” type in a report title, and then click “Continue.”
7. Click “OK” to generate the report.
8. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.

Case Summaries

Total Dollars Spent

Region	N	% of Total N	Mean	Sum	% of Total Sum
NorthEast	227	22.7%	380.74	86427	25.8%
MidAtlantic	167	16.7%	301.50	50351	15.0%
SouthEast	81	8.1%	322.86	26152	7.8%
MidWest	263	26.3%	357.40	93996	28.0%
SouthCentral	98	9.8%	367.62	36027	10.7%
West	164	16.4%	257.78	42276	12.6%
Total	1000	100.0%	335.23	335229	100.0%

Hints and Options

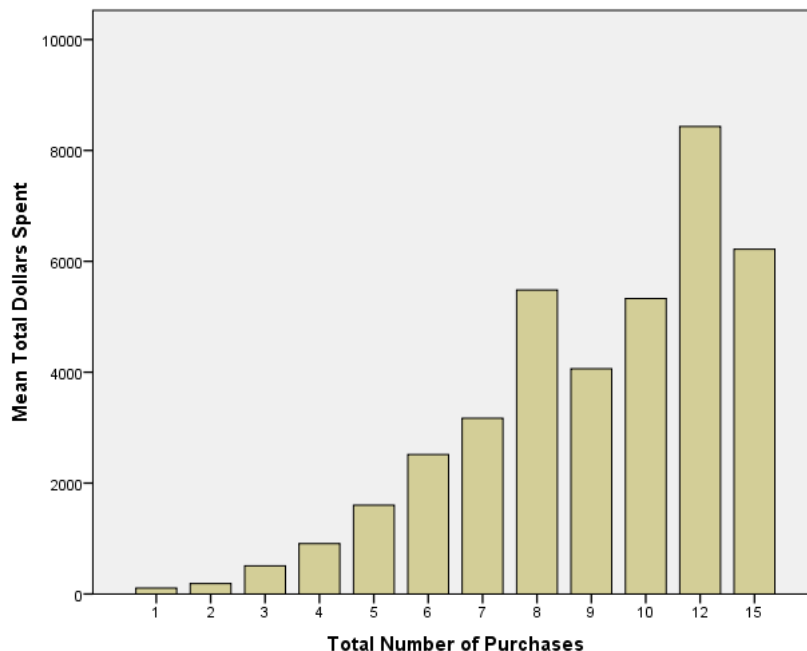
1. Select categorical variables for the grouping variable(s) and continuous variables for the Variable(s).
2. The *Options* box will allow you to add whatever title you choose.

Graphing the relationship between two variables

SPSS allows you to create many different types of graphs showing the relationship between variables. In this example, we will create a bar graph.

Suppose you want to graph the relationship between total dollar purchases and the total number of purchases.

1. From the pull-down menus, select **Graphs / Legacy Dialogs / Bar.**
2. Click on "Simple," and select "Summaries for groups of cases." Then click on "Define."
3. In the dialog box that appears, select NUMPURCH as the category axis.
4. For "Bars represent," click on "Other statistic".
5. Select TOTDOL for the variable list. "MEAN(Total Dollars Spent)" will appear in the variable list box, which indicates that each bar will represent the average dollar purchases by all customers who have made a particular number of purchases.
6. If you like, click on "Titles" and enter a title or subtitle for the graph.
7. Click "OK" to generate the bar graph.
8. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.



V. Data Manipulation – Creating New Variables

Computing a new variable

New variables can be created using the **Transform / Compute** procedure. For example, you may want to compute an average of several variables, or a total, or some other numerical function.

The PCsUnlimited dataset includes total number of purchases and total dollar purchases for each customer. To compute the average purchase dollar amount:

1. Make sure you are in the Data Editor window.
2. From the pull-down menus, select **Transform / Compute Variable**.
3. Type the new variable name (perhaps AVGDOL) in the “Target Variable” box.
4. Click on “Type&Label” below the “Target Variable” box.
5. In the dialog box that will appear, click on “Numeric” for variable type, and type a variable label (such as “Average Purchase \$ Amount”) if desired. Then click “Continue” to return to the main dialog box for computation.
6. In the “Numeric Expression” box, type TOTDOL / NUMPURCH.
7. Click on “OK.” The command will execute; depending on the file size and the speed of your PC, this could take from several seconds to several minutes. This step will add a new column to the dataset containing the AVGDOL variable.
8. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.

If you want to keep your variable for future use, remember to save your dataset!

Now, if you use the **Descriptives** you will find:

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Average Purchase \$ Amount	1000	14.00	1095.38	126.6833	138.68636
Valid N (listwise)	1000				

Hints and Options

1. You can create more sophisticated formulas including different functions of existing variables (using the function list in the dialog). You can also click on the “If.” button on the dialog box to assign values to variables only under certain conditions.
2. In our example, we added a new variable; however, you can also use **Transform / Compute Variable** to change the value of an existing variable.

Recoding a variable into a new variable

Suppose you want to divide your customers into two groups: one-time purchasers and repeat customers or those who have purchased more than once.

An easy way to do this in SPSS is to use the ***Transform / Recode/ Into Different Variables*** to create a new 'Yes/No' variable:

1. Make sure you are in the Data Editor window.
2. From the pull-down menus, select ***Transform / Recode Into Different Variables***.
3. Select NUMPURCH for "Input Variable → Output Variable."
4. Under 'Output variable,' type a name for the new variable (perhaps REPEAT) and a brief description. Click "Change."
5. The "Numeric Variable → Output Variable" box should show:
NUMPURCH → REPEAT
6. Click the "Old and New Values" box.
7. Click the "Output Variables are Strings" box on the lower right-hand side of the dialog box.
8. Under "Old Value," click "Value" and type a 1 in the box.
9. In the "New Value" box, type No.
10. Click "Add." This specification will set REPEAT to 'No' whenever NUMPURCH is equal to one (i.e., the customer made only one purchase)
11. Under "Old Value" click "Range: value through HIGHEST" and type a 2 in the box.
12. In the "New Value" box, type Yes.
13. Click "Add." This specification will set REPEAT to 'Yes' whenever NUMPURCH is 2 or greater (i.e., the customer has made 2 or more purchases).
14. The "Old→New" box should show two lines:
1→'No'
2 thru Highest→'Yes'
15. Click "Continue."
16. Click "OK." The new variables will be added. Depending on the file size, this can take from a few seconds to several minutes.
17. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.

You can now run a Crosstabs to see how many one-time versus repeat customers responded to the latest offer. If you want to keep your new variables for future use, remember to save the SPSS dataset!

Repeat customer? * Response Crosstabulation

			Response		Total
			no	yes	
Repeat customer?	No	Count	463	62	525
		% within Repeat customer?	88.2%	11.8%	100.0%
	Yes	Count	401	74	475
		% within Repeat customer?	84.4%	15.6%	100.0%
Total	Count		864	136	1000
	% within Repeat customer?		86.4%	13.6%	100.0%

Generating a decile or quintile variable

Some analyses require that the file be split into deciles (ten subsets of approximately equal size) or quintiles (five subsets of approximately equal size) according to the value of a variable. You can do this with the “Rank Cases” procedure.

Suppose we want to create a variable containing quintile assignments based on total dollar purchases:

1. Make sure you are in the Data Editor window..
2. From the pull-down menus, select **Transform / Rank Cases**.
3. Select “TOTDOL” for the variable list.
4. In the area labeled “Assign Rank 1 to,” select “Largest value.” This means that category 1 will be assigned to those customers with the highest dollar purchases – in other words, the biggest spenders.
5. Click on the “Rank types” button.
6. In the Rank Types dialog box, click on “Ntiles.” Type 5 in the box to have the file divided into quintiles.
7. If any other options are selected, click on the boxes to “un-select” them.
8. Click “Continue.”
9. Click “OK.” The operation can take from a few seconds to several minutes, depending on dataset size. A variable called NTOTDOL will be created, containing the quintile assignments (values 1-5).
10. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.

If you run a “Frequencies” procedure to count the number of records per quintile, you may find that the quintiles are not of exact equal size. In this example, they are very close:

Percentile Group of totdol					
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	200	20.0	20.0	20.0
	2	200	20.0	20.0	40.0
	3	196	19.6	19.6	59.6
	4	207	20.7	20.7	80.3
	5	197	19.7	19.7	100.0
	Total	1000	100.0	100.0	

Hints and Options:

1. In some cases SPSS may create only quartiles (4 categories) or “tri-tiles” (3 categories). This can happen if a large number of records in the dataset share the same value for the variable, or when there are very few values of the variable in the dataset. SPSS will not place records with the same values for a variable into separate N-tiles, which sometimes makes it impossible for it to create the number of categories you specified with approximately equal numbers of records.

VI. Selecting Records from a Dataset

Sometimes you want to do an analysis using only certain cases or observations in your dataset. Suppose, for example, that you want to restrict your analysis to those customers who live in the Northeast:

1. Be sure you are in the data editor window
2. From the pull-down menus, select **Data / Select Cases ...**
3. Select **If condition is satisfied**
4. Click on the **If** button
5. Enter the desired condition – in this case, region = 1 – then click on the **Continue** button
6. Finally click on the **OK** button to select the cases (by default the unselected cases will be 'filtered' (i.e. not used in the analysis), but not deleted).
7. Now, if I run **DESCRIPTIVES** on Numpurch and Totdol here is what I get.
8. Copy and paste the SPSS output and write one to two sentences (short interpretations/ explanations) of the results.

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Total Number of Purchases	227	1	10	1.98	1.453
Total Dollars Spent	227	50	8763	380.74	908.007
Valid N (listwise)	227				

9. Later, if you want to include all cases again, return to the **Select Cases** dialog box and select **All cases**.