

Psychology 2040 Laboratory 2

Comparing Distributions across Groups

Much of the work of statistics involves comparison of scores between two or more groups. For example, in order to demonstrate that a new method of therapy for depression is more effective than current methods, a clinical psychologist might give the old therapy to one group of patients and the new therapy to another group and then compare scores on a test of depression. Indeed, comparison of scores between groups is so commonplace in statistical methods, that it is very hard to find a statistical report without such a comparison. We will focus on comparison of groups using SPSS here.

There are two ways in which groups can be compared using SPSS. The first way splits the data file into as many pieces as there are groups. After splitting, the same statistics are computed for each group. This method is called the **Split File method**.

The second way uses SPSS procedures that are specifically designed to make comparisons of two or more groups. Since much of inferential statistics involves comparison of groups, many of the procedures for inferential statistics are constructed in this way.

I. . Using Split File to compare categorical distributions: Male and Female Job Categories

A. Open the **Employee Data.sav** data set.

File -> Open -> Data... -> (C:) -> Program Files -> IBM -> Statistics -> 20 -> Samples -> English -> Employee Data.sav

B. Open the Split File dialog box. Data -> Split File...

In this example, we'll split the file by Gender.

Check "Compare Groups"

Move "Gender" to the [Groups Based on:] box.

Run the Frequencies procedure and specify **jobcat** as the variable to be analyzed. Write a short description of the difference between the two distributions . . .

II. Using Split File to Compare Continuous Distributions: Male and Female Current Salaries

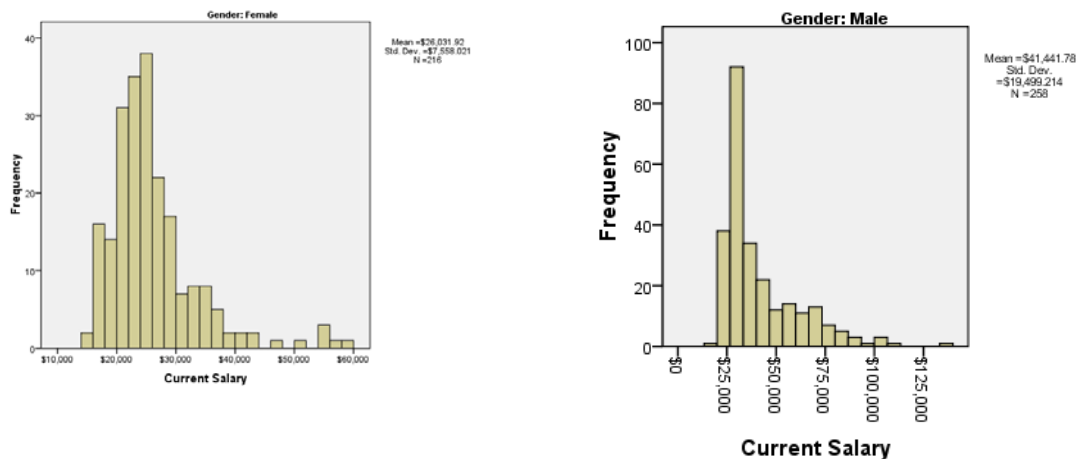
A. Comparison of histograms

1. Open the **Employee Data.sav** data set.
2. Use the Split File command to split the file by Gender. (If you're continuing from part I of this exercise, the Split File command is already in effect.) We'll use it to compare histograms of current salary of males and females
3. To create histograms, we'll use the Histograms procedure: **Graphs -> Histograms**....
4. Put Salary in the right field.

The Split File command causes two histograms to be printed - one for Males and one for Females. Note that the Histogram Scales are not the same, so it's hard to see which gender had higher salaries from a quick examination of the two histograms.

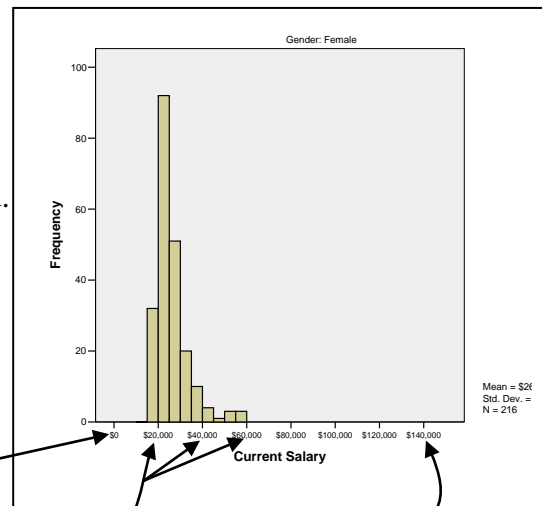
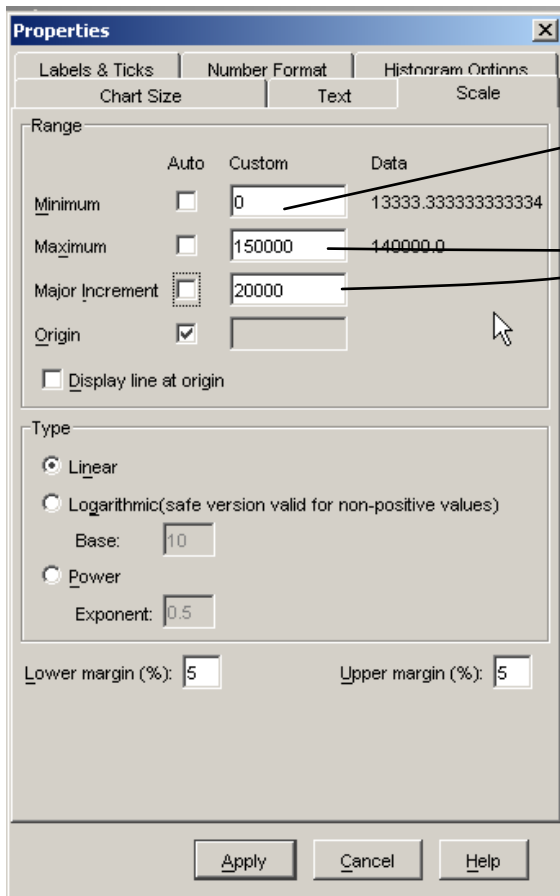
B. Making the histogram axes identical

The problem with the histograms we just created (shown below) is that they're hard to compare because the axes are not the same. We'll make the comparison more meaningful by putting identical values on the two axes.



1. First, make a mental note of the X-axis labels of each histogram. Note that the Female histogram ranges from 0 to \$60,000 while the Male histogram scale is from \$0 to about \$140,000. To make the histograms comparable, both will have to have the same scale.
2. Let's use \$0 and \$150,000 as the leftmost and rightmost scale labels. Those labels will contain both histograms, although, of course, there'll be some empty space in the Female graph.
3. We also have to choose the intermediate labels. Let's use a label every \$20,000.
4. Now double-click on one of the histograms.

5. Double-click on one of the values below the X-axis.
6. Click on the tab labeled “Scale”.
7. Uncheck the “Auto” boxes and enter 0 into the “Minimum” field. 150000 into the “Maximum” field. 20000 into the “Major Increment” field.
8. Click on “Apply”.



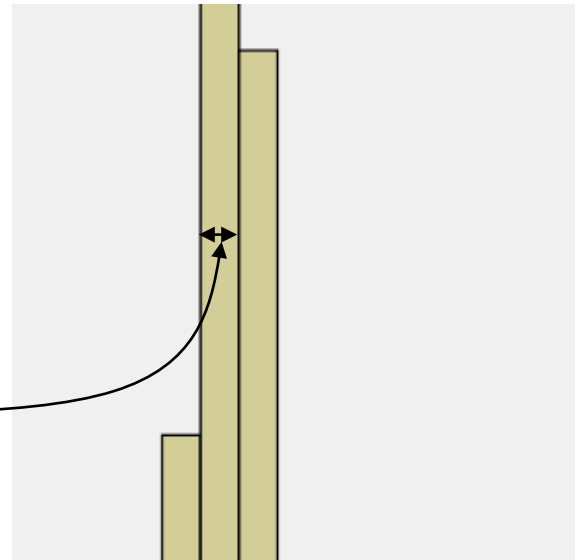
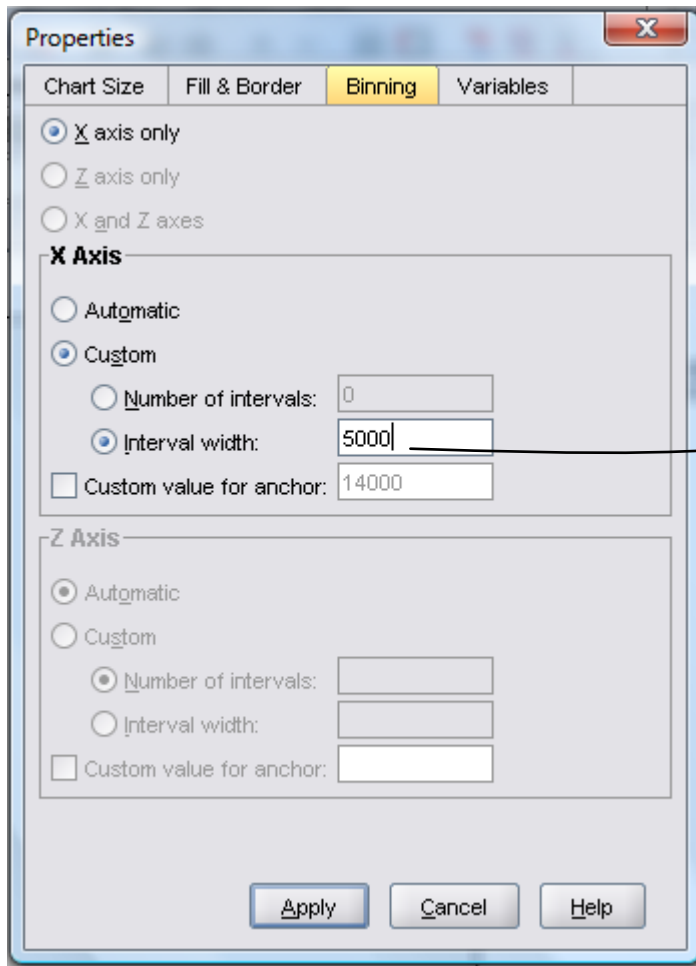
These specifications have taken care of the labeling of the axes.

But the **width of each column** must also be specified. Column width is called the Bin Size in SPSS.

To change the bin size.

1. Double click on the histogram.
2. Double-click on one of the columns.

3. Click on the “Binning” tab.
4. Click on the “Custom” circular check box.
5. Click on the “Interval Width” circular check box.
6. Finally, enter 5000 into the Interval Width field.



7. Click on “Apply” and then click on “Close” Note that the histogram has changed.

8. Now repeat all the above steps for the other histogram.

(Isn't this fun??? Why won't SPSS do all this for us?? I don't know.)

D .Describe the differences between salaries of males and females based on the identically scaled histograms.

III. Using Split File to Compare Histograms of Beginning salaries of Males and Female. (This is drill and practice for the procedures just covered.)

Carry out the above procedures for **beginning salaries** making identically formatted histograms.

Write a description of the difference between the two histograms.
Based on the examination of beginning salaries, what would you say is a logical choice as the reason for the differences in **current** salaries?

IV. Using the Split File method to compare TV hours across education levels.

A. Open the **GSS93 subset.sav** data file. (Recall – www.utc.edu/michael-biderman)

1. Find the variable named TVHours (Label=Hours per day watching TV).
2. Find the variable named Degree2: (Label=College Degree) to the right of TVHours.
3. Split the file by Degree2.
4. Compare histograms of TVHours between the groups.

There are four groups for DEGREE2- No, Yes, DK (for Don't Know) and NA (for Not Applicable). **Ignore the DK and NA groups.**

5. Make the histogram axes comparable for the Degree2 No and Yes groups.

Hint: Use **width=2** with **Minimum value=0** and **maximum value = 30**.

Write a description of the difference between the two histograms . . .

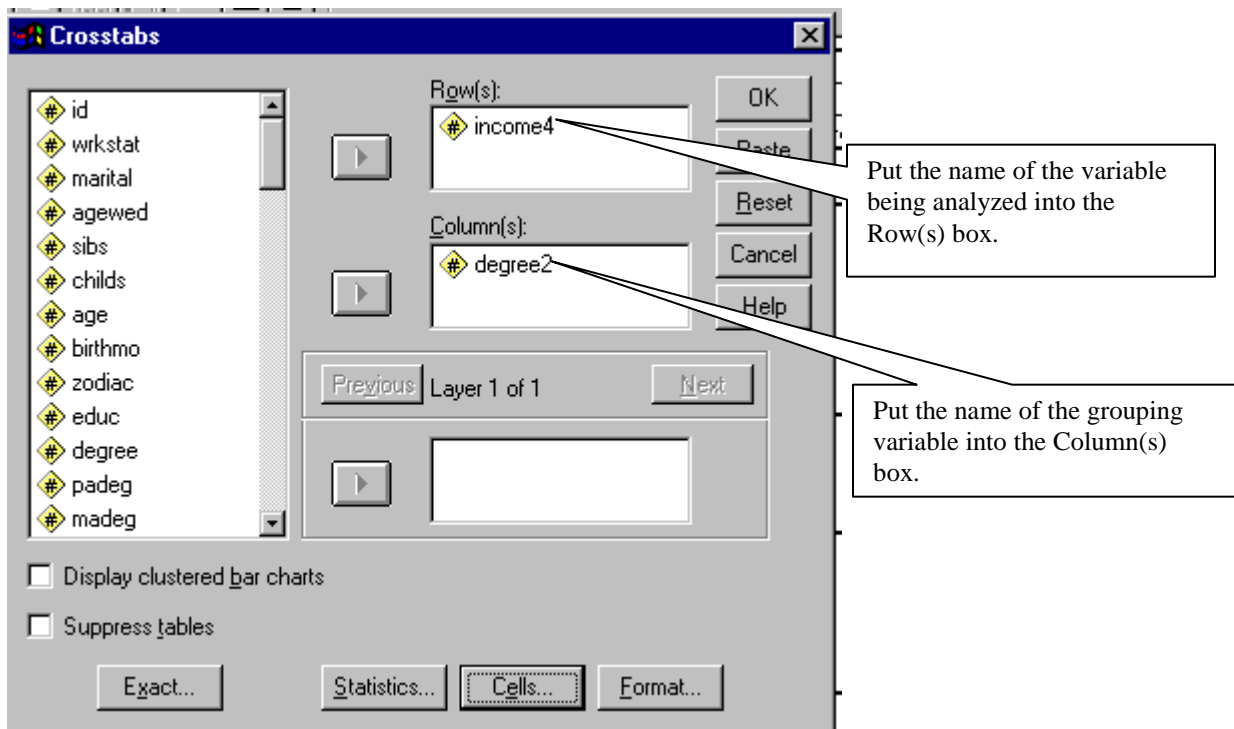
V. Comparisons of Groups using Procedures designed to Compare Groups.

Some procedures allow comparisons across groups without requiring Split File. Two of these are the **Explore** procedure and the **Crosstabs** procedure. Both are under the **Descriptives Statistics** option. **Crosstabs** is designed to compare categorical variables. **Explore** is best for comparison of continuous variables between groups.

A. Using Crosstabs for comparison of Income Category across Education Groups.

1. Use the **GSS93 Subset.sav** data file. Recall: www.utc.edu/michael-biderman

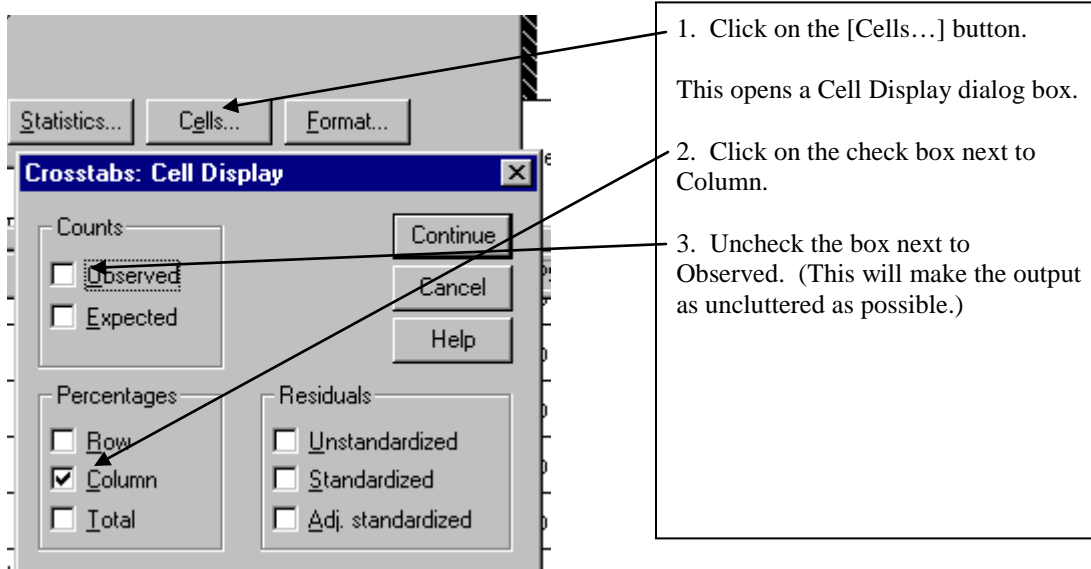
2. Analyze -> Descriptive Statistics -> Crosstabs



Write a description of the difference between the two columns. . . .

B. Refining the Crosstabs output.

The unrefined Crosstabs output displays only counts. Since there were many more respondents with less than college education than with college education, a comparison of counts is not as straightforward as it should be. Instead, it will be easier to compare the two groups by comparing percentages of persons in each group achieving the various income categories. This is easily done in Crosstabs.



Your table should look something like the following . . .

income4 Total Family Income * degree2 College Degree Crosstabulation

% within degree2 College Degree

		degree2 College Degree		Total
		0 No College degree	1 College degree	
income4 Total Family Income	1.00 24,999 or less	46.6%	13.8%	39.0%
	2.00 25,000 to 39,999	19.5%	21.9%	20.1%
	3.00 40,000 to 59,999	13.9%	20.2%	15.4%
	4.00 60,000 or more	19.9%	44.1%	25.5%
Total		100.0%	100.0%	100.0%

You should be able to summarize the difference in education between the two groups in a sentence or two.

C. Using Explore to compare continuous variables: Comparing Salaries.

The Explore procedure allows a variety of comparisons between groups. We'll use it extensively throughout the semester. Here we'll demonstrate how it can be used to create histograms for different groups.

We'll repeat the comparison of salaries of males and females from the Employee Data file.

1. Open the Employee Data file.
2. Analyze -> Descriptive Statistics -> Explore.

The image shows the SPSS Explore dialog box and its Plots sub-dialog box. The Explore dialog box has a list of variables on the left, including id, bdate, educ, jobcat, salbegin, jobtime, prevexp, and minority. The Dependent List contains 'salary' and the Factor List contains 'gender'. The Plots sub-dialog box is open, showing options for Boxplots (Factor levels together, Dependents together, None), Descriptive (Stem-and-leaf, Histogram), Normality plots with tests, and Spread vs. Level with Levene Test (None, Power estimation, Transformed, Untransformed). The Histogram checkbox is checked. Numbered callouts point to specific elements: 3. Put the variable being analyzed into the Dependent List box. 4. Put the variable defining the groups to be compared into the Factor List box. 5. Click on the [Plots...] button. 6. Check the Histogram box.

The Explore procedure automatically displays a collection of descriptive statistics for the whole sample and for each group. Don't worry if you haven't covered all of them yet.

Psychology 2040. Lab 2 Homework Assignment.

Note: Begin your answer to each problem on a new page.

1. Are there differences between **older and younger people in their** attitudes toward **spanking** children?

Use the GSS93.SAV file. (www.utc.edu/michael-biderman)

Split the file by the variable **AGECAT4** (Age Category).

Don't use AGE. You'll get 50 or more groups and won't be able to make any sense of the results.

Use the SPANKING (Favor Spanking to Discipline Child) variable as the dependent variable.

- A. Submit output of the Frequencies procedure including the mean and median.
- B. Submit a histogram of each group, with equal histogram labels.
- C. Write a 1-3 paragraph description of the results answering the question: Are there differences between older and younger people in their attitudes. If so, what are those differences?

2. Are there differences between **age categories** in preference for **big band music**?

Use the Crosstabs procedure. Put AGECAT4 into the Column(s) box. Put BIGBAND into the Row(s) box.

- A. Submit output of the Crosstabs procedure. Have Crosstabs print column percentages as shown on page 2-7 above.
- B. Write a 1-3 paragraph answer to the question: Are there differences between age categories in preference for big band music. If so, describe the nature of the differences..

3. In the Employee Data file, are there differences in **Jobtime** between males and females?

Use the **Explore** procedure.

- A. Submit the tables of descriptive statistics for males and females.
- B. Submit histograms for males and females. Make the histograms have identical axes.
- C. Write a 1-3 paragraph description answering the question: Are there differences in Jobtime between males and females? If so what is the nature of those differences?.

4. Is there a difference in **Politics** between those with a college degree and those without?

Use the GSS93 subset file.

Split the file by Degree2.

Run the Frequencies procedure with the variable named Politics as the dependent variable.

- A. Submit the output of the Frequencies procedure.
- B. Write a 1-3 paragraph answering the question: Is there a difference in Politics between those with a college degree and those without? If so, describe the difference.