



# Intro to SPSS

Fall 2012

# Intro to SPSS

---

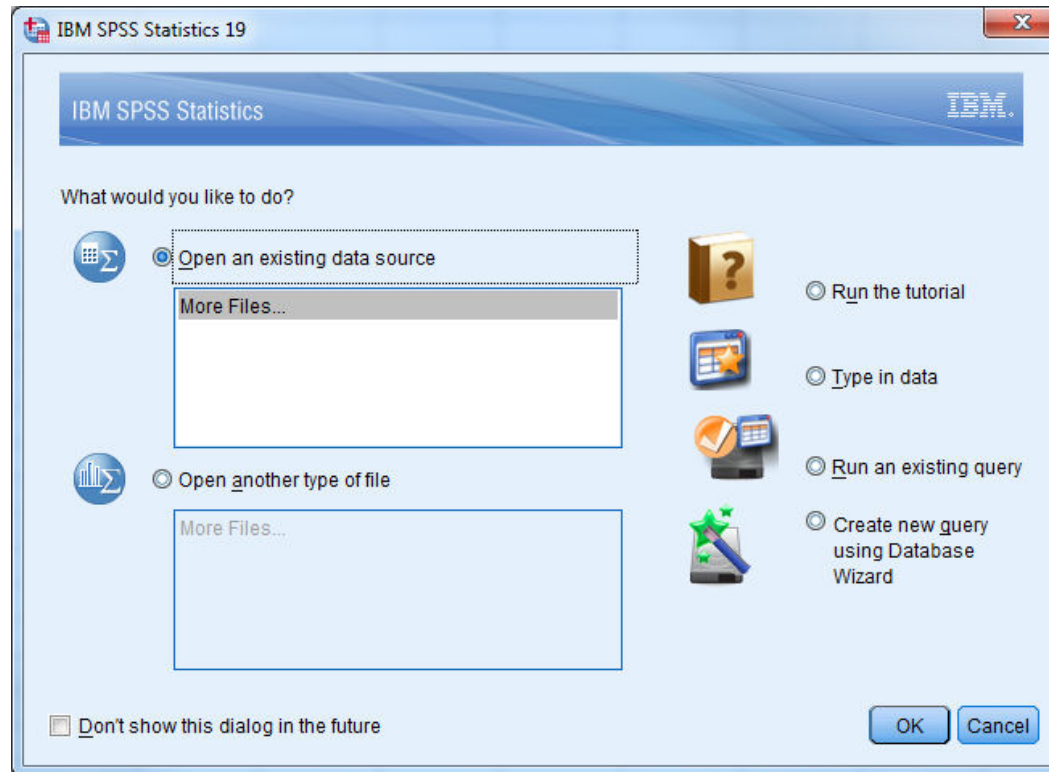
- SPSS is a statistical analysis program that allows:
  - Data management in a spreadsheet-like format
  - The ability to create graphs and tables
  - A broad range of statistical analyses – very detailed analyses
  - You will need to know or remember some basic statistics that I will go over for this section of the course
- SPSS is more *specialized* than Excel, therefore we need to provide data in a more precise way to take advantage of SPSS's functionality.

# SPSS

- 
- Goals for this section of the course include:
    - Becoming familiar with Statistical Packages and what they can do
    - Creating new Datasets
    - Importing & exporting Datasets
    - Manipulating data in a Dataset
    - Basic analysis of data (mainly descriptive statistics)
    - An overview of SPSS's advanced features
    - Examining the Help utility within SPSS

Note: This is not a statistics course such as Math 207. We will only concentrate on basic statistical concepts.

# Open SPSS



# Run the Tutorial

The screenshot shows a web browser window titled "Help - IBM SPSS Statistics". The address bar displays the URL: `127.0.0.1:60004/help/index.jsp?topic=/com.ibm.spss.statistics.tut/introtut2.htm`. The search bar contains the text "Search: GO Search scope: All topics".

The left sidebar, titled "Contents", lists the following categories and items:

- Help
- Tutorial
  - Introduction (highlighted)
  - Reading Data
  - Using the Data Editor
  - Working with Multiple Data Sources
  - Examining Summary Statistics for Individual Variables
  - Crosstabulation Tables
  - Creating and editing charts
  - Working with Output
  - Working with Syntax
  - Modifying Data Values
  - Time Saving Features
  - Customizing IBM SPSS Statistics
  - Automated Production
  - Scoring data with predictive models
  - Getting Started with Custom Tables
- Case Studies
- Statistics Coach
- Add-ons

The main content area is titled "Introduction" and contains the following text:

Introduction >  
**Introduction**

This tutorial will show you how to use many of the available features. It is designed to provide a step-by-step, hands-on guide. All of the files shown in the examples are installed with the tutorial so that you can follow along, performing the same analyses and obtaining the same results shown here.

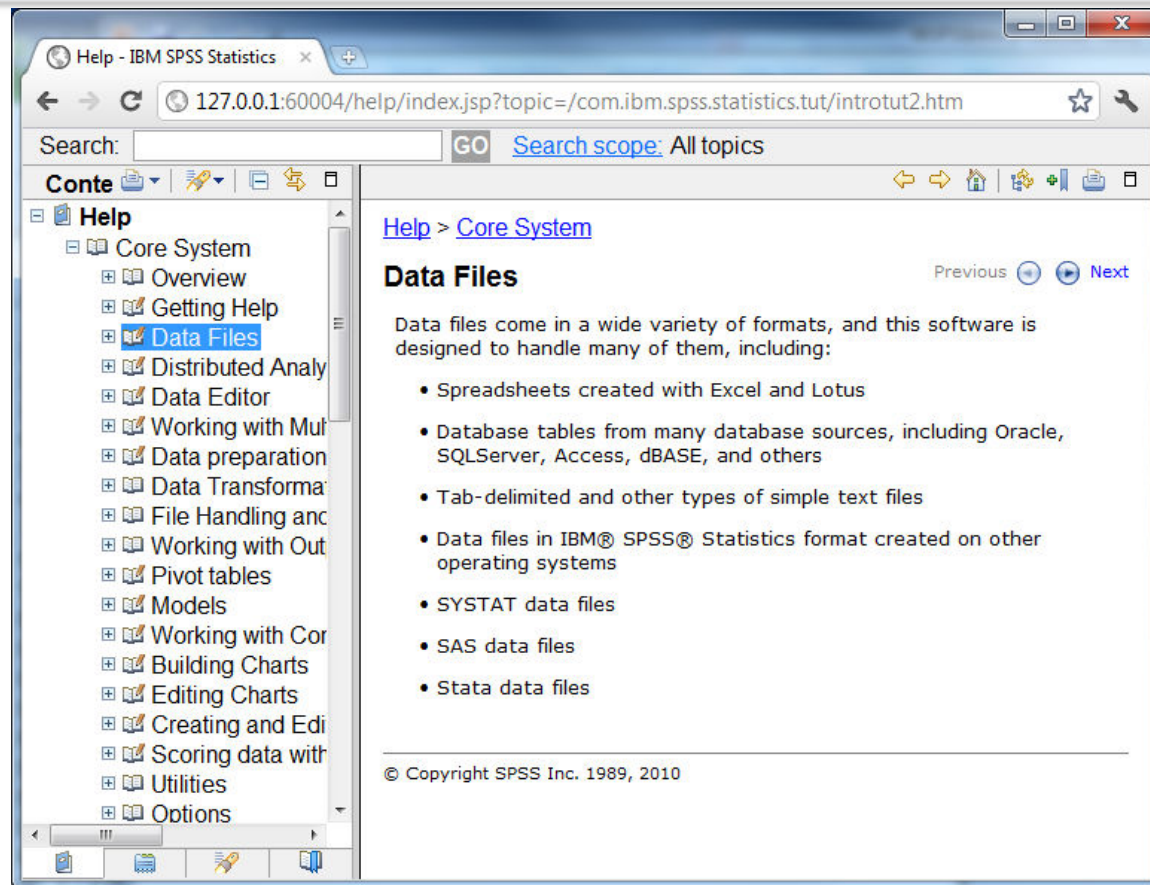
It is also designed to make it easy for you to start and stop anywhere you want. If you want to learn about only a few specific tasks, use the Contents and Index buttons (next to the Back and Next buttons) at the bottom of this window to find the information you need, or simply click the Next button (arrow pointing right) to step through all of the tutorials from beginning to end.

If you want detailed examples of various statistical analysis techniques, try the step-by-step [Case Studies](#).

© Copyright SPSS Inc. 1989, 2010

At the bottom of the page, the word "Introduction" is displayed in a large, light blue font, and a play button icon is visible in the bottom right corner.

# SPSS Help



[Help > Core System](#)

## Data Files

Previous [Next](#)

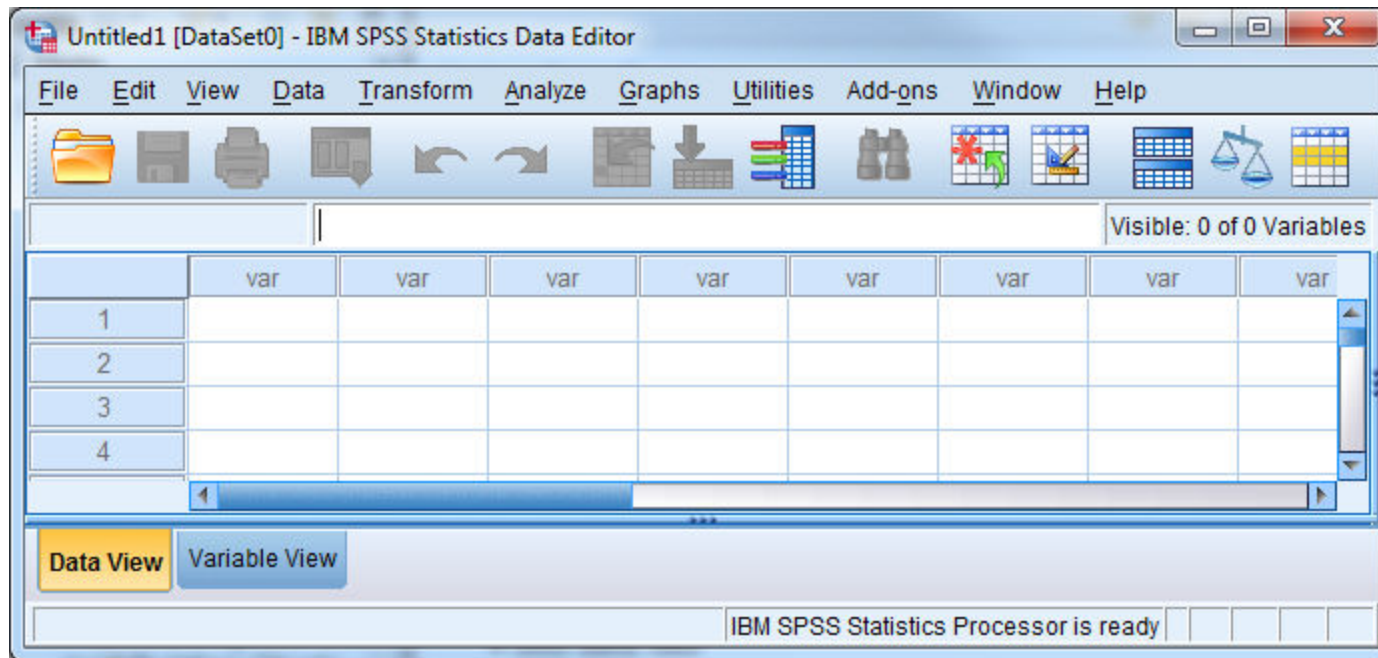
Data files come in a wide variety of formats, and this software is designed to handle many of them, including:

- Spreadsheets created with Excel and Lotus
- Database tables from many database sources, including Oracle, SQLServer, Access, dBASE, and others
- Tab-delimited and other types of simple text files
- Data files in IBM® SPSS® Statistics format created on other operating systems
- SYSTAT data files
- SAS data files
- Stata data files

© Copyright SPSS Inc. 1989, 2010

# Create a Simple Dataset

- SPSS looks somewhat like Excel BUT there are several important differences
- Select the Data View tab



# Excel versus SPSS Differences

---

- Column data pertains to a particular variable

List several examples of what a variable might be

- Row data is considered a case or an observation

List several examples of an observations

- A cell contains a value for a particular variable that is part of a part of a particular observation



# SPSS Views

---

- Data View – displays the actual values of the data set
- Variable View – contains the descriptions of each variable's attribute in the data file

List at least three attributes of a variable from the Variable View

# Dog Dataset Example

---

<b>Breed</b>	<b>Age</b>	<b>Weight</b>
Collie	2	23
Collie	3	35
Setter	5	45
Shepard	1	65
Setter	2	72

# Dataset Questions

---

- Using the SPSS Tutorial, SPSS Help, or Web define each of the following terms and give a real life example of each. SPSS contains the following data types (**measures**):
  - Categorical
    - Nominal
    - Ordinal
  - Scale

# Dataset Questions

---

- Using the SPSS, SPSS Tutorial, and SPSS Help:
  - What are the **types** available in the Variable View?

# SPSS

---

- We can build the Dog data sheet together
  - Variable View

– Enter Data

<b>Data</b>	<b>Type</b>	<b>Measure</b>
Breed		
Age		
Weight		

# Candy Dataset Example

---

<b>Brand</b>	<b>Name</b>	<b>ServingPerPkg</b>	<b>OzPerPkg</b>	<b>Calories</b>	<b>TotalFatInGrams</b>	<b>SatFatInGrams</b>
<b>M&amp;M/Mars</b>	<b>Snickers Peanut Butter</b>	<b>1.0</b>	<b>2.00</b>	<b>310</b>	<b>20</b>	<b>7.0</b>
<b>Hershey</b>	<b>Cookies 'n Mint</b>	<b>1.0</b>	<b>1.55</b>	<b>230</b>	<b>12.0</b>	<b>6.0</b>
<b>Hershey</b>	<b>Cadbury Dairy Milk</b>	<b>3.5</b>	<b>5.00</b>	<b>220</b>	<b>12.0</b>	<b>8.0</b>
<b>M&amp;M/Mars</b>	<b>Snickers</b>	<b>3.0</b>	<b>3.70</b>	<b>170</b>	<b>8.0</b>	<b>3.0</b>
<b>Charms</b>	<b>Sugar Daddy</b>	<b>1.0</b>	<b>1.70</b>	<b>200</b>	<b>2.5</b>	<b>2.5</b>

# More Dataset Questions

---

- For the given dataset, what is the type and measure for the data for each of the variables?  
Why?
  - Brand
  - Name
  - ServingPerPkg
  - OzPerPkg
  - Calories
  - TotalFatInGrams
  - SatFatInGrams

# Problem 8.1

Setup the Variable information

Input the data by hand

---

Create the dataset Candy8.1 from the Candy dataset using SPSS 19.0

1. Create the variables using the Variable View. Make sure that each variable has the correct **Type** and **Measure**.
2. Set the decimals column as follows: Brand: 0, Name: 0, ServingPerPkg: 1, OzPerPkg: 2, Calories: 0, TotalFatInGrams: 1, and SatFatInGrams: 1.
3. In the Values column, create the Value Labels for Brand where 1 = "M&M/Mars", 2 = "Hershey", and 3 = "Charms".
4. Change to Data View and enter the candy data. You will need to go back to Variable View and edit some of the settings. Do so as necessary.



# Types of Data Analysis

---

- When doing data analysis, we are interested in two types of summaries:
  - Statistical Summaries (e.g. descriptive, hypothesis testing)
  - Visual Summaries (e.g. tables, graphs)

# Areas of Statistics

---

- **Descriptive Statistics** - a branch of statistics dealing with description and summarization of collections of data
- **Inferential Statistics** - where inferences are made from samples of the population (e.g. smokers smoking a pack of cigarettes per day have a higher cholesterol). In this area we get into **Hypothesis testing**.

# Descriptive Statistics

---

- We are concerned, among other things, the following:
  - Mean
  - Median
  - Mode

# Problem 8.1 Continued

---

- Identify each of the following for Total Fat giving your answer to 1 decimal place:
  - Minimum:
  - Maximum:
  - Mean:
  - Std Deviation:

# Problem 8.2

---

A paint manufacturer tested two experimental brands of paint over a period of months to determine how long they would last without fading. Here are the results:

Brand A	Brand B	Report on the following
10	25	-Mean
20	35	-Median
60	40	-Mode
40	45	-Std Deviation
50	35	-Minimum
30	30	-Maximum

# Solution - Method 1

---

One way has two variable columns where the first is BrandA and the second is BrandB. Enter the above data and find the asked for information. Save this file as BrandMethod1.sav.

What are the type and measure values for:

BrandA \_\_\_\_\_ and

BrandB \_\_\_\_\_

# Solution – Method 2

---

The second way has two columns where the first column is a variable called Brand and the second column is called Fading. Create value labels where 1="BrandA" and 2="BrandB". Enter the information and find the asked for information. Save this file as BrandMethod2.sav.

What are the type and measure values for  
Brand \_\_\_\_\_ and  
Fading \_\_\_\_\_

What do the descriptive statistics tell us about the paint with regard to fading?

# Common Display of Statistical Info

---

- Two common ways to graphically display statistical information is through the use of bar charts and pie charts.
- A bar chart graphically displays a bar graph where the lengths of the bars are proportional to the values that they represent.



# Sample CS120 Dataset

---

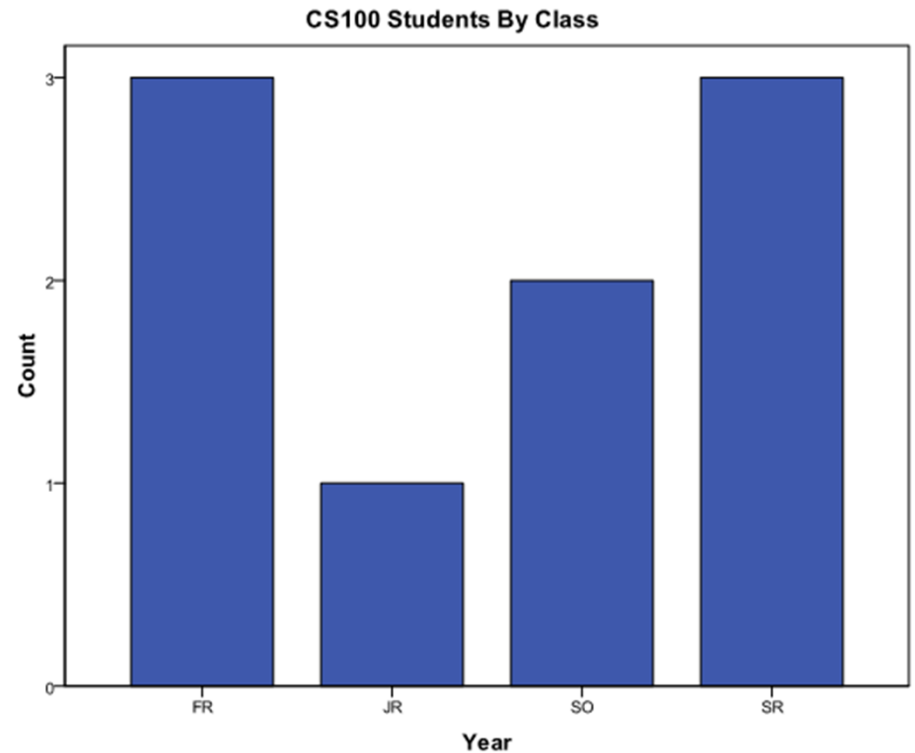
- Consider the following CS120 class information:

<b>ID</b>	<b>Year</b>	<b>Age</b>
0001	FR	18
0002	FR	18
0003	SR	22
0004	JR	22
0005	SO	19
0006	FR	19
0007	SR	23
0008	SO	19
0009	SR	22

If this data was in SPSS,  
what would be the Type  
and Measure for each  
variable?

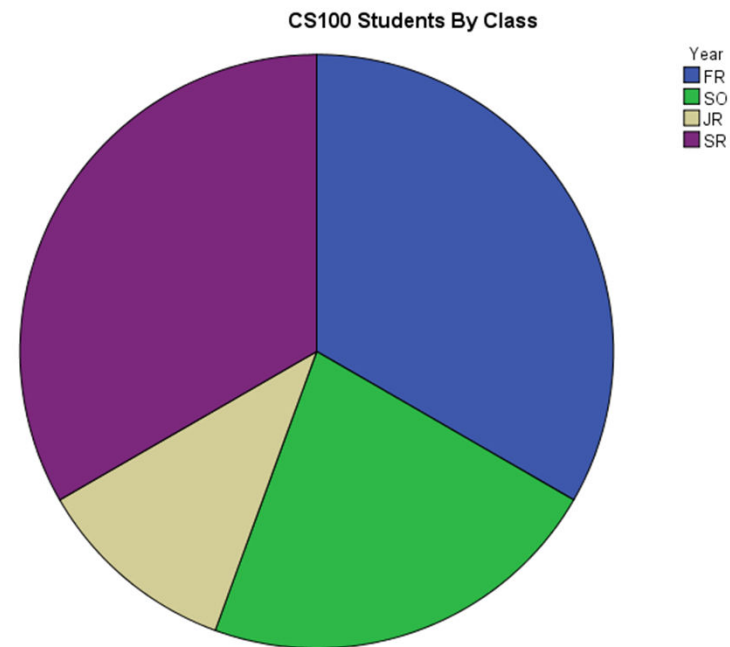
# Bar Chart

We could take the above information and show using a bar chart a graphical representation of the number of students that are FR, SO, JR, and SR.



# Pie Chart

- Notice with a pie chart we get a better visualization of the frequency of occurrence as a percent. The amount of arc in the above example is proportional to the represented quantity.



# Exercise

Copy CS120.sav from CS 130 Public to your Desktop.

---

- Let's discuss the variable definitions and data.
- Create a **Bar Chart** to show how many students in each Year took CS 120.
- Create a **Pie Chart** to show how many students in each Year took CS 120.
- Create a **Bar Chart** to show the **median age** of students in each Year.
- Create a Word document called **graphs.docx** that has all three graphs in the document properly labeled and looking professional. Let me see the results.
- Enter this data into Excel and build the three charts above.

# More SPSS practice

---

- Compare the mean, standard deviation, and median for age by year
- Add the Report to the Word document from earlier

		Report			
Age	Year	Mean	N	Std. Deviation	Median
	FR	18.33	3	.577	18.00
	SO	19.00	2	.000	19.00
	JR	22.00	1	.	22.00
	SR	22.33	3	.577	22.00
	Total	20.22	9	1.986	19.00