Welcome to Statistics 133

Statistics for the Business Sciences

Agenda

- Goals of the course
- Course syllabus
- Section 1.1
  - Displaying distributions with graphs

1.1 Displaying Data with Graphs

- Purpose of a data display
  - Presents an easy to ready summary of the data
  - Allows you to see and point out important features in the data
- Data displays should always precede any data analysis
  - Checks conditions needed for a data analysis
  - Helps guide you to the right analyses

Important Terms

- Variable
- Quantitative Variable
  - Special case: Time (Time Series Data)
- Categorical Variable
- Distribution

Data Displays for Categorical Data

- Data examples: gender, opinion, region, etc.
- Types of displays
  - Pie Charts
  - Bar Graphs

Important notes:

- Pie charts show percents (relative frequencies) and sum to 100%
- Bar graphs show counts (frequencies) in each category or percentage (relative frequency) in each category

Example: Gender of Stat 133 Students

- Data:
  - 285 men in Stat 133
  - 215 women in Stat 133
- Total number of students: 500
- What is the variable?
- Why is it categorical?
- Graph the distribution of the data and interpret (explain in words a non-statistician would understand)

Example: Gender of Stat 133 Students

- Distribution of the data:
- Interpret:

Data Displays for Quantitative Data

- Data examples: Unemployment rates, Stock prices, length of time for a service call
- Type of displays: Histogram
  - Similar to a bar graph except it uses a number line and the bars connect to each other in numerical order

What is a histogram?

- A graph that shows the distribution of a quantitative data set.
- Puts the data into numerical groupings
- Finds out how many data values lie in each grouping (# or % in each grouping)
- Makes a graph showing the numerical groupings on the X axis and the number (or %) of data values on the Y axis.

U. Rates Example

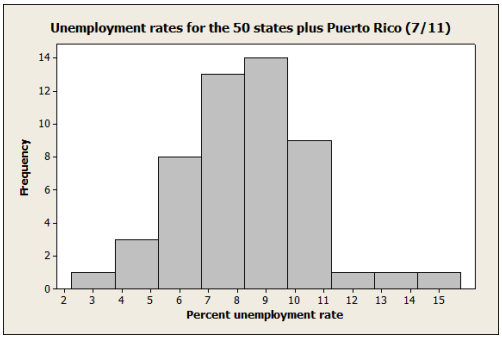
- Data: Unemployment Rates for all 50 states plus Puerto Rico (7/2011)

  5.8, 7.6, 8.5, 9.3, 6.1, 5.7, 7.5, 8.4, 9.8, 4.7, 10.9, 10.8, 7.8, 9.5, 5.5, 9.0, 3.3, 10.1, 8.0, 6.7, 9.5, 5.2, 12.9, 4.1, 7.7, 8.7, 10.4, 7.2, 10.9, 7.6, 7.2, 7.7, 7.6, 9.5, 6.5, 6.0, 8.5, 9.5, 9.4, 6.1, 10.1, 10.7, 8.1, 9.1, 8.5, 12.0, 8.2, 9.4, 7.7, 10.0

Terms

- Variable:
- Type of variable:
- How to show the distribution?

Unemployment Rates #1



Understanding the histogram

- X axis: what are the values?
- Y axis: how many values are there in each group?
- Groupings:
- Bars:
- Overall information:

Describing the data from a histogram

- Shape of the distribution of the data
- Center of the distribution of the data
- Variability in the distribtion of the data

The Shape of the Distribution:

– Symmetric:
  - Means:

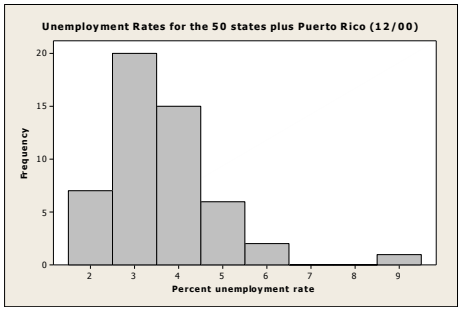
– Skewed right:
  - Means:

– Skewed left:
  - Means

The Shape of the Distribution

- Flat (special symmetric distribution):
  – Means:
- Many other shapes!
- What's the shape of the 2011 U. Rate data?

Unemployment Rates #2

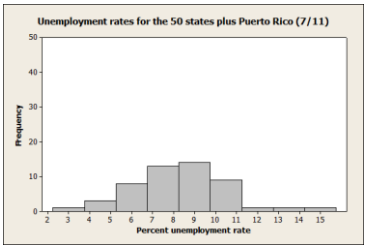


Interpret the 2011 histogram

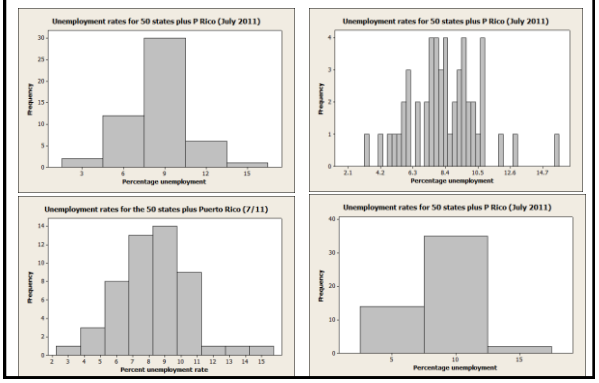
Compare U. Rates for 2011 to U. Rates for 2000

Note 1: Scale can affect histogram shape

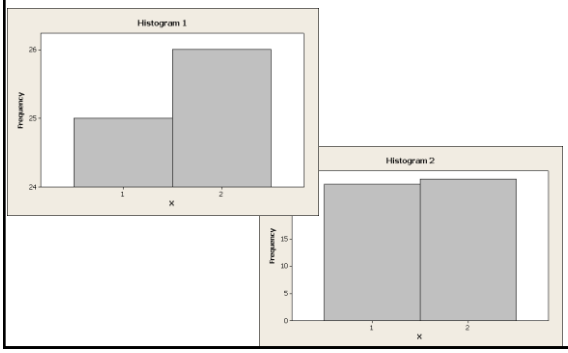
- Through use of scale (Y-axis) they can be made to look different.



Note 2: # Bars can affect histogram shape



Note #3: Starting point can affect histogram shape

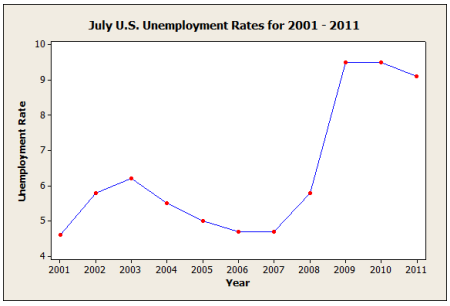


Variability in a Histogram

- 3, 3, 3, 3, 3

- 1, 2, 3, 4, 5

Data Displays for Time Series Data: Time Series Plots



Important Features of Time Series Data

- Pattern
- Variability

Patterns in Time Series Data

- Cycles:
- Trends:

Patterns in Time Series Data

- Flat (special symmetric distribution):
  - Means:
- Many other shapes!
- What's the pattern of the U. Rates 2000-2011 data?

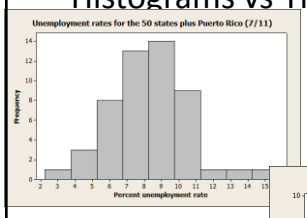
Variability in Time Series Data

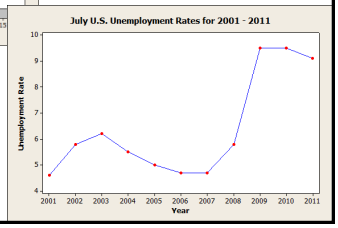
- Variability = Volatility
- 3, 3, 3, 3, 3
- 1, 2, 3, 4, 5
- 1, 2, 3, 2, 1

Important notes about Time Series Plots

- Always put time on the X axis
- Always put your quantitative variable of interest on the Y axis (price, population size, sales, etc.)
- Watch the scale on the Y axis
- Watch for time gaps in the X axis

Histograms vs Time Series Plots





Compare the formats

Histogram

- Data set:
- Each bar on the graph represents:

Time Series Plot

- Data set:
- Each dot on the graph represents:

Compare the axes

Histogram

- X axis:
  - Represents:
  - Units:
- Y axis:
  - Represents:
  - Units:

Time Series Plot

-
- X axis:
  - Represents:
  - Units:
- Y-axis:
  - Represents:
  - Units:

Compare variability

Histogram

HIGH VARIABILITY LOW VARIABILITY

Compare variability

Time Series Plots

HIGH VARIABILITY LOW VARIABILITY

Discuss U. Rates using these two graphs

Organizing data

- Identify type of data
- Determine appropriate graphs
- Make the graphs (computer)
- Understand the graphs
- Interpret the data using the graphs
- Watch for deceptive graphs