Introduction to Statistics

Chapter 4: Understanding and Comparing Distributions

Date: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Objectives:

## Students will be able to compare groups and look for patterns among groups over time.

## Students will understand that using histograms and stem-and-leaf plots are useful for comparing two distributions, but boxplots are more useful for comparing several groups.

## Students will understand the value of identifying outliers.

|  |  |
| --- | --- |
| **Main Idea** | **Notes** |
| **The Big Picture** | We can answer interesting questions about variables when we compare distributions  for \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  Here is a histogram of the *Average Wind Speed* for each day in 1989.  Picture 1   1. Where is the mean? 2. Describe the distribution of this data. (CUSS!) |
| **Windspeed: Making Boxplots** | Now here is the histogram & boxplot for daily wind speeds:  Picture 2  Explain what you see here… |
| **Comparing Groups with Histograms** | Sometimes it is more interesting to compare groups (here: spring/summer & fall/winter).  With histograms, we can note the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_, \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_,  and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ of the two distributions.  Figure5  Compare the two distributions (CUSS ***comparatively***). |
| **Comparing Groups with Boxplots** | Boxplots offer a nice balance of information and simplicity, displaying  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  We can plot them \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ for groups or  categories that we wish to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (in this case, by month).  Now, let’s say we are curious about the following questions: Are some months windier than others? Are wind speeds equally variable from month to month? Do some months show more variation?  Figure5   1. What do you see here? 2. Why do we now see outliers? 3. When do wind speeds tend to decrease? 4. When are winds the strongest and most variable? 5. Why do you think there could have been an outlier in November? |
| **What about Outliers?** | If there are clear outliers and you want to report the mean and standard deviation,  report them with \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and  with \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. The  differences may be quite revealing. |
| **Time Plots** | For some data sets, we are interested in how the data behave \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.  In these cases, it may be helpful to construct \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  of the data. Here we are looking at calendar ‘year days’, starting on day 1.  Figure5  What do you see here? |
| **Classwork:** | Comparing Groups Worksheet |
| **Re-expressing Skewed Data**  **Re-expressing Skewed Data**  **(Continued)** | When data are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ it can  be hard to summarize them using \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_,  and hard to decide whether extreme values are \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ or just  part of a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. How can we say  anything useful about such data?  Look at the display below:  Figure5  What do you see?  One way to make a skewed distribution more symmetric is to \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (transform) the data by applying some \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (example: logarithm).  Figure5   1. Why is this re-expressed histogram better than the last one? 2. We can now see that a typical log compensation is between about 6.5 and 7. What does this mean? |
| **Some Log Examples (so you know how to work with logs):** | Solve for the variable. |
| **Classwork/**  **Homework** | Chapter 4 Review for the Quiz  Study for the Chapter 4 Quiz (Next Class) |