Lecture 26
Summarizing Data

*STAT 225 Introduction to Probability Models*
April 18, 2014
Agenda

1. Summarizing Qualitative Data
2. Summarizing Quantitative Data
Summarizing Qualitative Data

- **Frequency Distribution**: the frequency or count of the occurrences within a particular category

Typically the above 3 distributions are summarized in table form and they can also be represented by a *bar graph* or *pie chart*.
Summarizing Qualitative Data

- **Frequency Distribution**: the frequency or count of the occurrences within a particular category
- **Relative/Percent Frequency Distribution**: the proportion/percentage of the occurrences within a particular category

Typically the above 3 distributions are summarized in table form and they can also be represented by a bar graph or pie chart
Summarizing Quantitative Data

- Use relative/percent frequency, cumulative distributions to summarize quantitative data
- Use dot plots, histograms, or stem and leaf plots to visualize quantitative data
Histograms and Skewness

- Symmetric Bell shaped
- Skewed to the Left
- Skewed to the Right
Histograms and Skewness cont’d

Differences between histograms and bar graphs:
- Histograms have a specified order on the $x$– axis

The Effect of Skew on the Mean and Median

- Skewed to the left $\Rightarrow$ the mean is less than the median
- Skewed to the right $\Rightarrow$ the mean is greater than the median
- Symmetric $\Rightarrow$ the mean is about the same as the median
Histograms and Skewness cont’d

Differences between histograms and bar graphs:
- Histograms have a specified order on the $x$– axis
- Histograms are for quantitative data only

The Effect of Skew on the Mean and Median

Skewed to the left $\Rightarrow$ the mean is less than the median
Skewed to the right $\Rightarrow$ the mean is greater than the median
Symmetric $\Rightarrow$ the mean is about the same as the median
Histograms and Skewness cont’d

Differences between histograms and bar graphs:
- Histograms have a specified order on the $x$-axis
- Histograms are for quantitative data only

The Effect of Skew on the Mean and Median
- Skewed to the left $\Rightarrow$ the mean is less than the median
Histograms and Skewness cont’d

Differences between histograms and bar graphs:
- Histograms have a specified order on the $x$-axis
- Histograms are for quantitative data only

The Effect of Skew on the Mean and Median
- Skewed to the left $\Rightarrow$ the mean is less than the median
- Skewed to the right $\Rightarrow$ the mean is greater than the median
Histograms and Skewness cont’d

Differences between histograms and bar graphs:
- Histograms have a specified order on the $x$– axis
- Histograms are for quantitative data only

The Effect of Skew on the Mean and Median
- Skewed to the left $\Rightarrow$ the mean is less than the median
- Skewed to the right $\Rightarrow$ the mean is greater than the median
- Symmetric $\Rightarrow$ the mean is about the same as the median
Line Graphs, Stem and Leaf Plots

- **Line graphs**: used to summarize time series data. A typical line graph has time on the $x$-axis and the variable on the $y$-axis.
Line Graphs, Stem and Leaf Plots

- **Line graphs**: used to summarize *time series data*. A typical line graph has time on the *x*-axis and the variable on the *y*-axis.

- **Stem and leaf plots**: a technique that orders quantitative data points and provides insight about the shape of the distribution. To make a stem-and-leaf plot, the last digit of the number is the leaf and the rest of the number is the stem. Additionally, any stem that is not used, but is within the range of the data, is kept in the plot.
Line Graph

Value of Sarah's Car

Year

Value in $

2001 2002 2003 2004 2005 2006 2007
Stem and Leaf Plots

Data: 24, 10, 13, 2, 28, 34, 65, 67, 55, 34, 25, 59, 8, 39, 61. First, arrange this data in an increasing order 2, 6, 10, 13, 24, 25, 28, 34, 34, 39, 55, 59, 61, 65, 67. Then we will use 0, 1, 2, 3, 4, 5, and 6 as stems. The plot is displayed below:

<table>
<thead>
<tr>
<th>Stems</th>
<th>Leaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2 6</td>
</tr>
<tr>
<td>1</td>
<td>0 3</td>
</tr>
<tr>
<td>2</td>
<td>4 5 8</td>
</tr>
<tr>
<td>3</td>
<td>4 4 9</td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5 9</td>
</tr>
<tr>
<td>6</td>
<td>1 5 7</td>
</tr>
</tbody>
</table>

Means 25

Means 55
Example 64

Suppose our data set is the numbers 1, 3, 5, 7, 12, 15, 17, 19, 21, 21, 21, 30, 33, 39, and 56. Create a stem-and-leaf plot of the data.

Solution.