STOR 155 Introductory Statistics

Lecture 2: Displaying Distributions with Graphs: HG & Time Plots
Recall

• Data:
  – Individuals
  – Variables
    • Categorical variables
    • Quantitative variables

• Distribution of variables

• Graphical tools for categorical data
  – Bar graph
  – Pie chart

• Graphical tools for quantitative data
  – Stemplot
Example: A study on litter size

- **Data**: (170 observations)

```
4 6 5 6 7 3 6 4 4 6 4 4 9 5 10 6 6 5 6
8 2 7 7 9 3 7 5 7 7 4 5 5 6 7 6 7 8
6 6 7 6 6 7 5 4 5 6 6 1 3 4 7 5 4 7 5
8 8 5 6 8 5 5 4 9 6 7 3 7 7 5 4 6 9 6
7 7 5 7 3 7 6 5 3 7 10 5 6 8 7 5 5 7 5
5 8 9 7 5 7 5 5 5 6 3 7 8 7 7 6 3 4 4
4 7 2 7 8 5 8 6 6 5 6 4 7 5 5 6 9 3 5
4 8 3 9 8 3 6 5 4 7 8 4 8 6 8 5 6 4 3
8 8 6 9 5 5 6 6 7 6 8 6 11 6 5 6 6 3
```
Stem-and-leaf plot for pups

0|1223333333333333344... (35)
0|555555555555555555555... (132)
1| 001
Histogram

• Partition the range of values of a quantitative variable into intervals and display only the count or percent of the observations that fall into each interval.
• You can choose any convenient number of intervals.
• Intervals must be of equal width (except at the two ends ?)
Relative frequency HG for the study on litter size
Data analysis in action: show steps in doing HG ...

### Table 1.1

Service times (seconds) for calls to a customer service center

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>77</td>
<td>289</td>
<td>128</td>
<td>59</td>
<td>19</td>
<td>148</td>
<td>157</td>
<td>203</td>
</tr>
<tr>
<td>126</td>
<td>118</td>
<td>104</td>
<td>141</td>
<td>290</td>
<td>48</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>372</td>
<td>140</td>
<td>438</td>
<td>56</td>
<td>44</td>
<td>274</td>
<td>479</td>
<td>211</td>
</tr>
<tr>
<td>179</td>
<td>1</td>
<td>68</td>
<td>386</td>
<td>2631</td>
<td>90</td>
<td>30</td>
<td>57</td>
</tr>
<tr>
<td>89</td>
<td>116</td>
<td>225</td>
<td>700</td>
<td>40</td>
<td>73</td>
<td>75</td>
<td>51</td>
</tr>
<tr>
<td>148</td>
<td>9</td>
<td>115</td>
<td>19</td>
<td>76</td>
<td>138</td>
<td>178</td>
<td>76</td>
</tr>
<tr>
<td>67</td>
<td>102</td>
<td>35</td>
<td>80</td>
<td>143</td>
<td>951</td>
<td>106</td>
<td>55</td>
</tr>
<tr>
<td>4</td>
<td>54</td>
<td>137</td>
<td>367</td>
<td>277</td>
<td>201</td>
<td>52</td>
<td>9</td>
</tr>
<tr>
<td>700</td>
<td>182</td>
<td>73</td>
<td>199</td>
<td>325</td>
<td>75</td>
<td>103</td>
<td>64</td>
</tr>
<tr>
<td>121</td>
<td>11</td>
<td>9</td>
<td>88</td>
<td>1148</td>
<td>2</td>
<td>465</td>
<td>25</td>
</tr>
</tbody>
</table>

Table 1-1

*Introduction to the Practice of Statistics, Fifth Edition*

© 2005 W.H. Freeman and Company
Data analysis in action: count (frequency HG)
Example: Call Center Data

- Financial firm call center
- Calls handled by Avi within 60 seconds
  - October: 666
  - December: 523
October

Histogram

Frequency

calling time

Frequency

5/12/10 Lecture 2
December

Histogram

Frequency

calling time

0 20 40 60 80 100 120

6 12 18 24 30 36 42 48 54 60

Frequency
Notes for Making Histogram

• Choose the number of classes sensibly (Fig 1.4, 1.8).

• Intervals must be of equal width.
• Areas of the bars are proportional to the frequency.
Examining Distributions

• Overall Pattern
  – Shape
  – Center (numerical, Lecture 3)
    • midpoint
  – Spread (numerical, Lecture 3)
    • range

• Deviations
  – Outliers: some values that fall outside the overall pattern.
Shapes of Distributions

• Graphs can help to determine shapes.
  – Modes: local peaks of a distribution.
    • Unimodal: one peak
    • Bimodal: two peaks
  – Symmetric or skewed?

Q: What detailed information is lost in a HG? What highlighted?
Shakespeare’s Words: Uni-modal
Tuition and fees: bimodal or trimodal
A bimodal histogram

A modal class

A modal class
Skewness

Right skewed

Left skewed
Iowa Test of Basic Skills vocabulary scores
A study on litter size
Bell-shaped Histograms
Summary: Shapes of Distributions

• **Symmetric**:  
  – histogram in which the right half is a mirror image of the left half.

• **Skewed to the right**:  
  – histogram in which the right tail is more stretched out than the left.(long tail to the right)

• **Skewed to the left**:  
  – histogram the left tail is more stretched out than the right.(long tail to the left)

• **Number of modal classes**:  
  – the number of distinct peaks in a histogram

• **Bell-shaped**:  
  – A histogram looks like a bell.
Time plots

- A time plot of a variable plots each obs against the time at which it was measured.
  - Time: x-axis
  - Variable: y-axis
  - Examples: stock price, unemployment rate, daily temperature
  - Great for identifying changing patterns over time.

- What to look for
  - Trend
  - Seasonal variations
  - Major deviations
Example: Number of Suicides in USA (1900-1970)
Call Center: Daily Call Volume in Sep. 2002

Time Plot of # of Calls for Agent By Date (in September)
Outliers

• Observations that lie outside the overall pattern of a distribution.

• Possible reasons:
  – error in data entry (most likely reason)
    • Equipment failure
    • Human error
    • Missing value code
  – extraordinary individuals (Jordan’s salary)
Handling Outliers

• Detect it using graphical and numerical methods.
• Check the data to make sure correct entry.
• Reducing influence of outlier
  – delete the observation (BE CAREFUL!)
  – Use transformations, robust methods.
Call Center: Daily Call Volume in 9/2002 (seasonality?)

Time Plot of # of Calls for Agent By Date (in September)
Take Home Message

• Examine distributions:
  – Overall pattern
    • Shape
      – Symmetric or skewed
      – How many modes?
      – Bell-shaped
  – Outliers

• Graphical tools for quantitative data
  – Histograms
  – Time plots