Descriptive statistics

Describing data with numbers: measures of location
What to describe?

• What is the “location” or “center” of the data? ("measures of location")

• How do the data vary? ("measures of variability")
Measures of Location

- Mean
- Median
- Mode
Mean

- Another name for average.
- If describing a population, denoted as \( \mu \), the greek letter “mu”.
- If describing a sample, denoted as \( \bar{x} \), called “x-bar”.
- Appropriate for describing measurement data.
- Seriously affected by unusual values called “outliers”.
Calculating Sample Mean

Formula: \( \bar{X} = \frac{\sum X_i}{n} \)

That is, add up all of the data points and divide by the number of data points.

Data (# of classes skipped): \( 2 \ 8 \ 3 \ 4 \ 1 \)

Sample Mean = \( \frac{2+8+3+4+1}{5} = 3.6 \)

Do not round! Mean need not be a whole number.
Median

• Another name for 50th percentile.
• Appropriate for describing measurement data.
• “Robust to outliers,” that is, not affected much by unusual values.
Calculating Sample Median

Order data from smallest to largest.

If odd number of data points, the median is the middle value.

Data (# of classes skipped): 2 8 3 4 1

Ordered Data: 1 2 3 4 8

Median
Calculating Sample Median

Order data from smallest to largest.

If even number of data points, the median is the average of the two middle values.

Data (# of classes skipped): \(2 \ 8 \ 3 \ 4 \ 1 \ 8\)

Ordered Data: \(1 \ 2 \ 3 \ 4 \ 8 \ 8\)

\[\text{Median} = \frac{3+4}{2} = 3.5\]
Mode

• The value that occurs most frequently.
• One data set can have many modes.
• Appropriate for all types of data, but most useful for categorical data or discrete data with only a few number of possible values.
In Minitab:

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone</td>
<td>139</td>
<td>121.6</td>
<td>60.0</td>
<td>88.1</td>
<td>217.7</td>
<td>18.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone</td>
<td>2.0</td>
<td>2000.0</td>
<td>30.0</td>
<td>120.0</td>
</tr>
</tbody>
</table>

N = number of data points

Sample mean

Sample median
In Minitab:

• Select Stat.
• Select Basic Statistics.
• Select Display Descriptive Statistics.
• Select variable(s) of interest.
• Select OK.
The most appropriate measure of location depends on …

the shape of the data’s distribution.
Most appropriate measure of location

• Depends on whether or not data are “symmetric” or “skewed”.
• Depends on whether or not data have one (“unimodal”) or more (“multimodal”) modes.
Symmetric and Unimodal
Symmetric and Unimodal

GPA

[Box plot diagram]
Symmetric and Unimodal

Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>92</td>
<td>3.0698</td>
<td>3.1200</td>
<td>3.0766</td>
<td>0.4851</td>
<td>0.0506</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>GPA</td>
<td>2.0200</td>
<td>3.9800</td>
<td>2.6725</td>
<td>3.4675</td>
</tr>
</tbody>
</table>
Symmetric and Bimodal
Symmetric and Bimodal

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>84</td>
<td>70.048</td>
<td>70.000</td>
<td>70.092</td>
<td>3.030</td>
</tr>
<tr>
<td>Females</td>
<td>89</td>
<td>64.798</td>
<td>65.000</td>
<td>64.753</td>
<td>2.877</td>
</tr>
<tr>
<td>All</td>
<td>176</td>
<td>67.313</td>
<td>67.000</td>
<td>67.291</td>
<td>4.017</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>SE</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>0.331</td>
<td>63.0</td>
<td>76.0</td>
<td>68.0</td>
<td>72.0</td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>0.305</td>
<td>56.0</td>
<td>77.0</td>
<td>63.0</td>
<td>67.0</td>
<td></td>
</tr>
<tr>
<td>All</td>
<td>0.303</td>
<td>56.0</td>
<td>77.0</td>
<td>64.0</td>
<td>70.0</td>
<td></td>
</tr>
</tbody>
</table>
Symmetric and Bimodal
Skewed Right

Number of Music CDs of Spring 1998 Stat 250 Students

Number of Music CDs

Frequency

0 100 200 300 400

20

10

0

0
Skewed Right

Number of CDs
Skewed Right

Descriptive Statistics

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDs</td>
<td>92</td>
<td>61.04</td>
<td>46.50</td>
<td>52.93</td>
<td>62.90</td>
<td>6.56</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDs</td>
<td>0.00</td>
<td>400.00</td>
<td>21.50</td>
<td>83.00</td>
</tr>
</tbody>
</table>
Skewed Left

![Bar chart showing skewed left distribution of grades.](chart.png)
Skewed Left
Skewed Left

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>TrMean</th>
<th>StDev</th>
<th>SE Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>grades</td>
<td>22</td>
<td>89.18</td>
<td>93.50</td>
<td>90.60</td>
<td>12.92</td>
<td>2.76</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Q1</th>
<th>Q3</th>
</tr>
</thead>
<tbody>
<tr>
<td>grades</td>
<td>50.00</td>
<td>100.00</td>
<td>87.00</td>
<td>98.00</td>
</tr>
</tbody>
</table>
Choosing Appropriate Measure of Location

- If data are symmetric, the mean, median, and mode will be approximately the same.
- If data are multimodal, report the mean, median and/or mode for each subgroup.
- If data are skewed, report the median.