DESCRIPTIVE STATISTICS
(Weiss, Chapters 2 & 3)

2.1 Variables and Data

• Variable

A characteristic that

• Data

Information obtained by

Two types of data

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Response</td>
<td>Response</td>
</tr>
</tbody>
</table>

Ex: Ex:
Quantitative Data

- **Discrete**
  Whose possible values form a finite (or countable infinite) set of numbers.
  
  Ex.

- **Continuous**
  Whose possible values form some interval of numbers.
  
  Ex.

**Important:**

The statistical technique employed depends on the type of data.

**Qualitative** - **Quantitative** -
Summary:

Variable
2.2 Grouping Data

The Bureau of Economic Analysis gathers information on the length of stay in Europe by US travelers. A sample of 36 US residents yield the following data:

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>41</td>
<td>16</td>
<td>6</td>
<td>21</td>
<td>1</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>31</td>
<td>20</td>
<td>27</td>
<td>17</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>32</td>
<td>2</td>
<td>48</td>
<td>8</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>44</td>
<td>1</td>
<td>56</td>
<td>5</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>15</td>
<td>10</td>
<td>18</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>14</td>
<td>12</td>
<td>64</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Group data using classes of equal width starting with the class 0 - (12).

0 - (12): 0 inclusive up to 12 exclusive - up to and not including.
<table>
<thead>
<tr>
<th>Length of Stay</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - (12)</td>
<td></td>
</tr>
<tr>
<td>36 - (48)</td>
<td>2</td>
</tr>
<tr>
<td>48 - (60)</td>
<td>2</td>
</tr>
<tr>
<td>60 - (72)</td>
<td>1</td>
</tr>
</tbody>
</table>

Guidelines for grouping data:

1. # classes small enough to provide effective summary & large enough to display relevant characteristics.

2. Each observation must belong to one and only one class.

3. Whenever feasible, all classes should have the same width.
Frequency & Relative-Frequency

- Frequency

  The number of

- Relative Frequency

  expressed as a decimal. Hence

  \[
  \text{Relative Frequency} = \frac{f_i}{n}
  \]

  where

  \( n \) : total number of measurements in the sample

  \( f_i \) : frequency for the \( i \)th category.

- Midpoint
<table>
<thead>
<tr>
<th>Length of Stay</th>
<th>Freq.</th>
<th>Relat. Freq.</th>
<th>Midpoint</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - (12)</td>
<td>15</td>
<td>0.417</td>
<td></td>
</tr>
<tr>
<td>12 - (24)</td>
<td>13</td>
<td>0.361</td>
<td></td>
</tr>
<tr>
<td>24 - (36)</td>
<td>3</td>
<td>0.083</td>
<td></td>
</tr>
<tr>
<td>36 - (48)</td>
<td>2</td>
<td>0.056</td>
<td>42</td>
</tr>
<tr>
<td>48 - (60)</td>
<td>2</td>
<td>0.056</td>
<td>54</td>
</tr>
<tr>
<td>60 - (72)</td>
<td>1</td>
<td>0.028</td>
<td>66</td>
</tr>
</tbody>
</table>

**Single-Value Grouping**

Sometimes more appropriate to use classes that each represents a single possible value:

- Discrete data with few observations,

- Qualitative data.
Example 2.8, pg 52 (50)

Students in an introductory stats class were asked for their political party affiliation:

- Democratic (D)
- Republican (R)
- Other (O)

Responses:

<table>
<thead>
<tr>
<th>D</th>
<th>R</th>
<th>O</th>
<th>R</th>
<th>R</th>
<th>R</th>
<th>R</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>D</td>
<td>O</td>
<td>R</td>
<td>D</td>
<td>O</td>
<td>O</td>
<td>R</td>
<td>D</td>
</tr>
<tr>
<td>D</td>
<td>R</td>
<td>O</td>
<td>D</td>
<td>R</td>
<td>R</td>
<td>O</td>
<td>R</td>
</tr>
<tr>
<td>D</td>
<td>O</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>R</td>
<td>O</td>
<td>D</td>
</tr>
<tr>
<td>O</td>
<td>R</td>
<td>D</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>R</td>
<td>D</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Party</th>
<th>Frequency</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### 2.3 Graphs and Charts

- **Histograms**

Ex. US travelers in Europe

<table>
<thead>
<tr>
<th>Length of Stay</th>
<th>Freq.</th>
<th>Relat. Freq.</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - (12)</td>
<td>15</td>
<td>0.417</td>
</tr>
<tr>
<td>12 - (24)</td>
<td>13</td>
<td>0.361</td>
</tr>
<tr>
<td>24 - (36)</td>
<td>3</td>
<td>0.083</td>
</tr>
<tr>
<td>36 - (48)</td>
<td>2</td>
<td>0.056</td>
</tr>
<tr>
<td>48 - (60)</td>
<td>2</td>
<td>0.056</td>
</tr>
<tr>
<td>60 - (72)</td>
<td>1</td>
<td>0.028</td>
</tr>
</tbody>
</table>
1. **Frequency Histogram**: Plot
   
   Classes
   Frequencies

2. **Relative-frequency Histogram**: Plot
   
   Classes
   Relat. Freq.

**Note**: Freq. and rel. freq. are both represented by vertical bars, whose height is the freq. or rel. freq. of the class.
Frequency Histogram - Travel example

- Inspection of the histogram reveals the
Relative-frequency Histogram - Travel ex.

US travelers

- Shape of both graphs are
Cumulative Frequency Distributions

Usually cumulate for

Number of observations < class end point

Ex. US travelers:

< 36 days
< 36 days

or relative cumulative frequency of

< 36 days
**Ogive - pg 67 (65)**

Plot relative cumulative frequencies against upper end points of classes.

**Ex. US travelers**

<table>
<thead>
<tr>
<th>Length of stay</th>
<th>Class end point</th>
<th>Freq</th>
<th>Cumul Freq</th>
<th>Relat Cumul Freq</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>0 -(12)</td>
<td>12</td>
<td>12</td>
<td>0.417</td>
<td></td>
</tr>
<tr>
<td>12-(24)</td>
<td>24</td>
<td>15</td>
<td>0.778</td>
<td></td>
</tr>
<tr>
<td>24-(36)</td>
<td>36</td>
<td>3</td>
<td>0.861</td>
<td></td>
</tr>
<tr>
<td>36-(48)</td>
<td>48</td>
<td>2</td>
<td>33</td>
<td>0.917</td>
</tr>
<tr>
<td>48-(60)</td>
<td>60</td>
<td>2</td>
<td>35</td>
<td>0.972</td>
</tr>
<tr>
<td>60-(72)</td>
<td>72</td>
<td>1</td>
<td>36</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Uses of ogives

- What is the length of stay that 50% of US travelers fall above (below) of? (known as Median.)

Read off the ogive:

- What is the minimum length of stay the top 10% of travelers stay in Europe?

Approximately
Percentiles

The pth percentile is a number with p% of the values below it and (100-p)% above it.

Ex. US travelers

First quartile  25th percentile ≈ 10th percentile
Median  50th percentile ≈
Third quartile  75th percentile ≈ 90th percentile
Graphical Displays of Qualitative Data

- **Pie chart**

  Widely used for showing fractions -

Ex. Recall example 2.8 pg 52 (50). Political affiliation of 40 students.

  Responses: 13 Democratic
              18 Republican
              9 Other

<table>
<thead>
<tr>
<th>Party</th>
<th>Relative Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democrats</td>
<td>0.325</td>
</tr>
<tr>
<td>Republican</td>
<td>0.450</td>
</tr>
<tr>
<td>Other</td>
<td>0.225</td>
</tr>
</tbody>
</table>
• Bar Chart

Like a histogram, except that its bars

[Diagram showing a bar chart with political affiliation categories: Democratic, Republican, Other]
2.4 Stem and Leaf Diagrams

Steps:

1. 

2. 

3. 

4. 

Ex. Consider the following set of student examination marks

<table>
<thead>
<tr>
<th>62</th>
<th>55</th>
<th>48</th>
<th>73</th>
<th>73</th>
<th>75</th>
<th>39</th>
</tr>
</thead>
<tbody>
<tr>
<td>82</td>
<td>94</td>
<td>81</td>
<td>85</td>
<td>68</td>
<td>65</td>
<td>29</td>
</tr>
<tr>
<td>61</td>
<td>51</td>
<td>69</td>
<td>76</td>
<td>89</td>
<td>78</td>
<td>77</td>
</tr>
<tr>
<td>70</td>
<td>67</td>
<td>73</td>
<td>56</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Ex. Examination marks

Stem-and-Leaf Display: Marks

Stem-and-leaf of Marks    N  =  25
Leaf Unit = 1.0

2 9
3 9
4 8
5 156
6 125789
7 03335678
8 1259
9 4

Stem-and-Leaf Display: Marks

Stem-and-leaf of Marks    N  =  25
Leaf Unit = 1.0

2 9
3
3 9
4
4 8
5 1
5 56
6 12
6 5789
7 0333
7 5678
8 12
8 59
9 4
2.5 Distribution shapes: Symmetry and Skewness

Shape

- important aspect of the dist’n of quantitative data
- plays a role in determining the correct method of statistical analysis.

Distribution Shapes?

1. Symmetric Dist’n
   A dist’n that can be divided into two pieces that are mirror images of one another
2. **Skewed Dist’n**

A dist’n that is not symmetric is either
• Population and Sample Distributions

The dist’n of a population is called the population dist’n.

The dist’n of a sample is called the sample dist’n.

IMPORTANT!!

For a particular population:

See figure 2.11, pg 76 (75), in textbook.