Descriptive Statistics I

I. Frequency Distributions

II. Measures of Central Tendency

Frequency Distributions

A. A first step in summarizing a set of observation
B. Definition:
   The number of scores in a given range
C. Use with both Qualitative & Quantitative data

D. Example: Qualitative Data

   Type of shoes
E. Quantitative Example

8.00 10.00 5.00 9.00 6.00 7.00 9.00
9.00 7.00 9.00 7.00 9.00 10.00 9.00 8.00
5.00 9.00 8.00 8.00 7.00 7.00 7.00 6.00
9.00 8.00

10.00 10.00
9.00 9.00 9.00 9.00 9.00 9.00 9.00 9.00 9.00
8.00 8.00 8.00 8.00 8.00
7.00 7.00 7.00 7.00 7.00 7.00 7.00
6.00 6.00
5.00 5.00
**Frequency Distributions**

### E. Quantitative Example

<table>
<thead>
<tr>
<th>X</th>
<th>f</th>
<th>cf</th>
<th>%Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>23</td>
<td>92</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>15</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>

\[ \sum f = 25 = n \]

---

**Frequency Distribution Graph: Histogram**

---

**Shapes of Frequency Distributions**

- Symmetrical
- Positive Skew
- Negative Skew
Number of words recalled on a memory test.

Reaction time on a Stroop color naming task.

Grades in a junior level course.
II. Measures of Central Tendency

The one score that would best capture, or summarize a group of scores.

Three ways to conceptualize this:
Mode
Median
Mean

Mode: \( M_o \)

The most frequent observation:
most common
typical value
Good for qualitative variables
(academic major)
Good for discrete variables
(number of children)

What is the modal footwear?

<table>
<thead>
<tr>
<th>Frequency</th>
<th>boots</th>
<th>clogs</th>
<th>leather shoes</th>
<th>sandals</th>
<th>sneakers</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
What is the mode?

Median: \( M_d \)
- The middle score
- The score that cuts the distribution into two equal halves
- The 50th percentile

Median

Three ways to calculate the median
1. Arrange scores high to low and pick the middle score.
   2 4 3 52 6
   2 3 4 6 52
Three ways to calculate the median

2. With an even number of scores, average the two middle scores.
   2 4 6 8

   Median = (4+6)/2 = 5

3. From a frequency distribution table, with tied scores near the middle, calculate the 50th percentile.

   Two examples:
### Easy Example

<table>
<thead>
<tr>
<th>X</th>
<th>freq.</th>
<th>Cumulative frequency</th>
<th>%Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>23</td>
<td>92</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>15</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>10</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>

What is the 50th Percentile?

### Harder Example

<table>
<thead>
<tr>
<th>X</th>
<th>f</th>
<th>cf</th>
<th>%Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>23</td>
<td>92</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>15</td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>5</td>
<td>20</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
</tbody>
</table>

What is the 50th Percentile?
Interpolation

8.5 - 60h
? - 50th
7.5 - 20h

\[
\frac{(50-20)}{(60-20)} = \frac{30}{40} = .75
\]

8.5 - 7.5 = 1.0

.75 x 1.0 = .75
7.5 + .75 = 8.25

Hardest Example

<table>
<thead>
<tr>
<th>X</th>
<th>f</th>
<th>cf</th>
<th>%Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>2</td>
<td>12</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>5</td>
<td>10</td>
<td>83.33</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>5</td>
<td>41.66</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>2</td>
<td>16.66</td>
</tr>
</tbody>
</table>

What is the 50th Percentile?
When to use the median

- Ordinal data
- When there are extreme scores
  - Example: 2, 4, 5, 7, 11, 259
    - Median = 6
    - Mean = 48
- With open ended questions or indeterminate values:
  - Example: How many times a day do you eat 1, 2, 3, 4, 5 or more (circle one)

Mean: $M$

The arithmetic average

\[ M = \frac{\sum x}{n} \]

Example: 3, 4, 6, 7
\[ \sum x = 20, \ M = \frac{20}{4} = 5 \]

Mean measures the balancing point of a distribution.
Mean: $M$

Calculating $M$ from a frequency distribution

<table>
<thead>
<tr>
<th>$X$</th>
<th>$f$</th>
<th>$xf$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
<td>20</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>72</td>
</tr>
<tr>
<td>8</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>42</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>

$\Sigma f = n = 25 \quad \Sigma x = 196$

$$M = \frac{\Sigma x}{n} = \frac{196}{25} = 7.84$$

Mean: $M$

Weighted means:

$$M = \frac{n_1M_1 + n_2M_2}{n_1 + n_2}$$

Mean: $M$

When to use the mean:

- Most frequently used measure.
- Must have interval data or better.
- Greatly influenced by extreme scores.

but
Measure of Central Tendency and the shape of the frequency distribution.

Symmetrical Distributions

\[ M = M_d = M \]

Skewed Distributions:

\[ M_s < M < M_d \]
Measure of Central Tendency and the shape of the frequency distribution.

Bimodal

Symmetrical bimodal distribution:

\[ M_1, M_2 = M, M \]

Age of onset of essential tremor is bimodal in some samples.
(Louis & Dogu, 2007)
Conclusion: Bias in the referral of young patients to the Columbia University Medical center.

Measures of Central Tendency:

wrap-up

Three different measures, each answering a different question, and suitable in different situations.