Chapter 5  
Measures of Variability  
PSY 295 – Oswald  

Outline  
• The general problem  
• Range and related statistics  
• Deviation scores  
• The variance and standard deviation  
• Boxplots  
• Thought questions  

The General Problem  
• Remember that central tendency (mean, median, mode)  
• Does the center represent ?  
  • modal # of cars owned  
  • average level of extraversion  
• Dispersion around the center helps us know  
  – around something  
  –  

Variation  
• Variability - The extent numbers in a data set are  
• When all elements measured receive the same scores (e.g., everyone in the data set has the same age, in years), there is  
  • e.g., Quality control =  
  • As the scores in a data set become more dissimilar  

Variation  
\[ X - \bar{X} \]  
<table>
<thead>
<tr>
<th>X</th>
<th>5</th>
<th>0.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>
\[ \sum X = 25 \quad n = 5 \quad \bar{X} \cdot 5 \]  

Variation  
\[ X - \bar{X} \]  
<table>
<thead>
<tr>
<th>X</th>
<th>6</th>
<th>+1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>-1.00</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>+1.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-1.00</td>
<td></td>
</tr>
</tbody>
</table>
\[ \sum X = 25 \quad n = 5 \quad \bar{X} \cdot 5 \]
Chapter 5  Measures of Variability

Variation

\[
\bar{X} = \frac{\sum X}{n}
\]

<table>
<thead>
<tr>
<th>X</th>
<th>(X - \bar{X})</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>+3.00</td>
</tr>
<tr>
<td>1</td>
<td>-4.00</td>
</tr>
<tr>
<td>9</td>
<td>+4.00</td>
</tr>
<tr>
<td>5</td>
<td>0.00</td>
</tr>
<tr>
<td>2</td>
<td>-3.00</td>
</tr>
</tbody>
</table>

\[\sum X = 25\quad n = 5\quad \bar{X} = 5\]

Avg. Deviation from the Mean?

• Let’s say we wanted to figure out the average deviation from the mean. You could

\[
\frac{\sum (X - \bar{X})}{N}
\]

• BUT We have a problem. This number will always (by definition of what a mean is)

Range and Related Statistics

• The range
  – Distance from
  – Very of variability – by extremes
  – Example: The youngest student in a class is 19 and the oldest is 46.

Range and Related Statistics

• The interquartile range (IQR)
  – Delete
  – IQR is range of what remains
  – May be

Trimmed Samples

• Delete of extreme scores
• Trimmed statistics are
Deviation Scores

- Definition
  - distance between an individual score and the # representing the center of the distribution
  - usually 
  $$(X - \bar{X})$$
- Importance

Estimators

- Mean
  - is an unbiased estimate of population mean
- Define unbiased
  - Long range average of statistic is equal to 

Estimators

- Variance
  - Is of $\sigma^2$
  - Using $N$ instead of $\sigma^2$ would be
  - Take square root of this estimate for , the standard deviation

Estimators

- Standard deviation
  - Square root of the of the variance
  - Easier to define than variance – variance deals with squared scores but the standard deviation is
  - Standard = , deviation =

Relationship Between Population and Sample

Variance

- Definitional formula
  - Population
  - Sample
Chapter 5  Measures of Variability

Calculation

\[
\begin{array}{cccccc}
X & 2 & 4 & 5 & 8 & 7 & 4 \\
(X - \bar{X}) & -3 & -1 & 0 & 3 & 2 & -1 \\
(X - \bar{X})^2 & 9 & 1 & 0 & 9 & 4 & 1 \\
\sum & 30 & & & & & \\
\end{array}
\]

\[s^2 = \frac{\sum (X - \bar{X})^2}{N - 1} = \]

\[\bar{X} = 5\]

Standard Deviation

- **Definitional formula**
  - The square root of the variance
  \[s = \sqrt{s^2} = \sqrt{\frac{\sum (X - \bar{X})^2}{N - 1}}\]
  - Why do this?
  - Why not just take the difference and ?

Computational Formula

The definitional formula is for
\[s^2 = \frac{\sum (X - \bar{X})^2}{N - 1}\]

...but it's the same as the computational formula which is
\[s^2 = \frac{\sum X^2 - (\sum X)^2}{N - 1}\]

which is often

Example w/ Computational Formula

- Always create a table that looks like this

<table>
<thead>
<tr>
<th>X</th>
<th>X^2</th>
<th>X</th>
<th>X^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>X_1</td>
<td>X_1^2</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>X_2</td>
<td>X_2^2</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>X_3</td>
<td>X_3^2</td>
<td>7</td>
<td>49</td>
</tr>
<tr>
<td>X_4</td>
<td>X_4^2</td>
<td>9</td>
<td>81</td>
</tr>
</tbody>
</table>

\[\sum X = \sum X^2 = \]

Computing Standard Deviation

- Values from the table are then entered into the formula as follows:

\[s^2 = \frac{\sum X^2 - (\sum X)^2}{N - 1}\]

\[s = \sqrt{\frac{\sum X^2 - (\sum X)^2}{N - 1}}\]

- Voila! (…and how would you get the variance?)

\[\sum X^2 = \]

\[\sum (X)^2 = \]
Get this down to a single #:

\[
s = \sqrt{\frac{(150-121)}{3}} \quad \Rightarrow \quad s = \sqrt{\frac{29}{3}}
\]

\[
s = \sqrt{9.6667} \quad \Rightarrow \quad s = 3.1091 \quad \Rightarrow \quad s = 3.11
\]

Degrees of Freedom

- Degrees of Freedom: The number of in a data set – or, the number of observations that are free to vary.

- For example, say there are 5 numbers that total 25 \( \sum x = 25, n = 5 \)
- Many combinations of numbers can total 25, can be any value
- The 5th number cannot vary if you make \( \sum x = 25 \)
- So here you have , because the five numbers are free to vary

Using for the variance and standard deviation (versus using N) gives us of the population variance and standard deviation.
- The , the more this will make a difference (e.g., when \( N=2 \) than when \( N=20,000 \)).

Why is Standard Deviation so Important?

- What does the standard deviation really tell us?
- Why would a sample’s standard deviation be small?
- Why would a sample’s standard deviation be large?

An Example

- You’re sitting in the student union with 4 of your friends. A student walks by, and being the superficial scientists you are, you and your friends view the student for about 2 seconds, and you rate this person’s “charisma” on a scale from 1 to 10 (where 1=not at all appealing and 10=very appealing)
Food for thought

1) What would it mean if all six of you rated this person a 9 out of 10?
2) What would it mean if all six of you rated this person a 5 out of 10?
3) What would it mean if the five of you produced the following ratings: 2, 4, 5, 8, 7, 4 (note that the mean rating would be 5)?

Why would scenario #3 happen instead of scenario #2? What factors would lead to these different ratings?

These questions form the basis of why

Computational Formula

\[ s^2 = \frac{\sum X^2 - \left(\frac{\sum X}{N}\right)^2}{N-1} = \frac{2^2 + 4^2 + 5^2 + 7^2 + 4^2 - 30^2}{6} = 4.80 \]

\[ s = \sqrt{\frac{\sum X^2 - \left(\frac{\sum X}{N}\right)^2}{N-1}} = \sqrt{4.8} = 2.19 \]

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Results

- took much longer to run on average (higher mean)
- Much greater
  - See following graphs for Anthrax and Mozart
  - Both X axes

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.

Example: Mice and Music

- Study by David Merrell
- Raised some mice in
- Raised some mice listening to
- Raised other mice listening
- Dependent variable is the after 4 weeks.
**Merrell’s Music Study Printout**

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Mean</th>
<th>N</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiet</td>
<td>307.2319</td>
<td>23</td>
<td>71.8267</td>
</tr>
<tr>
<td>Mozart</td>
<td>114.5833</td>
<td>24</td>
<td>36.1017</td>
</tr>
<tr>
<td>Anthrax</td>
<td>1825.8889</td>
<td>24</td>
<td>103.1392</td>
</tr>
<tr>
<td>Total</td>
<td>755.4601</td>
<td>71</td>
<td>777.9646</td>
</tr>
</tbody>
</table>

**Boxplots**

- The general problem
  - A display that shows dispersion of distribution
- Calculation steps
  - Find median
  - Find top and bottom 25% points (quartiles)
  - Eliminate top and bottom 2.5% (fences)
  - Draw boxes to quartiles and whiskers to fences, with remaining points as outliers
- Boxplots for comparing groups

**Thought Questions**

(these are NOT the only questions you need to know for HW/Exams)

- What do we look for in a measure of dispersion?
- What role do play in considering the mean and variance?
- If scores are on a 1-5 scale, how would the variance change if the mean was 1.1 vs. if the mean was 3.3?
- Why do we take the to get the standard deviation?
- What types of things can we conclude from the mice/music data?)