Assessing Variable Relationships

• Up to this point, we’ve been interested in exploring how groups differ on a variable of interest
  – $X\text{-bar}_1$ – $X\text{-bar}_2$
  – Here, the focus seems to be more on groups than on the variables themselves
  – Sometimes, we’re more interested in how variables relate to one another than how groups differ
Assessing Variable Relationships

• What are variable relationships?
  – Group comparisons
    • How do men vs. women differ on a measure of interest in sports?
    • Are there differences between scores on academic achievement tests between different ethnicities/cultures?
  – Variable comparisons
    • For our sample, how is self-esteem related to academic achievement?

Assessing Variable Relationships

• ANOVA and t-tests cannot assess variable relationships
• We must use a new set of techniques
  – Correlation
  – Regression
  – Multiple regression
Assessing Variable Relationships

• Correlation
  – A statistical procedure wherein we estimate the amount of variance shared by two variables
  • Correlation is bidirectional
  • Correlation ≠ causation

Assessing Variable Relationships

• Regression
  – A statistical procedure wherein we estimate the amount of variance in 1 variable predicted by a second variable
  • Regression assumes that 1 variable predicts the presence/level of another
  • Regression ≠ causation… unless you collect your data carefully
Assessing Variable Relationships

• Multiple Regression
  – A statistical procedure wherein we estimate the amount of variance in 1 variable predicted by set of variables
  • Multiple Regression assumes that several variables combine to predict the presence/level of another variable
  • Multiple Regression ≠ causation… unless you collect your data carefully

Assessing Variable Relationships

• Correlation – discussed in this class
• Regression – time permitting
• Multiple Regression – graduate school
Terminology of Correlation

• Types of variables
  – Random variables
    • Defined as variables outside of the experimenter's manipulation or control
      – Weight gain demonstrated by AN individuals
      – Rate of self-mutilation
      – GRE score
      – Hours of TV watched in a day/evening/week/month
    • “Data” – DV
    • Replication of the study will leave us with different values of random variables

• Types of variables
  – Fixed variables
    • Defined as variables “fixed” or set by the experimenter
      – Assignment to drug Tx or placebo control
      – Psychotherapy Tx assignment
      – Assigning individuals to watch 1, 2, or 3 hours of TV/night
    • “Categories” – IV
    • Fixed variables remain constant across replications of the experiment
Correlation vs. Regression

• Regression
  – 1 fixed variable (i.e. Tx group)
  – 1 random variable (i.e. response to medication, depression level, etc.)
  – Prediction of score based on group membership

• Correlation
  – 2 random variables (i.e. association of depression to self-injury)
  – relationship of 2 DVs

The Scatterplot

• A scatterplot is a visual representation of the relationship between two variables
  – A throw-back to middle-school mathematics
  – Graph paired scores for each individual
    • 1 variable on the x-axis (predictor)
    • 1 variable on the y-axis (criterion)
  – 3 classes of outcomes with the scatterplot
Positive Linear Relationship

- Lower-left hand corner $\rightarrow$ upper right-hand corner
- Correlation = 1

![Graph showing positive linear relationship]

Negative Linear Relationship

- Upper-left hand corner $\rightarrow$ lower right-hand corner
- Correlation = -1

![Graph showing negative linear relationship]
No Linear Relationship

- A swarm of plots with no apparent line connecting them
- Correlation $\approx 0$

The Regression Line

- When we calculate a regression or correlation, we generate a regression line
  - A line that estimates the best “line” to link all of the x-variables to the y-variables
  - Regression
    - Values of y that we would expect for a given value of x
  - Correlation
    - Line represents the degree of correlation—the closer observed data fits the line, the greater the correlation
The Road to Correlation: The Covariance

• The calculations for the correlation begin with a calculation of the covariance
  – The covariance represents the degree to which two variables differ, change, or vary together
  – “co” = together “vary” = differ

The Covariance

\[
\text{cov}_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{N - 1}
\]

Translation: The covariance is equal to the average amount of variance between x and y
The Covariance

\[
\text{cov}_{xy} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{N - 1}
\]

A more simple computational form of the equation… which should be rather familiar to you by now

Thus, to calculate the covariance, we need to know \(\sum X\), \(\sum Y\), & \(\sum XY\)

---

The Covariance: An Example

<table>
<thead>
<tr>
<th>#</th>
<th>X</th>
<th>Y</th>
<th>XY</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>7</td>
<td>35</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>10</td>
<td>80</td>
</tr>
<tr>
<td>(\sum X)</td>
<td>19</td>
<td>31</td>
<td>146</td>
</tr>
<tr>
<td>x-bar</td>
<td>3.8</td>
<td>6.2</td>
<td>29.8</td>
</tr>
<tr>
<td>n</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>(\sum X^2)</td>
<td>103</td>
<td>219</td>
<td>8058</td>
</tr>
<tr>
<td>s</td>
<td>2.77</td>
<td>2.58</td>
<td>657.93</td>
</tr>
</tbody>
</table>
Calculating the Covariance

\[ \text{cov}_{xy} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{N - 1} \]

\[ \text{cov}_{xy} = \frac{146 - (19)(31)}{5} \]

\[ \text{cov}_{xy} = \frac{146 - 589}{5} \]

\[ \text{cov}_{xy} = \frac{146 - 117.8}{4} \]

\[ \text{cov}_{xy} = \frac{28.2}{4} \]

\[ \text{cov}_{xy} = 7.05 \]

About the Covariance

- As noted before, the covariance serves as a measure of the degree to which two variables vary together.
- However, the covariance CANNOT be directly interpreted.
  - The size of the covariance is dependent on the size of the variances of \( x \) and \( y \).
  - The correlation is a conversion of the covariance to standardized form.
The Pearson Product-Moment Correlation

- The Pearson Product-Moment Correlation ($r$) is derived from the covariance
  - Correlations range from -1 to 1
    - If your correlation is less than -1 or greater than 1, you have calculated it incorrectly!

Calculating the Correlation

\[
 r = \frac{\text{COV}_{xy}}{S_x S_y}
\]

Simply divide the covariance by the product of the standard deviations of $x$ and $y$
Calculating the Correlation

\[ r = \frac{7.05}{(2.77)(2.58)} \quad r = \frac{7.05}{7.15} \quad r = .98 \]

However... is this correlation statistically significant? Does it differ from 0?

Testing the Significance of \( r \)

\[ t = \frac{r \sqrt{n - 2}}{\sqrt{1 - r^2}} \]

Tested on n-2 degrees of freedom
Testing the Significance of $r$

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

$$t = \frac{.98\sqrt{5-2}}{\sqrt{1-.98^2}}$$

$$t = \frac{.98\sqrt{3}}{\sqrt{1-.96}}$$

$$t = \frac{.98(1.73)}{\sqrt{.04}}$$

$$t = \frac{1.70}{.2}$$

$$t = 8.50$$

$t_{\text{obt}}(3) = 8.50$

$t_{\text{crit}}(3) = 3.18$

Reject $H_0$

The Effect Size of the Correlation

• Calculating the effect size for a correlation is a simple procedure

• The effect size reflects the percentage of variance shared between variables
  – $r^2 = .98 \times .98 = .96$
  – Thus, the x and y variables share 96% of their total variance
  • This is a large value—typically, we’re happy with correlations between .3 and .6 in research
  • $r > .80 = $multicolinearity
Unpacking the Meaning of $r$

- Remember: correlation does not mean causation
- Even when variables highly correlated, it may not be the case that the two variables are related to one another
  - The observed correlation may be caused by an association with a third variable
- Even IF causal… which direction?

Correlation: Uses

- Correlation serves as an excellent method with which we may explore variable relationships
  - We can NOT identify the origins of these relationships from correlation—only the size
  - Often used to determine other analyses that may be of interest
    - $t$-Tests to determine if different groups demonstrate differing levels of the variable of interest
Correlation: Presentation

• Correlations between variables are often presented as a group in what is known as a *correlation matrix*
  
  – Anxiety ↔ Life-Stress: $r = .35$
  – Anxiety ↔ Irrational Beliefs: $r = .50$
  – Life-Stress ↔ Irrational Beliefs: $r = .10$

Correlation Matrix

<table>
<thead>
<tr>
<th></th>
<th>Anxiety</th>
<th>Life-Stress</th>
<th>Irrational Beliefs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td>1</td>
<td>.35</td>
<td>.50</td>
</tr>
<tr>
<td>Life-Stress</td>
<td>.35</td>
<td>1</td>
<td>.10</td>
</tr>
<tr>
<td>Irrational Beliefs</td>
<td>.50</td>
<td>.10</td>
<td>1</td>
</tr>
</tbody>
</table>