Annotated Assignment

• psychinfo
• additional links on resources
• example

What know how to do

• means, standard deviations, sum of squares, variance, correlation
• Today, focus on (bivariate) regression
• Using association (correlations) to predict scores
  – Who will develop autism?
  – Who will succeed in graduate school?
  – Who will benefit from what kind of therapy?
  – Who will be “academically engaged”?
  – Who will be more likely to make upward or downward comparisons?
Linear regression

• Goal is to predict score on one scale based on the score for another scale
• Question is whether our prediction is any good?
• Should we care?
• What do equations tell us?
  – The relation between values of the predictor variable and the *predicted* values of the outcome variable

Terms

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Criterion/Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>Empathy</td>
<td>Satisfaction</td>
</tr>
<tr>
<td>Number of stressful events</td>
<td>Days to recover</td>
</tr>
</tbody>
</table>
Scattergram #1

Scattergram #2
(Bivariate) linear regression

- Predicted $Y = a + b \,(x)$
  - A person’s predicted score on the outcome equals the regression constant plus the raw score regression coefficient times that person’s score on the predictor
- Does psychologist’s degree of empathy predict therapist satisfaction?
  Predicted $Y = -.98 + .06 \,(x)$
- Does the number of stress events experienced in a week predict the number of days it takes to recover?
  Predicted $Y = 4.27 + .58 \,(x)$

Three approaches

1. Bivariate prediction from Z scores
2. Bivariate prediction from raw scores – Z score method
3. Bivariate prediction from raw scores – direct method
Bivariate prediction with Z scores

1. Determine the standardized regression coefficient.
   \[ \beta = r \]

2. Multiply the standardized regression coefficient by the person’s Z score on the predictor variable.
   \[ \text{Predicted } Z_y = (\beta)Z_x \]

Example

1. Determine the standardized regression coefficient.
   \[ \beta = .40 \]

2. Multiply the standardized regression Coefficient by the person’s Z score on the predictor variable.
   If \( Z_x = 1.8 \), predicted \( Z_y = (\beta)(Z_x) = (.40)(1.8) = .72 \)
   If \( Z_x = 0 \), predicted \( Z_y = (\beta)(Z_x) = (.40)(0) = 0 \)
   If \( Z_x = -1.8 \), predicted \( Z_y = (\beta)(Z_x) = (.40)(-1.8) = -.72 \)
Bivariate prediction with raw scores – Z score method

1. Change the person’s raw score on the predictor variable to a Z score.
   \[ Z_x = \frac{X - M_x}{SD_x} \]

2. Multiply the standardized regression coefficient by the person’s predictor variable Z score.
   Predicted \( Z_y \) = \( (\beta)(Z_x) \)

3. Change the person’s predicted Z score on the criterion variable to a raw score.
   Predicted \( Y = (SD_y)(Z_y) + M_y \)

Example

Imagine the \( r = .84 \), the mean for (mood) = 4, SD = 1.73 and the mean for (sleep) = 8, SD = 1.10

1. Change the person’s raw score on the predictor variable to a Z score.
   \[ Z_x = \frac{X - M_x}{SD_x} = \frac{10 - 8}{1.1} = \frac{2}{1.1} = 1.82 \]

2. Multiply the standardized regression coefficient by the person’s predictor variable Z score.
   Predicted \( Z_y = (\beta)(Z_x) = (.84)(1.82) = 1.53 \)

3. Change the person’s predicted Z score on the criterion variable to a raw score.
   Predicted \( Y = (SD_y)(Z_y) + M_y = (1.73)(1.53) + 4 + 2.65 + 4 = 6.65 \)
Bivariate prediction with raw scores –
direct method

1. Figure the regression constant: Using raw score prediction with z score, find predicted Y when X=0. This is a.

2. Figure the raw-score regression coefficient: Using raw score prediction with Z scores, find predicted Y when X=1, b is this number minus a.

3. Find the person’s predicted raw score on the criterion variable: Predicted Y = a + b (X)

Example

Imagine the r=.84, the mean for (mood)=4, SD=1.73 and the mean for (sleep)=8, SD=1.10

1. Figure the regression constant: Using raw score prediction with z score, find predicted Y when X=0. This is a.

   \[ Z_x = \frac{(X - M_x)}{SD_x} = \frac{(0 - 8)}{1.10} = -7.27 \]

   Predicted \( Z_y \) = (r)\( (Z_x) \) = (.84)(-7.27) = -6.11

   Predicted Y = (SD_y)(Predicted \( Z_y \)) + M_y = (1.73)(-6.11) + 4 = -6.57

   a = -6.57

1. Figure the raw-score regression coefficient: Using raw score prediction with Z scores, find predicted Y when X=1, b is this number minus a.

   \[ Z_x = \frac{(X - M_x)}{SD_x} = \frac{(1 - 8)}{1.10} = -6.26 \]

   Predicted \( Z_y \) = (r)\( (Z_x) \) = (.84)(-6.36) = -5.34

   Predicted Y = (SD_y)(Predicted \( Z_y \)) + M_y = (1.73)(-5.34) + 4 = -5.24

   b = Predicted Y - a = -5.24 - (-6.57) = 1.33

1. Find the person’s predicted raw score on the criterion variable: Predicted Y = a + b (X). Assume X = 10

   Predicted Y = a + (b) ZX = -6.57 + 1.33 (X) = -6.57 + (1.33)(10) = 6.74
One more alternative

- Formula for b: (on board)
- Formula for a: (on board)
- To calculate b:
  1. Change the scores for each variable to deviation scores.
  2. Figure the product of the deviation scores for each pair of scores.
  3. Add up all the products of the deviation scores.
  4. Square each deviation score for the predictor variable (x).
  5. Add up the squared deviation scores for the predictor variable (X).
  6. Divide the sum of the products of deviation scores from step 3 by the sum of the deviations for predictor variable from step 5.
- To calculate a:
  1. Multiply the regression coefficient b, by the mean of the X variable.
  2. Subtract the results of step 1 from the mean of the Y variable.

Figuring proportionate reduction in error for sleep & mood

<table>
<thead>
<tr>
<th>Actual Y</th>
<th>Mean Y</th>
<th>Error</th>
<th>Error squared</th>
<th>Predicted Y</th>
<th>Error</th>
<th>Error squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>2.74</td>
<td>1.26</td>
<td>1.59</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>5.40</td>
<td>-.40</td>
<td>.16</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>-1</td>
<td>-1</td>
<td>4.07</td>
<td>1.07</td>
<td>1.15</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>3</td>
<td>9</td>
<td>5.40</td>
<td>1.60</td>
<td>2.56</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>4.07</td>
<td>-.07</td>
<td>.00</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>-3</td>
<td>9</td>
<td>1.41</td>
<td>-.41</td>
<td>.17</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>-1</td>
<td>-1</td>
<td>4.07</td>
<td>-1.07</td>
<td>1.15</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>-2</td>
<td>4</td>
<td>2.47</td>
<td>-.74</td>
<td>.55</td>
</tr>
<tr>
<td>6</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>6.73</td>
<td>-1.73</td>
<td>.53</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>4.07</td>
<td>.93</td>
<td>.86</td>
</tr>
</tbody>
</table>

\[ SS_{total}=30 \quad SS_{Error}=8.72 \]
Proportionate reduction in error = Reduction in squared error using bivariate prediction rule over squared error using mean to predict

Proportionate reduction in error = Sum of squared error using mean to predict MINUS Sum of squared error using mean to predict MINUS Sum of squared error using bivariate prediction rule to predict

Proportionate reduction in error = \( \frac{SS_{Total} - SS_{Error}}{SS_{Total}} \)

---

“Variability pie”

Error is actual score – predicted score
If had no other information, use mean as comparison

R\(^2\) or proportionate reduction in error

Variance due to error

Variance accounted for by predictor

Total variability among outcome scores
Tips for success

• Predicted score should always be closer to zero than the actual score
• Make sure using correct means and SDs
• What is known goes on bottom of graph, what is predicted is vertical
• If prediction does no better than the mean,
  $SS_{\text{Error}} = SS_{\text{Total}}$
• $SS_{\text{Error}}$ is never larger than $SS_{\text{Total}}$

SPSS

• Should know how to get means, SDs, frequencies and histograms for any variable
• Should know how to determine the Cronbach’s alpha for any set of items
• Should know how to reverse score a single item and combine items into a single scale
• Should know how to correlate any two variables (represented as single item or as scale)
Where are we going?

- Multiple linear regression
- How to write about results