Linear Relationships

Joseph J. Ferrare, Ph.D.
School of Interdisciplinary Arts & Sciences
University of Washington Bothell
www.josephferrare.com
Bivariate Relationships as Conditional Distributions

For categorical variables, we have already explored conditional distributions through a contingency table:

- i.e., conditional distributions of $Y$ given categories of $X$
Bivariate Relationships as Linear Functions

For quantitative variables, we can explore how the conditional distribution of $Y$ varies across different values of $X$

- Numerous equations exist
- The simplest is the linear function
Linear Functions

\[ Y = \alpha + \beta X \]

\[ Y = 3 + 2X \]

\( \alpha \) = \( Y \) intercept

\( \beta \) = rate of change in \( Y \) as \( X \) increases (slope)
Regression Function

\[ E(Y) = \alpha + \beta X , \]

\[ E(Y) = \text{expected or mean value} \]
\[ \alpha = \text{intercept of } Y, \]
\[ \beta = \text{slope of } X \]
\[ X = \text{independent var.} \]

\[ Y = 10.4 + (0.302)X \]

- Note that \( \alpha \) is a constant in the equation
- \( \alpha \) is the expected value of \( Y \) when \( \beta = 0 \)
- \( \beta \) is the expected change in \( Y \) for a one-unit increase in \( X \)
Measuring Linear Association

Pearson’s Correlation, $r$

- Measures the strength and direction of a linear relationship between two variables, $X$ & $Y$
- Standardized version of the slope (i.e., does not depend on units of measurement)
- $-1 < r < 1$, positive $r$ reflects positive correlation, negative $r$ reflects negative correlation
- When $r = 0$ there is no correlation between $X$ and $Y$
- Strength of correlation is determined by the absolute value of $r$
- $r$ assumes a symmetric relationship (i.e., no causality)
- An increase in 1 SD in $X$ corresponds to a change of $r$ SDs in $Y$
Pearson’s Correlation

- $r = 1$, i.e., perfect positive linear relationship
- $r = 0$, i.e., no linear relationship
- $r = -1$, i.e., perfect negative linear relationship
Pearson’s Correlation

Correlation between an individual’s education and mother’s education, $r = 0.395$

- An increase in 1 SD of mother’s education corresponds to a 0.395 SD increase in an individual’s education
Pearson’s Correlation

\[ H_0: \text{No correlation, i.e. } \rho_{XY} = 0, \text{ where } \rho \text{ is the population correlation} \]
Pearson’s Correlation

H₀: No correlation, i.e. ρₓᵧ = 0,
Result: Reject H₀ - an increase of 1 SD of height is associated with 0.07 SD increase in education

Correlations

<table>
<thead>
<tr>
<th>Highest year of school completed</th>
<th>Pearson Correlation</th>
<th>Sig. (2-tailed)</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest year of school completed</td>
<td>1</td>
<td>.009</td>
<td>2345</td>
</tr>
<tr>
<td>R is how tall</td>
<td>.070**</td>
<td>1</td>
<td>1400</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).
Pearson’s Correlation

H₀: No correlation, i.e. \( \rho_{XY} = 0 \)
Result: Reject H₀ – an increase of 1 SD in father’s education is associated with a 0.689 SD increase in mother’s education.

<table>
<thead>
<tr>
<th></th>
<th>Highest year school completed, mother</th>
<th>Highest year school completed, father</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
<td>1</td>
<td>.689**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td></td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>2089</td>
<td>1561</td>
</tr>
<tr>
<td>Highest year school completed, father</td>
<td>Pearson Correlation</td>
<td>.689**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>1561</td>
<td>1687</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).
Pearson’s Correlation

H₀: No correlation, i.e. ρₓᵧ = 0
Result: Reject H₀ – an increase of 1 SD of siblings is associated with a 0.264 decrease in education

Correlations

<table>
<thead>
<tr>
<th></th>
<th>Highest year of school completed</th>
<th>Number of brothers and sisters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest year of school completed</td>
<td>Pearson Correlation: 1</td>
<td>-.264**</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed): 0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N: 2345</td>
<td>2340</td>
</tr>
<tr>
<td>Number of brothers and sisters</td>
<td>Pearson Correlation: -.264**</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sig. (2-tailed): .000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>N: 2340</td>
<td>2343</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).