Psychology 371: Personality Research
Part II: Psychometric Theory
Part I: Overview and the correlation coefficient

William Revelle

Department of Psychology
Northwestern University
Evanston, Illinois USA

April, 2019
Outline

What is psychometrics?
  An overview
  A latent variable approach to measurement

Science as Model fitting
  Model fitting
  Data and scaling
    Assigning Numbers to Observations
    Coomb’s Theory of Data
    Ordering people,
    Proximity rather than order

Ordering objects
  Difficulties and artifacts of scaling

Types of scales and how to describe data
  Describing data graphically

Central Tendency and variance
  Shape

Correlation

History
What is psychometrics?

*In physical science a first essential step in the direction of learning any subject is to find principles of numerical reckoning and methods for practicably measuring some quality connected with it. I often say that when you can measure what you are speaking about and express it in numbers you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the stage of science, whatever the matter may be. (Thomsom, 1891)*

What is psychometrics?

The character which shapes our conduct is a definite and durable ‘something’, and therefore . . . it is reasonable to attempt to measure it. (Galton, 1884)

The history of science is the history of measurement” (J. M. Cattell, 1893)

Whatever exists at all exists in some amount. To know it thoroughly involves knowing its quantity as well as its quality (E.L. Thorndike, 1918)
What is psychometrics?

We hardly recognize a subject as scientific if measurement is not one of its tools (Boring, 1929)

There is yet another [method] so vital that, if lacking it, any study is thought ... not be scientific in the full sense of the word. This further an crucial method is that of measurement. (Spearman, 1937)

One’s knowledge of science begins when he can measure what he is speaking about and express in numbers (Eysenck, 1973)

Psychometrics: the assigning of numbers to observed psychological phenomena and to unobserved concepts. Evaluation of the fit of theoretical models to empirical data.
Psychometric Theory: A conceptual Overview

\[ \text{Error} \quad X \quad \text{Latent X} \quad \text{Latent Y} \quad Y \quad \text{Error} \]

\[
\begin{align*}
\delta_1 & \rightarrow X_1 \\
\delta_2 & \rightarrow X_2 \\
\delta_3 & \rightarrow X_3 \\
\delta_4 & \rightarrow X_4 \\
\delta_5 & \rightarrow X_5 \\
\delta_6 & \rightarrow X_6 \\
\delta_7 & \rightarrow X_7 \\
\delta_8 & \rightarrow X_8 \\
\delta_9 & \rightarrow X_9 \\
\chi_1 & \rightarrow Y_1 \quad \epsilon_1 \\
\chi_2 & \rightarrow Y_2 \quad \epsilon_2 \\
\chi_3 & \rightarrow Y_3 \quad \epsilon_3 \\
\xi_1 & \rightarrow Y_4 \quad \epsilon_4 \\
\xi_2 & \rightarrow Y_5 \quad \epsilon_5 \\
\chi_1 & \rightarrow Y_6 \quad \epsilon_6 \\
\chi_2 & \rightarrow Y_7 \quad \epsilon_7 \\
\chi_3 & \rightarrow Y_8 \quad \epsilon_8 \\
\end{align*}
\]
Observed Variables

Error \quad X \quad Latent X \quad Latent Y \quad Y \quad Error

\delta_1 \quad X_1 \quad \chi_1 \quad \xi_1 \quad Y_1 \quad \epsilon_1
\delta_2 \quad X_2 \quad \chi_1 \quad \xi_1 \quad Y_2 \quad \epsilon_2
\delta_3 \quad X_3 \quad \chi_1 \quad \xi_1 \quad Y_3 \quad \epsilon_3
\delta_4 \quad X_4 \quad \chi_1 \quad \xi_1 \quad Y_4 \quad \epsilon_4
\delta_5 \quad X_5 \quad \chi_2 \quad \xi_2 \quad Y_5 \quad \epsilon_5
\delta_6 \quad X_6 \quad \chi_2 \quad \xi_2 \quad Y_6 \quad \epsilon_6
\delta_7 \quad X_7 \quad \chi_2 \quad \xi_2 \quad Y_7 \quad \epsilon_7
\delta_8 \quad X_8 \quad \chi_3 \quad \xi_3 \quad Y_8 \quad \epsilon_8
\delta_9 \quad X_9 \quad \chi_3 \quad \xi_3 \quad Y_8 \quad \epsilon_8
Latent Variables

\begin{align*}
\delta_1 & \rightarrow X1 \rightarrow \chi_1 \\
\delta_2 & \rightarrow X2 \\
\delta_3 & \rightarrow X3 \\
\delta_4 & \rightarrow X4 \\
\delta_5 & \rightarrow X5 \\
\delta_6 & \rightarrow X6 \\
\delta_7 & \rightarrow X7 \\
\delta_8 & \rightarrow X8 \\
\delta_9 & \rightarrow X9 \\
\chi_1 & \rightarrow Y1 \rightarrow \xi_1 \\
\chi_2 & \rightarrow Y2 \\
\chi_3 & \rightarrow Y3 \\
\xi_1 & \rightarrow Y4 \\
\xi_2 & \rightarrow Y5 \\
\xi_2 & \rightarrow Y6 \\
\xi_2 & \rightarrow Y7 \\
\xi_2 & \rightarrow Y8
\end{align*}
Theory

Latent X

Latent Y

Error X

\[ \chi_1 \]
\[ \xi_1 \]
\[ \chi_2 \]
\[ \xi_2 \]
\[ \chi_3 \]

Error Y

\[ \delta_1 \]
\[ \delta_2 \]
\[ \delta_3 \]
\[ \delta_4 \]
\[ \delta_5 \]
\[ \delta_6 \]
\[ \delta_7 \]
\[ \delta_8 \]
\[ \delta_9 \]

\[ X_1 \]
\[ X_2 \]
\[ X_3 \]
\[ X_4 \]
\[ X_5 \]
\[ X_6 \]
\[ X_7 \]
\[ X_8 \]
\[ X_9 \]

\[ Y_1 \]
\[ Y_2 \]
\[ Y_3 \]
\[ Y_4 \]
\[ Y_5 \]
\[ Y_6 \]
\[ Y_7 \]
\[ Y_8 \]

\[ \epsilon_1 \]
\[ \epsilon_2 \]
\[ \epsilon_3 \]
\[ \epsilon_4 \]
\[ \epsilon_5 \]
\[ \epsilon_6 \]
\[ \epsilon_7 \]
\[ \epsilon_8 \]
A theory of data and fundamentals of scaling
Correlation, Regression, Partial Correlation, Multiple Regression

\[ \begin{align*}
\text{Error} & \quad X & \quad \beta_{y.x} & \quad \beta_{x.y} & \quad r_{xy} & \quad r_{x4y4.x5} & \quad R_{y.x6x7x8} \\
\delta_1 & \quad X1 & \quad Y1 & \quad \epsilon_1 \\
\delta_2 & \quad X2 & \quad Y2 & \quad \epsilon_2 \\
\delta_3 & \quad X3 & \quad Y3 & \quad \epsilon_3 \\
\delta_4 & \quad X4 & \quad Y4 & \quad \epsilon_4 \\
\delta_5 & \quad X5 & \quad Y5 & \quad \epsilon_5 \\
\delta_6 & \quad X6 & \quad Y6 & \quad \epsilon_6 \\
\delta_7 & \quad X7 & \quad Y7 & \quad \epsilon_7 \\
\delta_8 & \quad X8 & \quad Y8 & \quad \epsilon_8 \\
\delta_9 & \quad X9 
\end{align*} \]
Measurement: A latent variable approach.
Reliability: How well does a test reflect one latent trait?

Error $\delta_1$ $\delta_2$ $\delta_3$ $\delta_4$ $\delta_5$ $\delta_6$ $\delta_7$ $\delta_8$ $\delta_9$

Latent X $\chi_1$ $\chi_2$ $\chi_3$

Latent Y $\xi_1$ $\xi_2$ $\xi_3$

Y $Y_1$ $Y_2$ $Y_3$ $Y_4$ $Y_5$ $Y_6$ $Y_7$ $Y_8$

Error $\epsilon_1$ $\epsilon_2$ $\epsilon_3$ $\epsilon_4$ $\epsilon_5$ $\epsilon_6$ $\epsilon_7$ $\epsilon_8$
Face, Concurrent, Predictive, Construct

Error \( \delta_1 \) \( \delta_2 \) \( \delta_3 \) \( \delta_4 \) \( \delta_5 \) \( \delta_6 \) \( \delta_7 \) \( \delta_8 \) \( \delta_9 \)

\( X_1 \) \( X_2 \) \( X_3 \) \( X_4 \) \( X_5 \) \( X_6 \) \( X_7 \) \( X_8 \) \( X_9 \)

\( \chi_2 \) \( \chi_3 \)

Face (Faith) \( \chi_2 \) \( \chi_3 \)

Concurrent \( \xi_2 \)

Predictive \( \xi_2 \)

Construct \( \xi_2 \)

Latent \( Y_1 \) \( Y_2 \) \( Y_3 \) \( Y_4 \) \( Y_5 \) \( Y_6 \) \( Y_7 \) \( Y_8 \)

Error \( \epsilon_1 \) \( \epsilon_2 \) \( \epsilon_3 \) \( \epsilon_4 \) \( \epsilon_5 \) \( \epsilon_6 \) \( \epsilon_7 \) \( \epsilon_8 \)
Psychometric Theory: Data, Measurement, Theory

Error X Latent X Latent Y Y Error

\[ \delta_1 \rightarrow X1 \rightarrow \chi_1 \rightarrow Y1 \rightarrow \epsilon_1 \]
\[ \delta_2 \rightarrow X2 \rightarrow \chi_1 \rightarrow Y2 \rightarrow \epsilon_2 \]
\[ \delta_3 \rightarrow X3 \rightarrow \chi_1 \rightarrow Y3 \rightarrow \epsilon_3 \]
\[ \delta_4 \rightarrow X4 \rightarrow \chi_2 \rightarrow Y3 \rightarrow \epsilon_4 \]
\[ \delta_5 \rightarrow X5 \rightarrow \chi_2 \rightarrow Y5 \rightarrow \epsilon_5 \]
\[ \delta_6 \rightarrow X6 \rightarrow \chi_2 \rightarrow Y6 \rightarrow \epsilon_6 \]
\[ \delta_7 \rightarrow X7 \rightarrow \chi_3 \rightarrow Y7 \rightarrow \epsilon_7 \]
\[ \delta_8 \rightarrow X8 \rightarrow \chi_3 \rightarrow Y8 \rightarrow \epsilon_8 \]
\[ \delta_9 \rightarrow X9 \rightarrow \chi_3 \rightarrow Y8 \rightarrow \epsilon_8 \]
Psychometric Theory: Data, Measurement, Theory

Error \rightarrow X \rightarrow \text{Latent X} \rightarrow \text{Latent Y} \rightarrow Y \rightarrow Error

Avoidance

Approach

Cog

Ability

\[ \delta_1 \rightarrow X_1 \]
\[ \delta_2 \rightarrow X_2 \]
\[ \delta_3 \rightarrow X_3 \]
\[ \delta_4 \rightarrow X_4 \]
\[ \delta_5 \rightarrow X_5 \]
\[ \delta_6 \rightarrow X_6 \]
\[ \delta_7 \rightarrow X_7 \]
\[ \delta_8 \rightarrow X_8 \]
\[ \delta_9 \rightarrow X_9 \]

\[ Y_1 \]
\[ Y_2 \]
\[ Y_3 \]
\[ Y_4 \]
\[ Y_5 \]
\[ Y_6 \]
\[ Y_7 \]
\[ Y_8 \]

\[ \epsilon_1 \]
\[ \epsilon_2 \]
\[ \epsilon_3 \]
\[ \epsilon_4 \]
\[ \epsilon_5 \]
\[ \epsilon_6 \]
\[ \epsilon_7 \]
\[ \epsilon_8 \]
Data = Model + Residual

• The fundamental equations of statistics are that
  • Data = Model + Residual
  • Residual = Data - Model
• The problem is to specify the model and then evaluate the fit of the model to the data as compared to other models
  • Fit = f(Data, Residual)
  • Typically: Fit = f\left(1 - \frac{\text{Residual}^2}{\text{Data}^2}\right)
  • Fit = f\left(1 - \frac{(\text{Data} - \text{Model})^2}{\text{Data}^2}\right)
• Even for something as simple as the mean is a model of the data. The residual left over after we remove the mean is the variance.
Psychometrics as model estimation and model fitting

We will explore a number of models

1. Modeling the process of data collection and of scaling
   - $X = f(\theta)$
   - How to measure $X$, properties of the function $f$

2. Correlation and Regression
   - $Y = \beta X$
   - $R_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$

3. Factor Analysis and Principal Components Analysis
   - $R = FF' + U^2 \quad R = CC'$

4. Reliability $\rho_{xx} = \frac{\sigma_\theta^2}{\sigma_X^2}$

5. Item Response Theory
   - $p(X|\theta, \delta) = f(\theta - \delta)$

6. Structural Equation Modeling
   - $\rho_{yy} Y = \beta \rho_{xx} X$
A theory of data and fundamentals of scaling

Error

X

Latent X

Latent Y

Y

Error

\( \delta_1 \)

\( X_1 \)

\( \chi_1 \)

\( \delta_2 \)

\( X_2 \)

\( \chi_2 \)

\( \delta_3 \)

\( X_3 \)

\( \chi_3 \)

\( \delta_4 \)

\( X_4 \)

\( \xi_1 \)

\( \delta_5 \)

\( X_5 \)

\( \xi_2 \)

\( \delta_6 \)

\( X_6 \)

\( \xi_3 \)

\( \delta_7 \)

\( X_7 \)

\( \xi_4 \)

\( \delta_8 \)

\( X_8 \)

\( \xi_5 \)

\( \delta_9 \)

\( X_9 \)

\( \xi_6 \)

\( \epsilon_1 \)

\( Y_1 \)

\( \epsilon_2 \)

\( Y_2 \)

\( \epsilon_3 \)

\( Y_3 \)

\( \epsilon_4 \)

\( Y_4 \)

\( \epsilon_5 \)

\( Y_5 \)

\( \epsilon_6 \)

\( Y_6 \)

\( \epsilon_7 \)

\( Y_7 \)

\( \epsilon_8 \)

\( Y_8 \)
Consider the following numbers, what do they represent?

**Table:** Numbers without context are meaningless. What do these numbers represent? Which of these numbers represent the same thing?

<table>
<thead>
<tr>
<th>Number 1</th>
<th>Number 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.71828</td>
<td>3.14159</td>
</tr>
<tr>
<td>24</td>
<td>86,400</td>
</tr>
<tr>
<td>37</td>
<td>98.7</td>
</tr>
<tr>
<td>365.25</td>
<td>365.25636305</td>
</tr>
<tr>
<td>31,557,600</td>
<td>31,558,150</td>
</tr>
<tr>
<td>3,412.1416</td>
<td>.4046856422</td>
</tr>
<tr>
<td>299,792,458</td>
<td>6.022141 * 10^{23}</td>
</tr>
<tr>
<td>42</td>
<td>X</td>
</tr>
</tbody>
</table>
Clyde Coombs and the Theory of Data

1. \( O = \) the set of objects
   - \( O = \{o_i, o_j \ldots o_n\} \)

2. \( S = \) the set of Individuals
   - \( S = \{s_i, s_j \ldots s_n\} \)

3. Two comparison operations
   - order ( \( x > y \) )
   - proximity ( \(|x - y| < \epsilon\) )

4. Two types of comparisons
   - Single dyads
     - \((s_i, s_j) (s_i, o_j) (o_i, o_j)\)
   - Pairs of dyads
     - \((s_i, s_j)(s_k, s_l) (s_i, o_j)(s_k, o_l) (o_i, o_j)(o_k, o_l)\)

Coombs (1964)
2 types of comparisons: Monotone ordering and single peak proximity

Order

Proximity

probability $A > B$

probability $|A - B| < \delta$
### Theory of Data and types of measures

**Table:** The theory of data provides a $3 \times 2 \times 2$ taxonomy for various types of measures

<table>
<thead>
<tr>
<th>Elements of Dyad</th>
<th>Number of Dyads</th>
<th>Comparison</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>People $\times$ People</td>
<td>1</td>
<td>Order</td>
<td>Tournament rankings</td>
</tr>
<tr>
<td>People $\times$ People</td>
<td>1</td>
<td>Proximity</td>
<td>Social Networks</td>
</tr>
<tr>
<td>Objects $\times$ Objects</td>
<td>1</td>
<td>Order</td>
<td>Similarities</td>
</tr>
<tr>
<td>Objects $\times$ Objects</td>
<td>1</td>
<td>Proximity</td>
<td>Ability Measurement</td>
</tr>
<tr>
<td>People $\times$ Objects</td>
<td>1</td>
<td>Order</td>
<td>Attitude Measurement</td>
</tr>
<tr>
<td>People $\times$ People</td>
<td>2</td>
<td>Proximity</td>
<td>Tournament rankings</td>
</tr>
<tr>
<td>People $\times$ People</td>
<td>2</td>
<td>Proximity</td>
<td>Social Networks</td>
</tr>
<tr>
<td>Objects $\times$ Objects</td>
<td>2</td>
<td>Order</td>
<td>Multidimensional scaling</td>
</tr>
<tr>
<td>Objects $\times$ Objects</td>
<td>2</td>
<td>Proximity</td>
<td>Ability Comparisons</td>
</tr>
<tr>
<td>People $\times$ Objects</td>
<td>2</td>
<td>Order</td>
<td>Preferential Choice</td>
</tr>
<tr>
<td>People $\times$ Objects</td>
<td>2</td>
<td>Proximity</td>
<td>Individual Differences in Multidimensional Scaling</td>
</tr>
</tbody>
</table>
Tournaments to order people (or teams)

1. Goal is to order the players by outcome to predict future outcomes

2. Complete Round Robin comparisons
   - Everyone plays everyone
   - Requires $N \times (N - 1)/2$ matches
   - How do you scale the results?

3. Partial Tournaments – Seeding and group play
   - World Cup
   - NCAA basketball
   - Is the winner really the best?
   - Can you predict other matches
Friendship as proximity

1. Chess or football provides a ranking based upon an ordering relationship \((p_i > p_j)\).

2. Alternatively, friendship groups are based upon closeness \((|p_i - p_j| < \delta)\)
   2.1 Do you know person j?
   2.2 Do you like person j? or as an alternative:
   2.3 Please list all your friends in this class (and is j included on the list)
   2.4 Would you be interested in having a date with person j?
   2.5 Would you like to have sex with person j?
   2.6 Would you marry person j?

3. Typically such data will be a rectangular matrix for there are asymmetries in closeness.
Moh’s hardness scale provides rank orders of hardness

Table: Mohs’ scale of mineral hardness. An object is said to be harder than X if it scratches X. Also included are measures of relative hardness using a sclerometer (for the hardest of the planes if there is a anisotropy or variation between the planes) which shows the non-linearity of the Mohs scale (Burchard, 2004).

<table>
<thead>
<tr>
<th>Mohs Hardness</th>
<th>Mineral</th>
<th>Scratch hardness</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Talc</td>
<td>0.59</td>
</tr>
<tr>
<td>2</td>
<td>Gypsum</td>
<td>0.61</td>
</tr>
<tr>
<td>3</td>
<td>Calcite</td>
<td>3.44</td>
</tr>
<tr>
<td>4</td>
<td>Fluorite</td>
<td>3.05</td>
</tr>
<tr>
<td>5</td>
<td>Apapatite</td>
<td>5.2</td>
</tr>
<tr>
<td>6</td>
<td>Orthoclase Feldspar</td>
<td>37.2</td>
</tr>
<tr>
<td>7</td>
<td>Quartz</td>
<td>100</td>
</tr>
<tr>
<td>8</td>
<td>Topaz</td>
<td>121</td>
</tr>
<tr>
<td>9</td>
<td>Corundum</td>
<td>949</td>
</tr>
<tr>
<td>10</td>
<td>Diamond</td>
<td>85,300</td>
</tr>
</tbody>
</table>
Ordering based upon external measures

Table: The Beaufort scale of wind intensity is an early example of a scale with roughly equal units that is observationally based. Although the units are roughly in equal steps of wind speed in nautical miles/hour (knots), the force of the wind is not linear with this scale, but rather varies as the square of the velocity.

<table>
<thead>
<tr>
<th>Force</th>
<th>Wind (Knots)</th>
<th>WMO Classification</th>
<th>Appearance of Wind Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Less than 1</td>
<td>Calm</td>
<td>Sea surface smooth and mirror-like</td>
</tr>
<tr>
<td>1</td>
<td>1-3</td>
<td>Light Air</td>
<td>Scaly ripples, no foam crests</td>
</tr>
<tr>
<td>2</td>
<td>4-6</td>
<td>Light Breeze</td>
<td>Small wavelets, crests glassy, no breaking</td>
</tr>
<tr>
<td>3</td>
<td>7-10</td>
<td>Gentle Breeze</td>
<td>Large wavelets, crests begin to break, scattered whitecaps</td>
</tr>
<tr>
<td>4</td>
<td>11-16</td>
<td>Moderate Breeze</td>
<td>Small waves 1-4 ft. becoming longer, numerous whitecaps</td>
</tr>
<tr>
<td>5</td>
<td>17-21</td>
<td>Fresh Breeze</td>
<td>Moderate waves 4-8 ft taking longer form, many whitecaps, some spray</td>
</tr>
<tr>
<td>6</td>
<td>22-27</td>
<td>Strong Breeze</td>
<td>Larger waves 8-13 ft, whitecaps common more spray</td>
</tr>
<tr>
<td>7</td>
<td>28-33</td>
<td>Near Gale</td>
<td>Sea heaps up, waves 13-20 ft, white foam streaks off breakers</td>
</tr>
<tr>
<td>8</td>
<td>34-40</td>
<td>Gale Moderately</td>
<td>high (13-20 ft) waves of greater length, edges of crests begin to break into spindrift, foam blown in streaks</td>
</tr>
<tr>
<td>9</td>
<td>41-47</td>
<td>Strong Gale</td>
<td>High waves (20 ft), sea begins to roll, dense streaks of foam, spray may reduce visibility</td>
</tr>
<tr>
<td>10</td>
<td>48-55</td>
<td>Storm</td>
<td>Very high waves (20-30 ft) with overhanging crests, sea white with densely blown foam, heavy rolling, lowered visibility</td>
</tr>
<tr>
<td>11</td>
<td>56-63</td>
<td>Violent Storm</td>
<td>Exceptionally high (30-45 ft) waves, foam patches cover sea, visibility more reduced</td>
</tr>
<tr>
<td>12</td>
<td>64+</td>
<td>Hurricane</td>
<td>Air filled with foam, waves over 45 ft, sea completely white with driving spray, visibility greatly reduced</td>
</tr>
</tbody>
</table>
Models of scaling objects

1. Assume each object \((a, b, ...z)\) has a scale value \((A, B, ...Z)\) with some noise for each measurement.

2. Probability of \(A > B\) increases with difference between \(a\) and \(b\).

3. \(P(A > B) = f(a - b)\)

4. Can we find a function, \(f\), such that equal differences in the latent variable \((a, b, c)\) lead to equal differences in the observed variable?

5. Several alternatives
   - Direct scaling on some attribute dimension (simple but flawed)
   - Indirect scaling by paired comparisons (more complicated but probably better)
Scaling of Objects: \( O \times O \) comparisons

1. Typical object scaling is concerned with order or location of objects

2. Subjects are assumed to be random replicates of each other, differing only as a source of noise

3. Absolute scaling techniques
   - Grant Proposals: 1 to 5
   - "On a scale from 1 to 10" this [object] is a X?
   - If A is 1 and B is 10, then what is C?
   - College rankings based upon selectivity
   - College rankings based upon "yield"
   - Zagat ratings of restaurants
   - A - F grading of papers
Absolute scaling: difficulties

1. "On a scale from 1 to 10" this [object] is a X?
   - sensitive to context effects
   - what if a new object appears?
   - Need unbounded scale

2. If A is 1 and B is 10, then what is C?
   - results will depend upon A, B
**Absolute scaling: artifacts**

1. College rankings based upon selectivity
   - accept/applied
   - encourage less able to apply

2. College rankings based upon "yield"
   - matriculate/accepted
   - early admissions guarantee matriculation
   - don’t accept students who will not attend

3. Proposed solution: college choice as a tournament
   - Consider all schools that accept a student
   - Which school does he/she choose?

* Avery, Glickman, Hoxby & Metrick (2013)
A revealed preference ordering Avery et al. (2013)
A revealed preference ordering Avery et al. (2013)

### Revealed Preference Ranking of Colleges Based on Matriculation Decisions

<table>
<thead>
<tr>
<th>Rank</th>
<th>College Name</th>
<th>Implied Prob. of &quot;Winning&quot; vs. College Listed</th>
<th>Rank Based on Matriculation (no Covariates)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Harvard University</td>
<td>0.93</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Caltech</td>
<td>0.92</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Yale University</td>
<td>0.92</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>MIT</td>
<td>0.89</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>Stanford University</td>
<td>0.90</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>Princeton University</td>
<td>0.90</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>Brown University</td>
<td>0.78</td>
<td>7</td>
</tr>
<tr>
<td>8</td>
<td>Columbia University</td>
<td>0.73</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>Amherst College</td>
<td>0.71</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>Dartmouth</td>
<td>0.72</td>
<td>10</td>
</tr>
<tr>
<td>11</td>
<td>Wellesley College</td>
<td>0.71</td>
<td>12</td>
</tr>
<tr>
<td>12</td>
<td>University of Pennsylvania</td>
<td>0.71</td>
<td>11</td>
</tr>
</tbody>
</table>
Weber-Fechner Law and non-linearity of scales

1. Early studies of psychophysics by Weber (1834b,a) and subsequently Fechner (1860) demonstrated that the human perceptual system does not perceive stimulus intensity as a linear function of the physical input.

2. The basic paradigm was to compare one weight with another that differed by amount $\Delta$, e.g., compare a 10 gram weight with an 11, 12, and 13 gram weight, or a 10 kg weight with a 11, 12, or 13 kg weight.

3. What was the $\Delta$ that was just detectable? The finding was that the perceived intensity follows a logarithmic function.

4. Examining the magnitude of the “just noticeable difference” or $JND$, Weber (1834b) found that

$$JND = \frac{\Delta Intensity}{Intensity} = constant.$$  (1)
Weber-Fechner Law and non-linearity of scales

1. An example of a logarithmic scale of intensity is the decibel measure of sound intensity.

2. Sound Pressure Level expressed in decibels (dB) of the root mean square observed sound pressure, $P_o$ (in Pascals) is

   $$L_p = 20 \log_{10} \frac{P_o}{P_{\text{ref}}}$$

3. where the reference pressure, $P_{\text{ref}}$, in the air is $20 \mu Pa$.

4. Just to make this confusing, the reference pressure for sound measured in the ocean is $1 \mu Pa$. This means that sound intensities in the ocean are expressed in units that are 20 dB higher than those units used on land.
The Just Noticeable Difference in Person perception

1. Although typically thought of as just relevant for the perceptual experiences of physical stimuli, Ozer (1993) suggested that the JND is useful in personality assessment as a way of understanding the accuracy and inter judge agreement of judgments about other people.

2. In addition, Sinn (2003) has argued that the logarithmic nature of the Weber-Fechner Law is of evolutionary significance for preference for risk and cites Bernoulli (1738) as suggesting that our general utility function is logarithmic.
Money and non linearity

... the utility resulting from any small increase in wealth will be inversely proportionate to the quantity of goods already possessed .... if ... one has a fortune worth a hundred thousand ducats and another one a fortune worth same number of semi-ducats and if the former receives from it a yearly income of five thousand ducats while the latter obtains the same number of semi-ducats, it is quite clear that to the former a ducat has exactly the same significance as a semi-ducat to the latter (Bernoulli, 1738, p 25).

Implies a log function for utility.
Econs and Humans

1. Simple expected value theory \( \implies \text{value} = \text{probability of event} \times \text{value of event} \)

2. Bernoulli theory of expected utility came to dominate choice theory and is fundamental to economics

3. Studied by comparing gambles and showing utility is non-linear with value
   - Would you rather have $80 or a 80% chance of $100 + 20% of $10?
   - expected value is 80 versus \(.8 \times 100 + .2 \times 10 = 82\)

4. Bernoulli value (from Kahneman, 2011)
   
   \[
   \begin{array}{cccccccccc}
   \text{Wealth (millions)} & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\
   \text{Utility units} & 10 & 30 & 48 & 60 & 70 & 78 & 84 & 90 & 96 & 100
   \end{array}
   \]
Kahneman and Tversky: Prospect Theory

Losses are more painful than gains are pleasant

Kahneman & Tversky (1979)
Better to skip lunch than be someone’s dinner.
## Four types of scales and their associated statistics

**Table:** Four types of scales and their associated statistics ([Rossi, 2007; Stevens, 1946](#)) The statistics listed for a scale are invariant for that type of transformation.

<table>
<thead>
<tr>
<th>Scale</th>
<th>Basic operations</th>
<th>Transformations</th>
<th>Invariant statistic</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nominal</td>
<td>equality</td>
<td>Permutations</td>
<td>Counts</td>
<td>Detection</td>
</tr>
<tr>
<td></td>
<td>( x_i = x_j )</td>
<td></td>
<td>Mode</td>
<td>Species classification</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>( \chi^2 ) and (( \phi )) correlation</td>
<td>Taxons</td>
</tr>
<tr>
<td>Ordinal</td>
<td>order</td>
<td>Monotonic (homeomorphic)</td>
<td>Median</td>
<td>Mhos Hardness scale</td>
</tr>
<tr>
<td></td>
<td>( x_i &gt; x_j )</td>
<td>( x' = f(x) )</td>
<td>Percentiles</td>
<td>Beaufort Wind (intensity)</td>
</tr>
<tr>
<td>Interval</td>
<td>differences</td>
<td>Linear (Affine)</td>
<td>Mean (( \mu ))</td>
<td>Temperature (( ^\circ F, ^\circ C ))</td>
</tr>
<tr>
<td></td>
<td>( (x_i - x_j) &gt; (x_k - x_l) )</td>
<td>( x' = a + bx )</td>
<td>Standard Deviation (( \sigma ))</td>
<td>Beaufort Wind (velocity)</td>
</tr>
<tr>
<td>Ratio</td>
<td>ratios</td>
<td>Multiplication (Similarity)</td>
<td>Pearson correlation (( r ))</td>
<td>Length, mass, time</td>
</tr>
<tr>
<td></td>
<td>( \frac{x_i}{x_j} &gt; \frac{x_k}{x_l} )</td>
<td>( x' = bx )</td>
<td>Regression (( \beta ))</td>
<td>Temperature (( ^\circ K ))</td>
</tr>
</tbody>
</table>

The Beaufort wind speed scale is interval with respect to the velocity of the wind, but only ordinal with respect to the effect of the wind. The Richter scale of earthquake intensity is a logarithmic scale of the energy released but...
Graphical and tabular summaries of data

1. The Tukey 5 number summary shows the important characteristics of a set of numbers
   - Maximum
   - 75th percentile
   - Median (50th percentile)
   - 25th percentile
   - Minimum

2. Graphically, this is the box plot
   - Variations on the box plot include confidence intervals for the median
The summary command gives the Tukey 5 numbers

```r
> summary(sat.act)
```

<table>
<thead>
<tr>
<th></th>
<th>gender</th>
<th>education</th>
<th>age</th>
<th>ACT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>1.000</td>
<td>0.000</td>
<td>13.00</td>
<td>3.00</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>1.000</td>
<td>3.000</td>
<td>19.00</td>
<td>25.00</td>
</tr>
<tr>
<td>Median</td>
<td>2.000</td>
<td>3.000</td>
<td>22.00</td>
<td>29.00</td>
</tr>
<tr>
<td>Mean</td>
<td>1.647</td>
<td>3.164</td>
<td>25.59</td>
<td>28.55</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>2.000</td>
<td>4.000</td>
<td>29.00</td>
<td>32.00</td>
</tr>
<tr>
<td>Max.</td>
<td>2.000</td>
<td>5.000</td>
<td>65.00</td>
<td>36.00</td>
</tr>
</tbody>
</table>
A box plot of the first 4 sat.act variables

A Tukey Boxplot

```
boxplot(sat.act[1:4], main = "A Tukey Boxplot")
```
A violin or density plot of the first 5 epi.bfi variables

Density plot

```
violinBy(epi.bfi[1:5], main = "A Tukey violin plot")
```
The `describe` function gives more descriptive statistics

```r
> describe(sat.act)
```

<table>
<thead>
<tr>
<th>vars</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>trimmed</th>
<th>mad</th>
<th>min</th>
<th>max</th>
<th>range</th>
<th>skew</th>
<th>kurtosis</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender</td>
<td>1</td>
<td>1.65</td>
<td>0.48</td>
<td>2</td>
<td>1.68</td>
<td>0.00</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-0.61</td>
<td>-1.62</td>
<td>0.02</td>
</tr>
<tr>
<td>education</td>
<td>2</td>
<td>3.16</td>
<td>1.43</td>
<td>3</td>
<td>3.31</td>
<td>1.48</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>-0.68</td>
<td>-0.07</td>
<td>0.05</td>
</tr>
<tr>
<td>age</td>
<td>3</td>
<td>25.59</td>
<td>9.50</td>
<td>22</td>
<td>23.86</td>
<td>5.93</td>
<td>13</td>
<td>65</td>
<td>52</td>
<td>1.64</td>
<td>2.42</td>
<td>0.36</td>
</tr>
<tr>
<td>ACT</td>
<td>4</td>
<td>28.55</td>
<td>4.82</td>
<td>29</td>
<td>28.84</td>
<td>4.45</td>
<td>3</td>
<td>36</td>
<td>33</td>
<td>-0.66</td>
<td>0.53</td>
<td>0.18</td>
</tr>
<tr>
<td>SATV</td>
<td>5</td>
<td>612.23</td>
<td>112.90</td>
<td>620</td>
<td>619.45</td>
<td>118.61</td>
<td>200</td>
<td>800</td>
<td>600</td>
<td>-0.64</td>
<td>0.33</td>
<td>4.27</td>
</tr>
<tr>
<td>SATQ</td>
<td>6</td>
<td>610.22</td>
<td>115.64</td>
<td>620</td>
<td>617.25</td>
<td>118.61</td>
<td>200</td>
<td>800</td>
<td>600</td>
<td>-0.59</td>
<td>-0.02</td>
<td>4.41</td>
</tr>
</tbody>
</table>
Multiple measures of central tendency

**mode**  The most frequent observation. Not a very stable measure, depends upon grouping. Can be used for categorical data.

**median**  The number with 50% above and 50% below. A powerful, if underused, measure. Not sensitive to transforms of the shape of the distribution, nor outliers. Appropriate for ordinal data, and useful for interval data.

**mean**  One of at least seven measures that assume interval properties of the data.
Multiple ways to estimate the mean

Arithmetic mean \( \bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \) \( \text{mean}(x) \)

Trimmed mean throws away the top and bottom t% of observations. This follows the principle that all data are normal at the middle. \( \text{mean}(x, \text{trim}=.1) \)

Winsorized mean Find the arithmetic mean after replacing the n lowest observations with the nth value, and the N largest values with the Nth largest. \( \text{winsor}(x, \text{trim}=.2) \)

Geometric Mean \( \bar{X}_{\text{geometric}} = \sqrt[N]{\prod_{i=1}^{N} X_i} = e^{\frac{\sum(ln(x))}{N}} \) (The anti-log of the mean log score). \( \text{geometric.mean}(x) \)

Harmonic Mean \( \bar{X}_{\text{harmonic}} = \frac{N}{\sum_{i=1}^{N} \frac{1}{X_i}} \) (The reciprocal of the mean reciprocal). \( \text{harmonic.mean}(x) \)

Circular Mean \( \bar{X}_{\text{circular}} = \tan^{-1} \left( \frac{\sum \cos(x)}{\sum \sin(x)} \right) \) \( \text{circular.mean}(x) \) (where x is in radians)

\( \text{circadian.mean} \) \( \text{circular.mean}(x) \) (where x is in hours)
Class size from the students’ point of view.

Table: Class size from the students’ point of view. Most students are in large classes; the median class size is 200 with a mean of 223.

<table>
<thead>
<tr>
<th>Class size</th>
<th>Number of classes</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>12</td>
<td>120</td>
</tr>
<tr>
<td>20</td>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td>100</td>
<td>2</td>
<td>200</td>
</tr>
<tr>
<td>200</td>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>400</td>
<td>1</td>
<td>400</td>
</tr>
</tbody>
</table>
Time in therapy

A psychotherapist is asked what is the average length of time that a patient is in therapy. This seems to be an easy question, for of the 20 patients, 19 have been in therapy for between 6 and 18 months (with a median of 12) and one has just started. Thus, the median client is in therapy for 52 weeks with an average (in weeks) \((1 \times 1 + 19 \times 52)/20\) or 49.4.

However, a more careful analysis examines the case load over a year and discovers that indeed, 19 patients have a median time in treatment of 52 weeks, but that each week the therapist is also seeing a new client for just one session. That is, over the year, the therapist sees 52 patients for 1 week and 19 for a median of 52 weeks. Thus, the median client is in therapy for 1 week and the average client is in therapy of \((52 \times 1 + 19 \times 52)/(52+19) = 14.6\) weeks.
Does teaching effect learning?

1. A leading research team in motivational and educational psychology was interested in the effect that different teaching techniques at various colleges and universities have upon their students. They were particularly interested in the effect upon writing performance of attending a very selective university, a less selective university, or a two year junior college.

2. A writing test was given to the entering students at three institutions in the Boston area. After one year, a similar writing test was given again. Although there was some attrition from each sample, the researchers report data only for those who finished one year. The pre and post test scores as well as the change scores were as shown below:
Types of teaching affect student outcomes?

**Table:** Three types of teaching and their effect on student outcomes

<table>
<thead>
<tr>
<th>School</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior College</td>
<td>1</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Non-selective university</td>
<td>5</td>
<td>27</td>
<td>22</td>
</tr>
<tr>
<td>Selective university</td>
<td>27</td>
<td>73</td>
<td>45</td>
</tr>
</tbody>
</table>

From these data, the researchers concluded that the quality of teaching at the selective university was much better than that of the less selective university or the junior college and that the students learned a great deal more. They proposed to study the techniques used there in order to apply them to the other institutions.
Teaching and math performance

Another research team in motivational and educational psychology was interested in the effect that different teaching at various colleges and universities affect math performance. They used the same schools as the previous example with the same design.

Table: Three types of teaching and their effect on student outcomes

<table>
<thead>
<tr>
<th>School</th>
<th>Pretest</th>
<th>Posttest</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Junior College</td>
<td>27</td>
<td>73</td>
<td>45</td>
</tr>
<tr>
<td>Non-selective university</td>
<td>73</td>
<td>95</td>
<td>22</td>
</tr>
<tr>
<td>Selective university</td>
<td>95</td>
<td>99</td>
<td>4</td>
</tr>
</tbody>
</table>

They concluded that the teaching at the junior college was far superior to that of the select university. What is wrong with this conclusion?
Effect of teaching, effect of students, or just scaling?

**Writing**

- Performance vs. time in school for Ivy, TC, and JC.

**Math**

- Performance vs. time in school for Ivy, TC, and JC.
The problem of scaling is ubiquitous

1. A leading cognitive developmentalist believed that there is a critical stage for learning spatial representations using maps. Children younger than this stage are not helped by maps, nor are children older than this stage.

2. He randomly assigned 3rd, 5th, and 7th grade students into two conditions (nested within grade), control, and map use. Performance was measured on a task of spatial recall (children were shown toys at particular locations in a set of rooms and then asked to find them again later.) Half the children were shown a map of the rooms before doing the task.

3. Their scores were

<table>
<thead>
<tr>
<th></th>
<th>No Map</th>
<th>Maps</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>3rd grade</td>
<td>5</td>
<td>27</td>
<td>22 Too young</td>
</tr>
<tr>
<td>5th grade</td>
<td>27</td>
<td>73</td>
<td>46 Critical period</td>
</tr>
<tr>
<td>7th grade</td>
<td>73</td>
<td>95</td>
<td>22 Too old</td>
</tr>
</tbody>
</table>
Map use is most effective at a particular developmental stage

Recall varies by age and exposure to maps

Recall

- maps
- nomaps

grade

3 4 5 6 7
0 20 40 60 80 100
Correlation and Regression

1. Developed in 1886 by Francis Galton
   • Further developments by Karl Pearson and Charles Spearman
2. Correlation/regression are the root concept of psychometrics
   • Other statistics, including factor analysis are ways of partitioning correlation matrices
   • Reliability theory is merely an application of factor analysis
Francis Galton 1822-1911

Francis Galton (1822-1911) was among the most influential psychologists of the 19th century. He did pioneering work on the correlation coefficient, behavior genetics and the measurement of individual differences. He introspectively examined the question of free will and introduced the lexical hypothesis to the study of personality and character. In addition to psychology, he did pioneering work in meteorology and introduced the scientific use of fingerprints. Whenever he could, he counted.
Karl Pearson 1857-1936

Carl (Karl) Pearson was among the most influential statisticians of the early 20th century. Founder of the statistics department at University College London. He developed the Pearson Product Moment Correlation Coefficient, its special case the $\phi$ coefficient, and the tetrachoric correlation. Major behavior geneticist and eugenicist.
Charles Spearman 1863-1945

Charles Spearman (1863-1945) was the leading psychometrician of the early 20th century. His work on the classical test theory, factor analysis, and the g theory of intelligence continues to influence psychometrics, statistics, and the study of intelligence. More than 100 years after their publication, his most influential papers remain two of the most frequently cited articles in psychometrics and intelligence.
Galton’s height data

Table: The relationship between the average of both parents (mid parent) and the height of their children. The basic data table is from Galton (1886) who used these data to introduce reversion to the mean (and thus, linear regression). The data are available as part of the UsingR or psych packages.

```r
> library(psych)
> data(galton)
> galton.tab <- table(galton)
> galton.tab[order(rank(rownames(galton.tab)),decreasing=TRUE),]
```

<table>
<thead>
<tr>
<th>child</th>
<th>parent</th>
<th>61.7</th>
<th>62.2</th>
<th>63.2</th>
<th>64.2</th>
<th>65.2</th>
<th>66.2</th>
<th>67.2</th>
<th>68.2</th>
<th>69.2</th>
<th>70.2</th>
<th>71.2</th>
<th>72.2</th>
<th>73.2</th>
<th>73.7</th>
</tr>
</thead>
<tbody>
<tr>
<td>73</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>72.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>71.5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>5</td>
<td>10</td>
<td>4</td>
<td>9</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>70.5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>12</td>
<td>18</td>
<td>14</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>69.5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>16</td>
<td>4</td>
<td>17</td>
<td>27</td>
<td>20</td>
<td>33</td>
<td>25</td>
<td>20</td>
<td>11</td>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>68.5</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>11</td>
<td>16</td>
<td>25</td>
<td>31</td>
<td>34</td>
<td>48</td>
<td>21</td>
<td>18</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>67.5</td>
<td>0</td>
<td>3</td>
<td>5</td>
<td>14</td>
<td>15</td>
<td>36</td>
<td>38</td>
<td>28</td>
<td>38</td>
<td>19</td>
<td>11</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>66.5</td>
<td>0</td>
<td>3</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>17</td>
<td>17</td>
<td>14</td>
<td>13</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>65.5</td>
<td>1</td>
<td>0</td>
<td>9</td>
<td>5</td>
<td>7</td>
<td>11</td>
<td>11</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>64.5</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>64</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
Galton’s height data

Figure: Galton’s data can be plotted to show the relationships between mid parent and child heights. Because the original data are grouped, the data points have been jittered to emphasize the density of points along the median. The bars connect the first, 2nd (median) and third quartiles. The dashed line is the best fitting linear fit, the ellipses represent one and two standard deviations from the mean.
Bivariate Regression

\[ \hat{y} = \beta_{y.x} x + \epsilon \]

\[ \beta_{y.x} = \frac{\sigma_{xy}}{\sigma_x^2} \]
Bivariate Regression

\[ \hat{y} = \beta_{y,x} x + \epsilon \]

\[ \beta_{y,x} = \frac{\sigma_{xy}}{\sigma_x^2} \]

\[ \hat{x} = \beta_{x,y} y + \delta \]

\[ \beta_{y,x} = \frac{\sigma_{xy}}{\sigma_y^2} \]
Bivariate Correlation is the geometric average of the two regressions

\[
\hat{x} = \beta_{x,y} y + \delta \\
\hat{y} = \beta_{y,x} x + \epsilon
\]

\[
\beta_{y,x} = \frac{\sigma_{xy}}{\sigma_y^2} \\
\beta_{y,x} = \frac{\sigma_{xy}}{\sigma_x^2}
\]

\[
r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}}
\]
The variance and the variance of a composite

1. If \( \mathbf{x}_1 \) and \( \mathbf{x}_2 \) are vectors of \( N \) observations centered around their mean (that is, deviation scores) their variances are \( V_{x_1} = \sum x^2_i / (N - 1) \) and \( V_{x_2} = \sum x^2_i / (N - 1) \), or, in matrix terms \( V_{x_1} = \mathbf{x}_1' \mathbf{x}_1 / (N - 1) \) and \( V_{x_2} = \mathbf{x}_2' \mathbf{x}_2 / (N - 1) \).

2. The variance of the composite made up of the sum of the corresponding scores, \( \mathbf{x} + \mathbf{y} \) is just

\[
V_{(x_1+x_2)} = \frac{\sum(x_i + y_i)^2}{N - 1} = \frac{\sum x^2_i + \sum y^2_i + 2 \sum x_i y_i}{N - 1} = \frac{(\mathbf{x} + \mathbf{y})'(\mathbf{x} + \mathbf{y})}{N - 1}. \quad (3)
\]

Or, more generally,

\[
\mathbf{S} = \begin{pmatrix}
V_{x_1} & C_{x_1x_2} & \cdots & C_{x_1x_n} \\
C_{x_1x_2} & V_{x_2} & \cdots & C_{x_2x_n} \\
\vdots & \vdots & \ddots & \vdots \\
C_{x_1x_n} & C_{x_2x_n} & \cdots & V_{xn}
\end{pmatrix}
\]
Sums as matrix products

\[ V_X = \sum \frac{X'X}{N - 1} = \frac{1'(X'X)1}{N - 1} . \]

\[ V_Y = \sum \frac{Y'Y}{N - 1} = \frac{1'(Y'Y)1}{N - 1} . \]

\[ C_{XY} = \sum \frac{X'Y}{N - 1} = \frac{1'(X'Y)1}{N - 1} . \]
Get the data from a remote data source

A nice feature of R is that you can read from remote data sets. The example dataset is on the personality-project.org server. Get it and describe it.

```r
> datafilename="http://personality-project.org/r/datasets/psychometrics.prob2.txt"
> mydata =read.table(datafilename,header=TRUE) #read the data file
> describe(mydata,skew=FALSE)
```

<table>
<thead>
<tr>
<th></th>
<th>var</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>trimmed</th>
<th>mad</th>
<th>min</th>
<th>max</th>
<th>range</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>1</td>
<td>1000</td>
<td>500.50</td>
<td>288.82</td>
<td>500.50</td>
<td>500.50</td>
<td>370.65</td>
<td>1.0</td>
<td>1000.00</td>
<td>999.00</td>
<td>9.13</td>
</tr>
<tr>
<td>GREV</td>
<td>2</td>
<td>1000</td>
<td>499.77</td>
<td>106.11</td>
<td>497.50</td>
<td>498.75</td>
<td>106.01</td>
<td>138.0</td>
<td>873.00</td>
<td>735.00</td>
<td>3.36</td>
</tr>
<tr>
<td>GREQ</td>
<td>3</td>
<td>1000</td>
<td>500.53</td>
<td>103.85</td>
<td>498.00</td>
<td>498.51</td>
<td>105.26</td>
<td>191.0</td>
<td>914.00</td>
<td>723.00</td>
<td>3.28</td>
</tr>
<tr>
<td>GREA</td>
<td>4</td>
<td>1000</td>
<td>498.13</td>
<td>100.45</td>
<td>495.00</td>
<td>498.67</td>
<td>99.33</td>
<td>207.0</td>
<td>848.00</td>
<td>641.00</td>
<td>3.18</td>
</tr>
<tr>
<td>Ach</td>
<td>5</td>
<td>1000</td>
<td>49.93</td>
<td>9.84</td>
<td>50.00</td>
<td>49.88</td>
<td>10.38</td>
<td>16.0</td>
<td>79.00</td>
<td>63.00</td>
<td>0.31</td>
</tr>
<tr>
<td>Anx</td>
<td>6</td>
<td>1000</td>
<td>50.32</td>
<td>9.91</td>
<td>50.00</td>
<td>50.43</td>
<td>10.38</td>
<td>14.0</td>
<td>78.00</td>
<td>64.00</td>
<td>0.31</td>
</tr>
<tr>
<td>Prelim</td>
<td>7</td>
<td>1000</td>
<td>10.03</td>
<td>1.06</td>
<td>10.00</td>
<td>10.02</td>
<td>1.48</td>
<td>7.0</td>
<td>13.00</td>
<td>6.00</td>
<td>0.03</td>
</tr>
<tr>
<td>GPA</td>
<td>8</td>
<td>1000</td>
<td>4.00</td>
<td>0.50</td>
<td>4.02</td>
<td>4.01</td>
<td>0.53</td>
<td>2.5</td>
<td>5.38</td>
<td>2.88</td>
<td>0.02</td>
</tr>
<tr>
<td>MA</td>
<td>9</td>
<td>1000</td>
<td>3.00</td>
<td>0.49</td>
<td>3.00</td>
<td>3.00</td>
<td>0.44</td>
<td>1.4</td>
<td>4.50</td>
<td>3.10</td>
<td>0.02</td>
</tr>
</tbody>
</table>
**Plot it using the `pairs.panels` function.**

Use the `pairs.panels` function to show a splom plot (use `gap=0` and `pch='.'`).

```r
> pairs.panels(mydata,pch=".",gap=0)  # pch='.' makes for a cleaner plot
```

![Splom Plot](image)
Plot a subset of the data using the \( \texttt{c()} \) function (concatenate).

Use the \texttt{pairs.panels} function to show a splom plot. Select a subset of variables using the \texttt{c()} function.

\[
\texttt{pairs.panels(mydata[c(2:4,6:8)],pch='.')}
\]
> pairs.panels(mydata[mydata$ID < 200,c(2:4,6:8)])
0 center the data

In order to do interaction terms in regressions, it is necessary to 0 center the data. We need to turn the result into a data.frame in order to use it in the regression function.

```r
> cent <- data.frame(scale(mydata,scale=FALSE))
> describe(cent,skew=FALSE)
```

<table>
<thead>
<tr>
<th></th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>trimmed</th>
<th>mad</th>
<th>min</th>
<th>max</th>
<th>range</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>1</td>
<td>1000</td>
<td>0 288.82</td>
<td>0.00</td>
<td>0.00</td>
<td>370.65</td>
<td>-499.50</td>
<td>499.50</td>
<td>999.00</td>
<td>9.13</td>
</tr>
<tr>
<td>GREV</td>
<td>2</td>
<td>1000</td>
<td>0 106.11</td>
<td>-2.27</td>
<td>-1.02</td>
<td>106.01</td>
<td>-361.77</td>
<td>373.23</td>
<td>735.00</td>
<td>3.36</td>
</tr>
<tr>
<td>GREQ</td>
<td>3</td>
<td>1000</td>
<td>0 103.85</td>
<td>-2.53</td>
<td>-2.02</td>
<td>105.26</td>
<td>-309.53</td>
<td>413.47</td>
<td>723.00</td>
<td>3.28</td>
</tr>
<tr>
<td>GREA</td>
<td>4</td>
<td>1000</td>
<td>0 100.45</td>
<td>-3.13</td>
<td>0.54</td>
<td>99.33</td>
<td>-291.13</td>
<td>349.87</td>
<td>641.00</td>
<td>3.18</td>
</tr>
<tr>
<td>Ach</td>
<td>5</td>
<td>1000</td>
<td>0 9.84</td>
<td>0.07</td>
<td>-0.05</td>
<td>10.38</td>
<td>-33.93</td>
<td>29.07</td>
<td>63.00</td>
<td>0.31</td>
</tr>
<tr>
<td>Anx</td>
<td>6</td>
<td>1000</td>
<td>0 9.91</td>
<td>-0.32</td>
<td>0.11</td>
<td>10.38</td>
<td>-36.32</td>
<td>27.68</td>
<td>64.00</td>
<td>0.31</td>
</tr>
<tr>
<td>Prelim</td>
<td>7</td>
<td>1000</td>
<td>0 1.06</td>
<td>-0.03</td>
<td>0.00</td>
<td>1.48</td>
<td>-3.03</td>
<td>2.97</td>
<td>6.00</td>
<td>0.03</td>
</tr>
<tr>
<td>GPA</td>
<td>8</td>
<td>1000</td>
<td>0 0.50</td>
<td>0.02</td>
<td>0.00</td>
<td>0.53</td>
<td>-1.50</td>
<td>1.38</td>
<td>2.88</td>
<td>0.02</td>
</tr>
<tr>
<td>MA</td>
<td>9</td>
<td>1000</td>
<td>0 0.49</td>
<td>0.00</td>
<td>0.00</td>
<td>0.44</td>
<td>-1.60</td>
<td>1.50</td>
<td>3.10</td>
<td>0.02</td>
</tr>
</tbody>
</table>

The standard deviations and ranges have not changed. However, the means are all 0. We use the scale function with the scale=FALSE option.
The standardized data

Alternatively, we could standardize it.

```r
> z.data <- data.frame(scale(my.data))
> describe(z.data)
```

<table>
<thead>
<tr>
<th>var</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>median trimmed mad</th>
<th>min</th>
<th>max</th>
<th>range</th>
<th>skew</th>
<th>kurtosis</th>
<th>se</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>1.28</td>
<td>-1.73</td>
<td>1.73</td>
<td>3.46</td>
<td>0.03</td>
</tr>
<tr>
<td>GREV</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>-0.02</td>
<td>-0.01</td>
<td>1.00</td>
<td>-3.41</td>
<td>3.52</td>
<td>6.93</td>
<td>0.09</td>
</tr>
<tr>
<td>GREQ</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>-0.02</td>
<td>-0.02</td>
<td>1.01</td>
<td>-2.98</td>
<td>3.98</td>
<td>6.96</td>
<td>0.08</td>
</tr>
<tr>
<td>GREA</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>-0.03</td>
<td>0.01</td>
<td>0.99</td>
<td>-2.90</td>
<td>3.48</td>
<td>6.38</td>
<td>-0.06</td>
</tr>
<tr>
<td>Ach</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0.01</td>
<td>-0.01</td>
<td>1.05</td>
<td>-3.45</td>
<td>2.95</td>
<td>6.40</td>
<td>0.02</td>
</tr>
<tr>
<td>Anx</td>
<td>6</td>
<td>1</td>
<td>0</td>
<td>-0.03</td>
<td>0.01</td>
<td>1.05</td>
<td>-3.67</td>
<td>2.79</td>
<td>6.46</td>
<td>-0.14</td>
</tr>
<tr>
<td>Prelim</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>-0.02</td>
<td>0.00</td>
<td>1.40</td>
<td>-2.86</td>
<td>2.81</td>
<td>5.67</td>
<td>-0.01</td>
</tr>
<tr>
<td>GPA</td>
<td>8</td>
<td>1</td>
<td>0</td>
<td>0.03</td>
<td>0.01</td>
<td>1.06</td>
<td>-3.00</td>
<td>2.74</td>
<td>5.74</td>
<td>-0.29</td>
</tr>
<tr>
<td>MA</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0.90</td>
<td>-3.23</td>
<td>3.04</td>
<td>6.27</td>
<td>-0.07</td>
</tr>
</tbody>
</table>

Or, we can standardize it by dividing though by the standard deviation. We use the `scale` function to do this for us.
Show how the correlations do not change with standardization

Find the correlations using the `lowerCor` function. This, by default, uses pairwise Pearson correlations and rounds to two decimals. Compare with the standard `cor` function.

```r
> lowerCor(my.data)

          ID   GREV  GREQ GREA  Ach Anx Prelm GPA  MA
ID  1.000
GREV -0.01 1.00
GREQ  0.00 0.73 1.00
GREA -0.01 0.64 0.60 1.00
Ach  0.00 0.01 0.01 0.45 1.00
Anx  0.00 0.01 0.01 -0.39 -0.56 1.00
Prelm 0.02 0.43 0.38 0.57 0.30 -0.23 1.00
GPA  0.00 0.42 0.37 0.52 0.28 -0.22 0.42 1.00
MA  0.00 0.32 0.29 0.45 0.26 -0.22 0.36 0.31 1.00

> lowerCor(z.data)

          ID   GREV  GREQ GREA  Ach Anx Prelm GPA  MA
ID  1.000
GREV -0.01 1.00
GREQ  0.00 0.73 1.00
GREA -0.01 0.64 0.60 1.00
Ach  0.00 0.01 0.01 0.45 1.00
Anx  0.00 0.01 0.01 -0.39 -0.56 1.00
Prelm 0.02 0.43 0.38 0.57 0.30 -0.23 1.00
GPA  0.00 0.42 0.37 0.52 0.28 -0.22 0.42 1.00
MA  0.00 0.32 0.29 0.45 0.26 -0.22 0.36 0.31 1.00
```
Show that the two matrices do not differ using the `lowerUpper` function

```r
r <- lowerCor(my.data)  # find the original correlations
z <- lowerCor(z.data)   # find the z transformed correlations
lu <- lowerUpper(r,z,diff=TRUE)  # combine into one matrix and take the difference
round(lu,2)
```

<table>
<thead>
<tr>
<th></th>
<th>ID</th>
<th>GREV</th>
<th>GREQ</th>
<th>GREA</th>
<th>Ach</th>
<th>Anx</th>
<th>Prelim</th>
<th>GPA</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ID</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>GREV</td>
<td>-0.01</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>GREQ</td>
<td>0.00</td>
<td>0.73</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>GREA</td>
<td>-0.01</td>
<td>0.64</td>
<td>0.60</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Ach</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.45</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Anx</td>
<td>-0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.39</td>
<td>-0.56</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Prelim</td>
<td>0.02</td>
<td>0.43</td>
<td>0.38</td>
<td>0.57</td>
<td>0.30</td>
<td>-0.23</td>
<td>NA</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>GPA</td>
<td>0.00</td>
<td>0.42</td>
<td>0.37</td>
<td>0.52</td>
<td>0.28</td>
<td>-0.22</td>
<td>0.42</td>
<td>NA</td>
<td>0.00</td>
</tr>
<tr>
<td>MA</td>
<td>-0.01</td>
<td>0.32</td>
<td>0.29</td>
<td>0.45</td>
<td>0.26</td>
<td>-0.22</td>
<td>0.36</td>
<td>0.31</td>
<td>NA</td>
</tr>
</tbody>
</table>
Scatter Plot Matrix showing correlation and LOESS regression

UgradGPA

GRE.Q

GRE.V

0.33
0.34
0.46

76 / 137
The effect of selection on the correlation

- Consider what happens if we select a subset
  - The “Oregon” model
  - \((\text{GPA} + (\text{V}+\text{Q})/200) > 11.6\)
- The range is truncated, but even more important, by using a compensatory selection model, we have changed the sign of the correlations.
Regression and restriction of range

Although the correlation is very sensitive, regression slopes are relatively insensitive to restriction of range.
**R code for regression figures**

```r
gradq <- subset(gradf, gradf[2] > 700) # choose the subset
with(gradq, lm(GRE.V ~ GRE.Q)) # do the regression

Call:
lm(formula = GRE.V ~ GRE.Q)

Coefficients:
(Intercept)      GRE.Q
     258.1549     0.4977

# show the graphic
op <- par(mfrow = c(1, 2)) # two panel graph
with(gradf, {
  plot(GRE.V ~ GRE.Q, xlim = c(200, 800), main = 'Original data', pch = 16)
  abline(lm(GRE.V ~ GRE.Q))
})
text(300, 500, 'r = .46  b = .56')
with(gradq, {
  plot(GRE.V ~ GRE.Q, xlim = c(200, 800), main = 'GRE Q > 700', pch = 16)
  abline(lm(GRE.V ~ GRE.Q))
})
text(300, 500, 'r = .18  b = .50')
op <- par(mfrow = c(1, 1)) # switch back to one panel
```

79 / 137
Show many correlations with a heat map using `cor.plot`.

Big 5 Inventory Items from SAPA
Alternative versions of the correlation coefficient

Table: A number of correlations are Pearson $r$ in different forms, or with particular assumptions. If $r = \frac{\sum x_i y_i}{\sqrt{\sum x_i^2 \sum y_i^2}}$, then depending upon the type of data being analyzed, a variety of correlations are found.

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>symbol</th>
<th>X</th>
<th>Y</th>
<th>Assumptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson</td>
<td>$r$</td>
<td>continuous</td>
<td>continuous</td>
<td></td>
</tr>
<tr>
<td>Spearman</td>
<td>rho ($\rho$)</td>
<td>ranks</td>
<td>ranks</td>
<td></td>
</tr>
<tr>
<td>Point bi-serial</td>
<td>$r_{pb}$</td>
<td>dichotomous</td>
<td>continuous</td>
<td></td>
</tr>
<tr>
<td>Phi</td>
<td>$\phi$</td>
<td>dichotomous</td>
<td>dichotomous</td>
<td></td>
</tr>
<tr>
<td>Bi-serial</td>
<td>$r_{bis}$</td>
<td>dichotomous</td>
<td>continuous</td>
<td>normality</td>
</tr>
<tr>
<td>Tetrachoric</td>
<td>$r_{tet}$</td>
<td>dichotomous</td>
<td>dichotomous</td>
<td>bivariate normality</td>
</tr>
<tr>
<td>Polychoric</td>
<td>$r_{pc}$</td>
<td>categorical</td>
<td>categorical</td>
<td>bivariate normality</td>
</tr>
</tbody>
</table>
The $\phi$ coefficient is just a Pearson r on dichotomous data

**Table:** The basic table for a phi, $\phi$ coefficient, expressed in raw frequencies in a four fold table is taken from Pearson & Heron (1913)

<table>
<thead>
<tr>
<th></th>
<th>Success</th>
<th>Failure</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accept</td>
<td>A</td>
<td>B</td>
<td>$R_1 = A + B$</td>
</tr>
<tr>
<td>Reject</td>
<td>C</td>
<td>D</td>
<td>$R_2 = C + D$</td>
</tr>
<tr>
<td>Total</td>
<td>$C_1 = A + C$</td>
<td>$C_2 = B + D$</td>
<td>$n = A + B + C + D$</td>
</tr>
</tbody>
</table>

In terms of the raw data coded 0 or 1, the *phi coefficient* can be derived directly by direct substitution, recognizing that the only non zero product is found in the A cell

\[
n \sum X_i Y_i - \sum X_i \sum Y_i = nA - R_1 C_1
\]

\[
\phi = \frac{AD - BC}{\sqrt{(A + B)(C + D)(A + C)(B + D)}}. \tag{4}
\]
Correlation size ≠ causal importance

**Table:** The relationship between sex and pregnancy (hypothetical data)

<table>
<thead>
<tr>
<th></th>
<th>Pregnant</th>
<th>Not Pregnant</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercourse</td>
<td>2</td>
<td>1,041</td>
<td>1,043</td>
</tr>
<tr>
<td>No intercourse</td>
<td>0</td>
<td>6,257</td>
<td>6,257</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>7,298</td>
<td>7,300</td>
</tr>
</tbody>
</table>

\[
\text{Phi} = .04
\]

> sex <- c(2, 1041, 0, 6257)
> phi(sex)

[1] 0.04
The biserial correlation estimates the latent correlation

\[ r = 0.9 \quad \text{rpb = 0.71} \quad \text{rbis = 0.89} \]

\[ r = 0.6 \quad \text{rpb = 0.48} \quad \text{rbis = 0.6} \]

\[ r = 0.3 \quad \text{rpb = 0.23} \quad \text{rbis = 0.28} \]

\[ r = 0 \quad \text{rpb = 0.02} \quad \text{rbis = 0.02} \]
The tetrachoric correlation estimates the latent correlation
Correlation size ≠ causal importance – tetrachoric correlation

Table: The relationship between sex and pregnancy (hypothetical data)

<table>
<thead>
<tr>
<th></th>
<th>Pregnant</th>
<th>Not Pregnant</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercourse</td>
<td>2</td>
<td>1,041</td>
<td>1,043</td>
</tr>
<tr>
<td>No intercourse</td>
<td>0</td>
<td>6,257</td>
<td>6,257</td>
</tr>
<tr>
<td>Total</td>
<td>2</td>
<td>7,298</td>
<td>7,300</td>
</tr>
<tr>
<td>Phi</td>
<td>.04</td>
<td></td>
<td>.95</td>
</tr>
</tbody>
</table>

> sex <- c(2, 1041, 0, 6257)
> phi(sex)
[1] 0.04

> tetrachoric(sex, correct=FALSE)
Call: tetrachoric(x = sex, correct = FALSE)
tetrachoric correlation
[1] 0.95

with tau of
[1] -3.5 -1.1
Pearson $r$ versus tetrachoric correlation on dichotomous ability data

```r
> tet <- tetrachoric(ability)
Loading required package: mvtnorm
Loading required package: parallel
> per <- lowerCor(ability)
> per.tet <- lowerUpper(tet$rho,per)
> per.tet.diff <- lowerUpper(tet$rho,per,diff=TRUE)
> round(per.tet[1:8,1:8],2)
```

<table>
<thead>
<tr>
<th></th>
<th>reason.4</th>
<th>reason.16</th>
<th>reason.17</th>
<th>reason.19</th>
<th>letter.7</th>
<th>letter.33</th>
<th>letter.34</th>
<th>letter.58</th>
</tr>
</thead>
<tbody>
<tr>
<td>reason.4</td>
<td>NA</td>
<td>0.28</td>
<td>0.40</td>
<td>0.30</td>
<td>0.28</td>
<td>0.23</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>reason.16</td>
<td>0.45</td>
<td>NA</td>
<td>0.32</td>
<td>0.25</td>
<td>0.27</td>
<td>0.20</td>
<td>0.26</td>
<td>0.21</td>
</tr>
<tr>
<td>reason.17</td>
<td>0.61</td>
<td>0.51</td>
<td>NA</td>
<td>0.34</td>
<td>0.29</td>
<td>0.26</td>
<td>0.29</td>
<td>0.29</td>
</tr>
<tr>
<td>reason.19</td>
<td>0.46</td>
<td>0.40</td>
<td>0.53</td>
<td>NA</td>
<td>0.25</td>
<td>0.25</td>
<td>0.27</td>
<td>0.25</td>
</tr>
<tr>
<td>letter.7</td>
<td>0.45</td>
<td>0.43</td>
<td>0.47</td>
<td>0.40</td>
<td>NA</td>
<td>0.34</td>
<td>0.40</td>
<td>0.33</td>
</tr>
<tr>
<td>letter.33</td>
<td>0.37</td>
<td>0.32</td>
<td>0.42</td>
<td>0.39</td>
<td>0.52</td>
<td>NA</td>
<td>0.37</td>
<td>0.28</td>
</tr>
<tr>
<td>letter.34</td>
<td>0.46</td>
<td>0.41</td>
<td>0.47</td>
<td>0.43</td>
<td>0.60</td>
<td>0.56</td>
<td>NA</td>
<td>0.32</td>
</tr>
<tr>
<td>letter.58</td>
<td>0.47</td>
<td>0.35</td>
<td>0.48</td>
<td>0.40</td>
<td>0.51</td>
<td>0.43</td>
<td>0.50</td>
<td>NA</td>
</tr>
</tbody>
</table>

```r
> round(per.tet.diff[1:8,1:8],2)
```

<table>
<thead>
<tr>
<th></th>
<th>reason.4</th>
<th>reason.16</th>
<th>reason.17</th>
<th>reason.19</th>
<th>letter.7</th>
<th>letter.33</th>
<th>letter.34</th>
<th>letter.58</th>
</tr>
</thead>
<tbody>
<tr>
<td>reason.4</td>
<td>NA</td>
<td>0.17</td>
<td>0.21</td>
<td>0.17</td>
<td>0.16</td>
<td>0.14</td>
<td>0.17</td>
<td>0.18</td>
</tr>
<tr>
<td>reason.16</td>
<td>0.45</td>
<td>NA</td>
<td>0.19</td>
<td>0.15</td>
<td>0.16</td>
<td>0.13</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td>reason.17</td>
<td>0.61</td>
<td>0.51</td>
<td>NA</td>
<td>0.19</td>
<td>0.18</td>
<td>0.16</td>
<td>0.18</td>
<td>0.19</td>
</tr>
<tr>
<td>reason.19</td>
<td>0.46</td>
<td>0.40</td>
<td>0.53</td>
<td>NA</td>
<td>0.14</td>
<td>0.14</td>
<td>0.15</td>
<td>0.15</td>
</tr>
<tr>
<td>letter.7</td>
<td>0.45</td>
<td>0.43</td>
<td>0.47</td>
<td>0.40</td>
<td>NA</td>
<td>0.18</td>
<td>0.20</td>
<td>0.18</td>
</tr>
<tr>
<td>letter.33</td>
<td>0.37</td>
<td>0.32</td>
<td>0.42</td>
<td>0.39</td>
<td>0.52</td>
<td>NA</td>
<td>0.19</td>
<td>0.15</td>
</tr>
<tr>
<td>letter.34</td>
<td>0.46</td>
<td>0.41</td>
<td>0.47</td>
<td>0.43</td>
<td>0.60</td>
<td>0.56</td>
<td>NA</td>
<td>0.18</td>
</tr>
<tr>
<td>letter.58</td>
<td>0.47</td>
<td>0.35</td>
<td>0.48</td>
<td>0.40</td>
<td>0.51</td>
<td>0.43</td>
<td>0.50</td>
<td>NA</td>
</tr>
</tbody>
</table>
Pearson $r$ versus polychoric correlation on 6 alternative BFI data

```r
> poly <- polychoric(bfi[1:10])
> pearson <- cor(bfi[1:10], use="pairwise")
> poly.pear <- lowerUpper(poly$rho, pearson)
> poly.pear.diff <- lowerUpper(poly$rho, pearson, diff=TRUE)
> poly.pear
> round(poly.pear, 2)
A1  A2  A3  A4  A5  C1  C2  C3  C4  C5
A1  NA -0.34 -0.27 -0.15 -0.18 0.03 0.02 -0.02 0.13 0.05
A2 -0.41  NA  0.49  0.34  0.39 0.09 0.14 0.19 0.15 -0.12
A3 -0.32  0.56  NA  0.36  0.50 0.10 0.14 0.13 -0.12 -0.16
A4 -0.18  0.39  0.41  NA  0.31 0.09 0.23 0.13 -0.15 -0.24
A5 -0.23  0.45  0.57  0.36  NA 0.12 0.11 0.13 -0.13 -0.17
C1  0.00  0.12  0.12  0.11  0.16  NA 0.43 0.31 -0.34 -0.25
C2  0.01  0.16  0.16  0.27  0.14 0.48  NA 0.36 -0.38 -0.30
C3 -0.02  0.23  0.16  0.17  0.15 0.34 0.40  NA -0.34 -0.34
C4  0.15 -0.19 -0.16 -0.20 -0.17 -0.40 -0.43 -0.38  NA  0.48
C5  0.06 -0.16 -0.19 -0.28 -0.20 -0.29 -0.33 -0.38  0.53  NA
> round(poly.pear.diff, 2)
A1  A2  A3  A4  A5  C1  C2  C3  C4  C5
A1  NA -0.07 -0.06 -0.03 -0.05 -0.02 -0.01  0.00  0.02  0.01
A2 -0.41  NA  0.07  0.05  0.06  0.02  0.02  0.03 -0.05 -0.03
A3 -0.32  0.56  NA  0.05  0.07  0.03  0.02  0.03 -0.04 -0.03
A4 -0.18  0.39  0.41  NA  0.05  0.02  0.04  0.04 -0.04 -0.04
A5 -0.23  0.45  0.57  0.36  NA  0.04  0.03  0.02 -0.04 -0.03
C1  0.00  0.12  0.12  0.11  0.16  NA  0.06  0.04 -0.06 -0.04
C2  0.01  0.16  0.16  0.27  0.14  0.48  NA  0.04 -0.05 -0.03
C3 -0.02  0.23  0.16  0.17  0.15  0.34  0.40  NA -0.04 -0.04
C4  0.15 -0.19 -0.16 -0.20 -0.17 -0.40 -0.43 -0.38  NA  0.05
C5  0.06 -0.16 -0.19 -0.28 -0.20 -0.29 -0.33 -0.38  0.53  NA
```
Spearman vs. Pearson on BFI data

```r
> spear <- cor(bfi[1:10], use="pairwise", method="spearman")
> spear.pear <- lowerUpper(spear, pearson, diff=TRUE)
> round(spear.pear, 2)
```

<table>
<thead>
<tr>
<th></th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>NA</td>
<td>-0.03</td>
<td>-0.03</td>
<td>-0.01</td>
<td>-0.04</td>
<td>-0.05</td>
<td>-0.03</td>
<td>-0.02</td>
<td>0.02</td>
<td>0.01</td>
</tr>
<tr>
<td>A2</td>
<td>-0.37</td>
<td>NA</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.03</td>
<td>-0.03</td>
</tr>
<tr>
<td>A3</td>
<td>-0.30</td>
<td>0.50</td>
<td>NA</td>
<td>0.00</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
<td>0.02</td>
<td>-0.03</td>
<td>-0.02</td>
</tr>
<tr>
<td>A4</td>
<td>-0.16</td>
<td>0.34</td>
<td>0.36</td>
<td>NA</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>-0.03</td>
<td>-0.01</td>
</tr>
<tr>
<td>A5</td>
<td>-0.22</td>
<td>0.40</td>
<td>0.53</td>
<td>0.31</td>
<td>NA</td>
<td>0.02</td>
<td>0.02</td>
<td>0.01</td>
<td>-0.03</td>
<td>-0.02</td>
</tr>
<tr>
<td>C1</td>
<td>-0.02</td>
<td>0.11</td>
<td>0.12</td>
<td>0.10</td>
<td>0.15</td>
<td>NA</td>
<td>0.02</td>
<td>0.01</td>
<td>-0.04</td>
<td>-0.01</td>
</tr>
<tr>
<td>C2</td>
<td>-0.01</td>
<td>0.14</td>
<td>0.15</td>
<td>0.25</td>
<td>0.13</td>
<td>0.45</td>
<td>NA</td>
<td>0.01</td>
<td>-0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>C3</td>
<td>-0.04</td>
<td>0.21</td>
<td>0.16</td>
<td>0.15</td>
<td>0.14</td>
<td>0.32</td>
<td>0.37</td>
<td>NA</td>
<td>-0.01</td>
<td>-0.01</td>
</tr>
<tr>
<td>C4</td>
<td>0.15</td>
<td>-0.18</td>
<td>-0.16</td>
<td>-0.16</td>
<td>-0.38</td>
<td>-0.40</td>
<td>-0.35</td>
<td>NA</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>0.06</td>
<td>-0.15</td>
<td>-0.18</td>
<td>-0.26</td>
<td>-0.19</td>
<td>-0.26</td>
<td>-0.30</td>
<td>-0.35</td>
<td>0.49</td>
<td>NA</td>
</tr>
</tbody>
</table>
Comments on these alternative correlations

1. The assumption is that there was an underlying bivariate, normal distribution that was somehow artificially dichotomized.

2. But some things are in fact dichotomous, not normally distributed
   - Alive/Dead
   - Vaccinated/Not vaccinated

3. Polychromic and tetrachoric correlations are found by iteratively fitting bivariate normal distributions with varying correlations until the best fit for a n x n table is found.

4. This is done using the tetrachoric or polychoric functions. They are not fast! (In comparison to Pearson r).
Cautions about correlations—The Anscombe data set

Consider the following 8 variables

<table>
<thead>
<tr>
<th>var</th>
<th>n</th>
<th>mean</th>
<th>sd</th>
<th>median</th>
<th>trimmed</th>
<th>mad</th>
<th>min</th>
<th>max</th>
<th>range</th>
<th>skew</th>
<th>kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>x1</td>
<td>1</td>
<td>9.0</td>
<td>3.32</td>
<td>9.00</td>
<td>9.00</td>
<td>4.45</td>
<td>4.00</td>
<td>14.00</td>
<td>10.00</td>
<td>0.00</td>
<td>-1.2</td>
</tr>
<tr>
<td>x2</td>
<td>2</td>
<td>9.0</td>
<td>3.32</td>
<td>9.00</td>
<td>9.00</td>
<td>4.45</td>
<td>4.00</td>
<td>14.00</td>
<td>10.00</td>
<td>0.00</td>
<td>-1.2</td>
</tr>
<tr>
<td>x3</td>
<td>3</td>
<td>9.0</td>
<td>3.32</td>
<td>9.00</td>
<td>9.00</td>
<td>4.45</td>
<td>4.00</td>
<td>14.00</td>
<td>10.00</td>
<td>0.00</td>
<td>-1.2</td>
</tr>
<tr>
<td>x4</td>
<td>4</td>
<td>9.0</td>
<td>3.32</td>
<td>8.00</td>
<td>8.00</td>
<td>0.00</td>
<td>8.00</td>
<td>19.00</td>
<td>11.00</td>
<td>2.47</td>
<td>11.0</td>
</tr>
<tr>
<td>y1</td>
<td>5</td>
<td>7.5</td>
<td>2.03</td>
<td>7.58</td>
<td>7.49</td>
<td>1.82</td>
<td>4.26</td>
<td>16.84</td>
<td>10.58</td>
<td>-0.05</td>
<td>-0.5</td>
</tr>
<tr>
<td>y2</td>
<td>6</td>
<td>7.5</td>
<td>2.03</td>
<td>8.14</td>
<td>7.79</td>
<td>1.47</td>
<td>3.10</td>
<td>9.26</td>
<td>6.16</td>
<td>-0.98</td>
<td>0.8</td>
</tr>
<tr>
<td>y3</td>
<td>7</td>
<td>7.5</td>
<td>2.03</td>
<td>7.11</td>
<td>7.15</td>
<td>1.53</td>
<td>6.39</td>
<td>12.74</td>
<td>7.35</td>
<td>1.38</td>
<td>4.3</td>
</tr>
<tr>
<td>y4</td>
<td>8</td>
<td>7.5</td>
<td>2.03</td>
<td>7.04</td>
<td>7.20</td>
<td>1.90</td>
<td>5.25</td>
<td>12.50</td>
<td>7.25</td>
<td>1.12</td>
<td>3.1</td>
</tr>
</tbody>
</table>
Cautions, Anscombe continued

With regressions of

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 3.0000909 | 1.1247468 | 2.667348 | 0.025734051 |
| x1        | 0.5000909 | 0.1179055 | 4.241455 | 0.002169629 |

[[2]]

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 3.000909  | 1.1253024 | 2.666758 | 0.025758941 |
| x2        | 0.500000  | 0.1179637 | 4.238590 | 0.002178816 |

[[3]]

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 3.0024545 | 1.1244812 | 2.670080 | 0.025619109 |
| x3        | 0.4997273 | 0.1178777 | 4.239372 | 0.002176305 |

[[4]]

| Estimate | Std. Error | t value | Pr(>|t|) |
|----------|------------|---------|----------|
| (Intercept) | 3.0017273 | 1.1239211 | 2.670763 | 0.025590425 |
| x4        | 0.4999091 | 0.1178189 | 4.243028 | 0.002164602 |
Cautions about correlations: Anscombe data set

Anscombe's 4 Regression data sets

---

1. $y_1$ vs $x_1$
2. $y_2$ vs $x_2$
3. $y_3$ vs $x_3$
4. $y_4$ vs $x_4$
Further cautions about correlations—the problem of levels

1. Correlations taken at one level of analysis can be unrelated to those at another level

2. \[ r_{xy} = \eta_{x_{wg}} \times \eta_{y_{wg}} \times r_{xy_{wg}} + \eta_{x_{bg}} \times \eta_{y_{bg}} \times r_{xy_{bg}} \]

3. Where \( \eta \) is the correlation of the data with the within group values, or the group means.

4. The within group and between group correlations can even be of different sign!

5. The withinBetween data set is an example of this problem.

6. The statsBy function will find the within and between group correlations for this kind of multi-level design.
Cautions about correlations: Within versus between groups
Bias, or just Simpson’s Paradox?

Table: Hypothetical Admissions data showing sex discrimination

<table>
<thead>
<tr>
<th></th>
<th>Admit</th>
<th>Reject</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>40</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

\[
\text{Phi} = \frac{(VP - \text{HR} \times \text{SR})}{\sqrt{\text{HR} \times (1-\text{HR}) \times \text{SR} \times (1-\text{SR})}} = 0.60
\]

polychoric rho = 0.81
Calculate the $\phi$ and tetrachoric correlations

> admit <- c(40,10,10,40)
> phi(admit)
[1] 0.6
> phi2poly(.6,.5,.5)
[1] 0.8090178
> tetrachoric(admit)
Call: tetrachoric(x = admit)
tetrachoric correlation
[1] 0.81

1. Input the four cell counts
2. Find the $\phi$ coefficient
3. Convert this to a tetrachoric correlation by specifying the marginals
4. Or, just call tetrachoric with these cell entries
Sex discrimination by department shows opposite effect

Table: Hypothetical Admissions data showing sex discrimination

<table>
<thead>
<tr>
<th></th>
<th>Admit</th>
<th>Reject</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>40</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>40</td>
<td>50</td>
</tr>
<tr>
<td>Total</td>
<td>50</td>
<td>50</td>
<td>100</td>
</tr>
</tbody>
</table>

Table: Males: unselective

<table>
<thead>
<tr>
<th></th>
<th>Admit</th>
<th>Reject</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>40</td>
<td>5</td>
<td>45</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>45</td>
<td>5</td>
<td>50</td>
</tr>
</tbody>
</table>

\[ \phi = -0.11 \quad \rho = -0.95 \]

Table: Females: selective

<table>
<thead>
<tr>
<th></th>
<th>Admit</th>
<th>Reject</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>0</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
<td>40</td>
<td>45</td>
</tr>
<tr>
<td>Total</td>
<td>5</td>
<td>45</td>
<td>50</td>
</tr>
</tbody>
</table>

\[ \phi = -0.11 \quad \rho = -0.95 \]
The ubiquitous correlation coefficient

Table: Alternative Estimates of effect size. Using the correlation as a scale free estimate of effect size allows for combining experimental and correlational data in a metric that is directly interpretable as the effect of a standardized unit change in x leads to r change in standardized y.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Estimate</th>
<th>r equivalent</th>
<th>as a function of r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson correlation</td>
<td>$r_{xy} = \frac{C_{xy}}{\sigma_x \sigma_y}$</td>
<td>$r_{xy}$</td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td>$b_{y,x} = \frac{C_{xy}}{\sigma_x^2}$</td>
<td>$r = \frac{b_{y,x} \sigma_y}{\sigma_x}$</td>
<td>$b_{y,x} = r \frac{\sigma_x}{\sigma_y}$</td>
</tr>
<tr>
<td>Cohen’s d</td>
<td>$d = \frac{X_1 - \bar{X}_2}{\sigma_x}$</td>
<td>$r = \frac{d}{\sqrt{d^2 + 4}}$</td>
<td>$d = \frac{2r}{\sqrt{1 - r^2}}$</td>
</tr>
<tr>
<td>Hedge’s g</td>
<td>$g = \frac{X_1 - \bar{X}_2}{s_x}$</td>
<td>$r = \frac{g}{\sqrt{g^2 + 4(df/N)}}$</td>
<td>$g = \frac{2r}{\sqrt{df/N}}$</td>
</tr>
<tr>
<td>t - test</td>
<td>$t = \frac{d \sqrt{df}}{2}$</td>
<td>$r = \sqrt{t^2/(t^2 + df)}$</td>
<td>$t = \sqrt{\frac{r^2 df}{1 - r^2}}$</td>
</tr>
<tr>
<td>F-test</td>
<td>$F = \frac{d^2 df}{4}$</td>
<td>$r = \sqrt{F/(F + df)}$</td>
<td>$F = \frac{r^2 df}{1 - r^2}$</td>
</tr>
<tr>
<td>Chi Square</td>
<td>$d = \frac{ln(OR)}{1.81}$</td>
<td>$r = \frac{ln(OR)}{1.81 \sqrt{(ln(OR)/1.81)^2 + 4}}$</td>
<td>$\chi^2 = r^2 n$</td>
</tr>
<tr>
<td>Odds ratio</td>
<td>$r_{equivalent} = r$ with probability p</td>
<td>$r = r_{equivalent}$</td>
<td>$ln(OR) = \frac{3.62r}{\sqrt{1 - r^2}}$</td>
</tr>
</tbody>
</table>
Correlation as the average of regressions

Galton’s insight was that if both $x$ and $y$ were on the same scale with equal variability, then the slope of the line was the same for both predictors and was measure of the strength of their relationship. Galton (1886) converted all deviations to the same metric by dividing through by half the interquartile range, and Pearson (1896) modified this by converting the numbers to standard scores (i.e., dividing the deviations by the standard deviation). Alternatively, the geometric mean of the two slopes ($b_{xy}$ and $b_{yx}$) leads to the same outcome:

$$r_{xy} = \sqrt{b_{xy}b_{yx}} = \sqrt{\frac{\text{Cov}_{xy}\text{Cov}_{yx}}{\sigma_x^2\sigma_y^2}} = \frac{\text{Cov}_{xy}}{\sqrt{\sigma_x^2\sigma_y^2}} = \frac{\text{Cov}_{xy}}{\sigma_x\sigma_y} \quad (5)$$

which is the same as the covariance of the standardized scores of $X$ and $Y$.

$$r_{xy} = \text{Cov}_{z_xz_y} = \text{Cov} \frac{x}{\sigma_x} \frac{y}{\sigma_y} = \frac{\text{Cov}_{xy}}{\sigma_x\sigma_y} \quad (6)$$
Error of correlation

The slope $b_{y,x}$ was found so that it minimizes the sum of the squared residual, but what is it? That is, how big is the variance of the residual?

$$V_r = \frac{\sum_{i=1}^{n} (y - \hat{y})^2}{n} = \frac{\sum_{i=1}^{n} (y - b_{y,x} x)^2}{n}$$

$$V_r = \frac{\sum_{i=1}^{n} (y^2 + b_{y,x}^2 x^2 - 2 b_{y,x} xy)}{n}$$

$$V_r = V_y + \frac{\text{Cov}_{xy}^2}{V_x} - 2 \frac{\text{Cov}_{xy}^2}{V_x} = V_y - \frac{\text{Cov}_{xy}^2}{V_x}$$

$$V_r = V_y - r_{xy}^2 V_y = V_y (1 - r_{xy}^2) \quad (7)$$

That is, the variance of the residual in $Y$ or the variance of the error of prediction of $Y$ is the product of the original variance of $Y$ and one minus the squared correlation between $X$ and $Y$. The squared correlation between $x$ and $y$ is thus an index of the amount of variance in $Y$ that is linearly predicted by $X$. This squared correlation is known as the index of determination.
Variance and correlations

The various relationships between correlations, predicted scores, the variance of the predicted scores, and the variances of the residuals may be seen in the following table ([19]).

**Table:** The basic relationships between Variance, Covariance, Correlation and Residuals

<table>
<thead>
<tr>
<th></th>
<th>Variance</th>
<th>Covariance with X</th>
<th>Covariance with Y</th>
<th>Correlation with X</th>
<th>Correlation with Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>$V_x$</td>
<td>$V_x$</td>
<td>$C_{xy}$</td>
<td>1</td>
<td>$r_{xy}$</td>
</tr>
<tr>
<td>Y</td>
<td>$V_y$</td>
<td>$C_{xy}$</td>
<td>$V_y$</td>
<td>$r_{xy}$</td>
<td>1</td>
</tr>
<tr>
<td>$\hat{Y}$</td>
<td>$r_{xy}^2 V_y$</td>
<td>$C_{xy} = r_{xy} \sigma_x \sigma_y$</td>
<td>$r_{xy} V_y$</td>
<td>1</td>
<td>$r_{xy}$</td>
</tr>
<tr>
<td>$Y_r = Y - \hat{Y}$</td>
<td>$(1 - r_{xy}^2)V_y$</td>
<td>0</td>
<td>$(1 - r_{xy}^2)V_y$</td>
<td>0</td>
<td>$\sqrt{1 - r^2}$</td>
</tr>
</tbody>
</table>
Set theoretic approach: Partitioning the variance in Y

\[ \beta_{y,x} \]

\[ X_1 \rightarrow Y \rightarrow \epsilon_1 \]

\[ \beta_{y,x} = \frac{\sigma_{xy}}{\sigma_x^2} \]

\[ \hat{y} = \beta_{y,x} x \]

\[ r_{xy} = \frac{\sigma_{xy}}{\sqrt{\sigma_x^2 \sigma_y^2}} \]

\[ V_r = V_y + \frac{\text{Cov}_{xy}^2}{V_x} - 2 \frac{\text{Cov}_{xy}^2}{V_x} \]

\[ V_r = V_y - \frac{\text{Cov}_{xy}^2}{V_x} \]

\[ V_r = V_y - r_{xy}^2 V_y \]

\[ V_r = V_y (1 - r_{xy}^2) \]

Variance in Y predicted by \( X = r_{xy}^2 \sigma_y^2 \)
Distance in the observational space

Because X and Y are vectors in the space defined by the observations, the covariance between them may be thought of in terms of the average squared distance between the two vectors in that same space. That is, following Pythagorus, the distance, \( d \), is simply the square root of the sum of the squared distances in each dimension (for each pair of observations), or, if we find the average distance, we can find the square root of the sum of the squared distances divided by \( n \):

\[
d^2_{xy} = \frac{1}{n} \sum_{i=1}^{n} (x_i - y_i)^2.
\]

which is the same as

\[
d^2_{xy} = V_x + V_y - 2C_{xy}
\]

\[
d_{xy} = \sqrt{2 \times (1 - r_{xy})}.
\]

(8)
Distance, correlations, and the law of cosines

Compare this to the trigonometric law of cosines,

\[ c^2 = a^2 + b^2 - 2ab \cdot \cos(ab), \]

and we see that the distance between two vectors is the sum of their variances minus twice the product of their standard deviations times the cosine of the angle between them. That is, the correlation is the cosine of the angle between the two vectors. The next figure shows these relationships for two Y vectors. The correlation, \( r_1 \), of \( X \) with \( Y_1 \) is the cosine of \( \theta_1 = \) the ratio of the projection of \( Y_1 \) onto \( X \). From the Pythagorean Theorem, the length of the residual \( Y \) with \( X \) removed (\( Y.x \)) is \( \sigma_y \sqrt{1 - r^2} \).
A geometric version of correlation

Correlations as cosines

\[ r_1, r_2, \theta_1, \theta_2, y_1, y_2 \]
The Ideal model of predicting $Y$ from $X_1$ and $X_2$

Variance in $Y$ predicted by $X_1$ and $X_2$ if $X_1$ and $X_2$ are independent. $\hat{V}_y = V_y r_{x_1y}^2 + V_y r_{x_2y}^2$
The usual case of predicting Y from $X_1$ and $X_2$

Variance in Y predicted by $X_1$ and $X_2$ if $X_1$ and $X_2$ - overlapping predictions

\[ \hat{V}_y = V_y r_{x_1y}^2 + V_y r_{x_2y}^2 - \text{overlap} \]

But what is the overlap?
Multiple correlations

\[ X \]

\[ r_{X_1Y} \]

\[ X_1 \]

\[ r_{X_1X_2} \]

\[ X_2 \]

\[ r_{X_2Y} \]

\[ Y \]
Multiple Regression

\[ X_1 \rightarrow Y \rightarrow \epsilon \]

\[ X_2 \rightarrow Y \rightarrow \epsilon \]

\[ r_{x_1 x_2} \]

\[ \beta_{y,x_1} \]

\[ \beta_{y,x_2} \]
Multiple Regression: decomposing correlations

\[ X \rightarrow r_{x_1y} \rightarrow Y \]

\[ X_1 \rightarrow \beta_{y,x_1} \rightarrow Y \]

\[ X_2 \rightarrow \beta_{y,x_2} \rightarrow Y \]

\[ r_{x_1y} \]

\[ r_{x_1x_2} \]

\[ r_{x_2y} \]
Multiple Regression: decomposing correlations

\[ r_{x_1y} = \beta_{y.x_1} + r_{x_1x_2}\beta_{y.x_2} \]

\[ r_{x_2y} = \beta_{y.x_2} + r_{x_1x_2}\beta_{y.x_1} \]
Multiple Regression: decomposing correlations

\[ r_{x_1y} = \beta_{y.x_1} + r_{x_1x_2}\beta_{y.x_2} \]

\[ r_{x_2y} = \beta_{y.x_2} + r_{x_1x_2}\beta_{y.x_1} \]

\[ \beta_{y.x_1} = \frac{r_{x_1y} - r_{x_1x_2}r_{x_2y}}{1 - r^2_{x_1x_2}} \]

\[ \beta_{y.x_2} = \frac{r_{x_2y} - r_{x_1x_2}r_{x_1y}}{1 - r^2_{x_1x_2}} \]
Multiple Regression: decomposing correlations

\[ r_{xy} = \beta_y.x_1 + r_{x_1x_2} \beta_y.x_2 \]

\[ r_{xy} = \beta_y.x_2 + r_{x_1x_2} \beta_y.x_1 \]

\[ \beta_{y.x_1} = \frac{r_{xy} - r_{x_1x_2} r_{x_2y}}{1 - r_{x_1x_2}^2} \]

\[ \beta_{y.x_2} = \frac{r_{x_2y} - r_{x_1x_2} r_{x_1y}}{1 - r_{x_1x_2}^2} \]

\[ R^2 = r_{xy} \beta_{y.x_1} + r_{xy} \beta_{y.x_2} \]
What happens with 3 predictors? The correlations

\[ X \]

\[ Y \]

\[ X_1 \]

\[ X_2 \]

\[ X_3 \]

\[ r_{x_1x_2} \]

\[ r_{x_1x_3} \]

\[ r_{x_2x_3} \]

\[ r_{x_1y} \]

\[ r_{x_2y} \]

\[ r_{x_3y} \]
What happens with 3 predictors? \( \beta \) weights

\[
\begin{align*}
X_1 & \quad r_{x_1x_2} \\
X_2 & \quad r_{x_2x_3} \\
X_3 & \\
\beta_{y.x_1} & \quad \beta_{y.x_2} \\
\beta_{y.x_3} & \\
Y & \\
\epsilon_3
\end{align*}
\]
What happens with 3 predictors?

\[ r_{x_1y} = \beta_{y.x_1} + r_{x_1x_2}\beta_{y.x_1} + r_{x_1x_3}\beta_{y.x_3} \]

direct

\[ r_{x_2y} = \ldots \]

indirect

\[ r_{x_3y} = \ldots \]

The math gets tedious
Multiple regression and linear algebra

- Multiple regression requires solving multiple, simultaneous equations to estimate the direct and indirect effects.
  - Each equation is expressed as a $r_{x_i y}$ in terms of direct and indirect effects.
  - Direct effect is $\beta_{y.x_i}$
  - Indirect effect is $\sum_{j \neq i} beta_{y.x_j} r_{x_j y}$
- How to solve these equations?
- Tediously, or just use linear algebra.
3 special cases of regression

Orthogonal predictors

$$X_1$$

$$_{r_{X1Y}}$$

$$Y$$

$$X_2$$

$$_{r_{X2Y}}$$

Correlated predictors

$$X_1$$

$$_{r_{X1Y}}$$

$$Y$$

$$X_2$$

$$_{r_{X2Y}}$$

Suppressive predictors

$$X_1$$

$$_{r_{X1Y}}$$

$$Y$$

$$X_2$$

$$_{r_{X1X2}}$$
Three basic cases

- **Independent**
  - $X_1$ and $X_2$ are independent of $Y$.
- **Correlated**
  - $X_1$ and $X_2$ are correlated with each other and $Y$.
- **Suppressor**
  - $X_2$ suppresses the effect of $X_1$ on $Y$.
3 special cases of regression

Orthogonal predictors

Correlated predictors

Suppressive predictors

Orthogonal predictors:

\[ \beta_{y,x_1} = r_{x_1y} \]

\[ \beta_{y,x_2} = r_{x_2y} \]

\[ R^2 = r_{x_1y} \beta_{y,x_1} + r_{x_2y} \beta_{y,x_2} \]

Correlated predictors:

\[ \beta_{y,x_1} = \frac{r_{x_1y} - r_{x_1x_2} r_{x_2y}}{1 - r_{x_1x_2}^2} \]

\[ \beta_{y,x_2} = \frac{r_{x_2y} - r_{x_1x_2} r_{x_1y}}{1 - r_{x_1x_2}^2} \]

Suppressive predictors:

\[ \beta_{y,x_1} = \frac{r_{x_1y} - r_{x_1x_2} r_{x_2y}}{1 - r_{x_1x_2}^2} \]

\[ \beta_{y,x_2} = \frac{r_{x_2y} - r_{x_1x_2} r_{x_1y}}{1 - r_{x_1x_2}^2} \]

\[ R^2 = r_{x_1y} \beta_{y,x_1} + r_{x_2y} \beta_{y,x_2} \]
Three basic cases: Theoretical examples

Independent
- PA
- NA
- Depression

Correlated
- Anxiety
- NA
- Depression
- Tension
- Anxiety

Suppressor
Find the regression of rated Prelim score on GREV

```r
> mod1 <- lm(GPA~GREV, data=mydata)
> summary(mod1)

Call:
  lm(formula = GPA ~ GREV, data = mydata)

Residuals:
     Min      1Q  Median      3Q     Max
-1.45807 -0.32322  0.00107  0.32811  1.44850

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.011729  0.069434   43.38  <2e-16 ***
GREV         0.001984  0.000136    14.60  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4558 on 998 degrees of freedom
Multiple R-squared: 0.176, Adjusted R-squared: 0.1751
F-statistic: 213.1 on 1 and 998 DF,  p-value: < 2.2e-16
```
Regression on z transformed data

```r
> mod2 <- lm(GPA~GREV, data=z.data)
> summary(mod2)

Call:
  lm(formula = GPA ~ GREV, data = z.data)

Residuals:
     Min       1Q   Median       3Q      Max
-2.90526 -0.64404  0.00213  0.65377  2.88619

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.888e-17  2.872e-02   0.00       1
GREV           4.195e-01  2.873e-02  14.60  <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9082 on 998 degrees of freedom
Multiple R-squared: 0.176,  Adjusted R-squared: 0.1751
F-statistic: 213.1 on 1 and 998 DF,  p-value: < 2.2e-16
```

Note that the slope is the same as the correlation.
> mod3 <- lm(GPA~GREV,data=cent)
> summary(mod3)

Call:
  lm(formula = GPA ~ GREV, data = cent)

Residuals:
        Min       1Q   Median       3Q      Max
-1.45807 -0.32322  0.00107  0.32811  1.44850

Coefficients:
                       Estimate  Std. Error   t value  Pr(>|t|)
(Intercept)          -3.332e-17  1.441e-02  0.00000      1
GREV                1.984e-03  1.359e-04  14.60000   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4558 on 998 degrees of freedom
Multiple R-squared: 0.176,    Adjusted R-squared: 0.1751
F-statistic: 213.1 on 1 and 998 DF,  p-value: < 2.2e-16

Note that the slope of the centered data is in the same units as the raw data, just the intercept has changed.
Multiple Regression: decomposing correlations

\[ r_{x_1y} = \beta_{y.x_1} + r_{x_1x_2} \beta_{y.x_2} \]

\[ r_{x_2y} = \beta_{y.x_2} + r_{x_1x_2} \beta_{y.x_1} \]

\[ \beta_{y.x_1} = \frac{r_{x_1y} - r_{x_1x_2} r_{x_2y}}{1 - r_{x_1x_2}^2} \]

\[ \beta_{y.x_2} = \frac{r_{x_2y} - r_{x_1x_2} r_{x_1y}}{1 - r_{x_1x_2}^2} \]

\[ R^2 = r_{x_1y} \beta_{y.x_1} + r_{x_2y} \beta_{y.x_2} \]
2 predictors

> summary(lm(GPA ~ GREV + GREQ, data= cent))

Call:
 lm(formula = GPA ~ GREV + GREQ, data = cent)

Residuals:
       Min        1Q      Median        3Q        Max
-1.424420 -0.332280  0.006160  0.324650  1.437650

Coefficients:    Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  -2.651e-17  1.435e-02  0.0000  1.00000
GREV          1.534e-03  1.976e-04  7.7600  2.10e-14 ***
GREQ          6.314e-04  2.019e-04  3.1270  0.00182 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 1

Residual standard error: 0.4538 on 997 degrees of freedom
Multiple R-squared: 0.184,
Adjusted R-squared: 0.1823
F-statistic: 112.4 on 2 and 997 DF,  p-value: < 2.2e-16
Multiple R with z transformed data

Do the same regression, but on the z transformed data. The units are now in correlation units.

```r
> z.data <- data.frame(scale(my.data))
> summary(lm(GPA ~ GREV + GREQ, data = z.data))
```

Call:
`lm(formula = GPA ~ GREV + GREQ, data = z.data)`

Residuals:
```
           Min        1Q  Median         3Q        Max
-2.838211 -0.662083  0.012281  0.646879  2.864572
```

Coefficients:
```
                      Estimate  Std. Error t value  Pr(>|t|)  
(Intercept)         3.205e-17  2.860e-02   0.000 1.0000000
GREV                 3.242e-01  4.179e-02  7.760   2.10e-14 ***
GREQ                 1.306e-01  4.179e-02  3.127   0.00182 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1  1
```

Residual standard error: 0.9043 on 997 degrees of freedom
Multiple R-squared: 0.184,   Adjusted R-squared: 0.1823
F-statistic: 112.4 on 2 and 997 DF,  p-value: < 2.2e-16
The 3 correlations produce the beta weights

```r
> R.small <- cor(my.data[,c(2,3,8)])
> round(R.small,2)
GREV  GREQ  GPA
GREV 1.00 0.73 0.42
GREQ 0.73 1.00 0.37
GPA  0.42 0.37 1.00
> solve(R.small[1:2,1:2])
GREV  GREQ
GREV 2.133188 -1.554768
GREQ -1.554768 2.133188
> beta <- solve(R.small[1:2,1:2],
               R.small[3,1:2])
> beta
GREV  GREQ
0.3242492 0.1306439
> beta.1 <- (.42 - .73*.37)/(1-.73^2)
> beta.1
[1] 0.3209163
> beta.2 <- (.37 - .73 *.42)/(1-.73^2)
> beta.2
[1] 0.1357311
```

- Find the correlation matrix
- Display it to two decimals
- Find the inverse of GREV and GREQ correlations
- Show them
- Find the beta weights by solving the matrix equation
- show them
- Find the beta weights by using the formula
- Show them
**3 predictors, no interactions**

Use three predictors, but print it with only 2 decimals

```r
> print(summary(lm(GPA ~ GREV + GREQ + GREA , data= cent)),digits=3)
```

Call:
`lm(formula = GPA ~ GREV + GREQ + GREA, data = cent)`

Residuals:

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>1Q</th>
<th>Median</th>
<th>3Q</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Residuals</td>
<td>-1.2668</td>
<td>-0.3038</td>
<td>0.0073</td>
<td>0.3051</td>
<td>1.3022</td>
</tr>
</tbody>
</table>

Coefficients:

|                     | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------------|----------|------------|---------|----------|
| (Intercept)         | -6.89e-17 | 1.35e-02  | 0.00    | 1.00000  |
| GREV                | 6.66e-04  | 2.00e-04  | 3.32    | 0.00092  *** |
| GREQ                | 7.75e-05  | 1.96e-04  | 0.40    | 0.69233  |
| GREA                | 2.08e-03  | 1.81e-04  | 11.52   | < 2e-16  *** |

Signif. codes:  0 Ô****Ô  0.001 Ô***Ô  0.01 Ô**Ô  0.05 Ô*Ô  0.1 Ô Ô  1

Residual standard error: 0.427 on 996 degrees of freedom
Multiple R-squared: 0.28, Adjusted R-squared: 0.278
F-statistic: 129 on 3 and 996 DF,  p-value: <2e-16
3 predictors, no interactions

Use three predictors, but just the middle 200 subjects

```r
> mod4 <- lm(GPA ~ GREV + GREQ + GREA, data = cent[400:600,])
> summary(mod4)

Call:
  lm(formula = GPA ~ GREV + GREQ + GREA, data = cent[400:600, ])

Residuals:
   Min     1Q  Median     3Q    Max
-1.03553 -0.30799 -0.00889  0.29320  1.20228

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)           0.0397399  0.0310412  1.280   0.202
GREV                  0.0004706  0.0004530  1.039   0.300
GREQ                  0.0005236  0.0004515  1.160   0.248
GREA                  0.0017904  0.0004360  4.107 5.88e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4394 on 197 degrees of freedom
Multiple R-squared: 0.2259, Adjusted R-squared: 0.2141
F-statistic: 19.16 on 3 and 197 DF,  p-value: 6.051e-11
```
Interaction terms are just products in regression

• To interpret all effects, the data need to be 0 centered.
  • This makes the main effects orthogonal to the interaction term.
  • Otherwise, need to compare model with and without interactions

• Graph the results in non-standardized form

• Consider a real data set of SAT V, SAT Q and Gender

```r
> data(sat.act)
> colors=c("black","red")  #choose some nice colors
> symb=c(19,25)
> colors=c("black","red")  #choose some nice colors
> with(sat.act,plot(SATQ~SATV,pch=symb[gender], col=colors[gender], bg=colors[gender],cex=.6,main="SATQ varies by SATV and gender"))
> by(sat.act,sat.act$gender,function(x)
  abline(lm(SATQ~SATV,data=x))
```
An example of an interaction plot

SATQ varies by SATV and gender

```
data(sat.act)
c.sat <- data.frame(scale(sat.act,scale=FALSE))
> summary(lm(SATQ~SATV * gender,data=c.sat))
```

Call:  
```
lm(formula = SATQ ~ SATV * gender, data = c.sat)
```

Residuals:
```
  Min       1Q   Median       3Q      Max
-294.423  -49.876    5.577   53.210   291.100
```

Coefficients:
```
               Estimate Std. Error t value Pr(>|t|)  
(Intercept)  -0.26696   3.31211  -0.081   0.936  
SATV          0.65398   0.02926  22.350  < 2e-16 ***  
gender       -36.71820   6.91495  -5.310 1.48e-07 ***  
SATV:gender  -0.05835   0.06086  -0.959   0.338  
```

Signif. codes:  
```
  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 86.79 on 683 degrees of freedom  
(13 observations deleted due to missingness)

Multiple R-squared: 0.4391, Adjusted R-squared: 0.4367
F-statistic: 178.3 on 3 and 683 DF, p-value: < 2.2e-16
Interaction of Anxiety with Verbal

> mod5 <- lm(GPA ~ GREV * Anx, data=cent)
> summary(mod5)

Call:
`lm(formula = GPA ~ GREV * Anx, data = cent)`

Residuals:
```
    Min     1Q    Median     3Q    Max
-1.49677 -0.31527  -0.00054  0.31223  1.32156
```

Coefficients:
```
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)             -2.375e-04  1.395e-02  -0.017   0.986
GREV                     1.996e-03  1.316e-04  15.167  < 2e-16 ***
Anx                      -1.131e-02  1.414e-03  -7.997 3.51e-15 ***
GREV:Anx                 2.219e-05  1.377e-05   1.612   0.107
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 0.4412 on 996 degrees of freedom
Multiple R-squared: 0.2294, Adjusted R-squared: 0.227
F-statistic: 98.81 on 3 and 996 DF, p-value: < 2.2e-16
Testing for the significance of correlations

```r
> corr.test(sat.act)
Call: corr.test(x = sat.act)
Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>gender</th>
<th>education</th>
<th>age</th>
<th>ACT</th>
<th>SATV</th>
<th>SATQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender</td>
<td>1.00</td>
<td>0.09</td>
<td>-0.02</td>
<td>-0.04</td>
<td>-0.02</td>
<td>-0.17</td>
</tr>
<tr>
<td>education</td>
<td>0.09</td>
<td>1.00</td>
<td>0.55</td>
<td>0.15</td>
<td>0.05</td>
<td>0.03</td>
</tr>
<tr>
<td>age</td>
<td>-0.02</td>
<td>0.55</td>
<td>1.00</td>
<td>0.11</td>
<td>-0.04</td>
<td>-0.03</td>
</tr>
<tr>
<td>ACT</td>
<td>-0.04</td>
<td>0.15</td>
<td>0.11</td>
<td>1.00</td>
<td>0.56</td>
<td>0.59</td>
</tr>
<tr>
<td>SATV</td>
<td>-0.02</td>
<td>0.05</td>
<td>-0.04</td>
<td>0.56</td>
<td>1.00</td>
<td>0.64</td>
</tr>
<tr>
<td>SATQ</td>
<td>-0.17</td>
<td>0.03</td>
<td>-0.03</td>
<td>0.59</td>
<td>0.64</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Sample Size

<table>
<thead>
<tr>
<th></th>
<th>gender</th>
<th>education</th>
<th>age</th>
<th>ACT</th>
<th>SATV</th>
<th>SATQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>687</td>
</tr>
<tr>
<td>education</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>687</td>
</tr>
<tr>
<td>age</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>687</td>
</tr>
<tr>
<td>ACT</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>687</td>
</tr>
<tr>
<td>SATV</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>700</td>
<td>687</td>
</tr>
<tr>
<td>SATQ</td>
<td>687</td>
<td>687</td>
<td>687</td>
<td>687</td>
<td>687</td>
<td>687</td>
</tr>
</tbody>
</table>

Probability values (Entries above the diagonal are adjusted for multiple tests.)

<table>
<thead>
<tr>
<th></th>
<th>gender</th>
<th>education</th>
<th>age</th>
<th>ACT</th>
<th>SATV</th>
<th>SATQ</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender</td>
<td>0.00</td>
<td>0.17</td>
<td>1.00</td>
<td>1.00</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>education</td>
<td>0.02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>age</td>
<td>0.58</td>
<td>0.00</td>
<td>0.00</td>
<td>0.03</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

135 / 137
Various tests of significance

1. Is the correlation different from 0? cor.test, corr.test (for more than two variables)
2. Does a correlation differ from another correlation, r.test with or without a third variable.
3. Does a correlation matrix differ from an Identity matrix? cortest
4. Bootstrapping confidence intervals for correlations cor.ci
**The correlation coefficient**

1. Perhaps the most powerful and useful statistic ever developed
2. Special cases of the correlation are used throughout statistics.
3. The basic concepts of correlation are very straightforward
4. Many ways to be misled with correlations.


Galton, F. (1886). Regression towards mediocrity in hereditary


