**INTERPRETATION OF $r$**

- **if we calculate a value of $r$**
- **How do we know what it means?**
- **How do we compare $r$ values for different data sets?**

| $|r|$  | Interpretation       |
|------|----------------------|
| 0.9  | Very high correlation |
| 0.7  | High correlation      |
| 0.5  | Moderate correlation  |
| 0.3  | Low positive correlation |
| 0.0  | Little if any correlation |

**SCALE OF $r$**

- Values of $r$ are **ordinal** measures of correlation.
- Higher $r$ values indicate larger correlation.
- Equal spacings of $r$ values may not indicate equal spacings of correlation.

Thus, $r = 0.90$ is **not** twice as correlated as $r = 0.45$.

The difference in correlation between $r = 0.90$ and $r = 0.75$ is **not** the same as the difference in correlation between $r = 0.60$ and $r = 0.45$.

**VARIANCE**

- We can interpret $r$ in terms of variance.
- Correlation coefficient indicates relationships between variables.
- Also indicates proportion of **individual differences** that can be associated with individual differences of another variable.

**VARIANCE**

The idea is embedded in mathematical models.

Assume you want to **predict** the final exam score when you know the SAT score.

Line predicts score (could go in reverse too).

**VARIATION**

Deviation of a final exam score from the mean value can be due to deviation accounted for by SAT scores, or due to something else.
VARIATION

it turns out that
\[ r^2 = \frac{s_y^2}{s_y^2 + s_a^2} \]

- \( s_y^2 \) is the total variance in \( Y \)
- \( s_a^2 \) is the variance in \( Y \) associated with variance in \( X \)

thus, \( r^2 \) is the proportion of variance in \( Y \) accounted for with variance in \( X \)

we are skipping the mathematical details (thank you!)

called the coefficient of determination

CORRELATION

we have studied the correlational coefficient called the Pearson \( r \)

one limitation is that it only works for quantitative data (interval or ratio)

sometimes we want to calculate correlations of ordinal (ranked) data

e.g.,
does ranking near the top of SAT scores correlate with ranking near the top of Final Exam scores?
we might not know the actual scores

SPEARMAN \( \rho \)

special case of the Pearson \( r \)

formulas looks much different

\[ \rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)} \]

where

- \( n \) = number of paired ranks
- \( d \) = difference between paired ranks

CALCULATION

Ties take average rank

<table>
<thead>
<tr>
<th>( X )</th>
<th>( Y )</th>
<th>X rank</th>
<th>Y rank</th>
<th>( d )</th>
</tr>
</thead>
<tbody>
<tr>
<td>550</td>
<td>56</td>
<td>4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>320</td>
<td>55</td>
<td>8</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>735</td>
<td>65</td>
<td>1</td>
<td>2</td>
<td>-1</td>
</tr>
<tr>
<td>405</td>
<td>42</td>
<td>14</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>694</td>
<td>64</td>
<td>2</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>429</td>
<td>45</td>
<td>11</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>655</td>
<td>56</td>
<td>6.5</td>
<td>5.5</td>
<td>1</td>
</tr>
<tr>
<td>598</td>
<td>59</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>615</td>
<td>56</td>
<td>3</td>
<td>5.5</td>
<td>-2.5</td>
</tr>
<tr>
<td>473</td>
<td>42</td>
<td>13</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>289</td>
<td>38</td>
<td>18</td>
<td>14</td>
<td>-4</td>
</tr>
<tr>
<td>515</td>
<td>59</td>
<td>9</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>580</td>
<td>37</td>
<td>15</td>
<td>15</td>
<td>0</td>
</tr>
<tr>
<td>539</td>
<td>42</td>
<td>10</td>
<td>12</td>
<td>-2</td>
</tr>
<tr>
<td>665</td>
<td>53</td>
<td>6.5</td>
<td>8</td>
<td>-1.5</td>
</tr>
</tbody>
</table>

if we calculated \( r \) from the ranked data we would get \( r = 0.93 \)

when we calculated \( r \) from the raw scores we got \( r = 0.90 \)
(text is misleading)

graph of ranked data

\[ \rho = 0.93 \]
CAUSALITY

in behavioral sciences we often look for causalties to try to determine how to make decisions
SAT scores and Final Exam scores in a statistics class may be highly correlated but no one would claim that doing well on the SAT causes someone to get a good grade in statistics
getting a good grade and statistics may be caused by being smart
getting a high SAT score may be caused by being smart
being smart causes both variables to be highly correlated

CAUSATION

e.g.
in some places (war) there is a high correlation between bullets in the brain and being dead
that does not mean that being dead causes bullets to form in the brain
(in fact it is probably the other way around)
causation cannot be established through quantitative methods
establishment of causation requires understanding of the variables and their roles

TV, Teens, Drinking

“High school students who watch lots of television and music videos are more likely to start drinking alcohol than other youngsters while those who rent movies are at less risk.” Associated Press 3 November 1998.
found correlation between TV and music video watching and drinking among 9th graders
implied that glamorization of drinking on TV led to increased drinking in viewers

TV, Teens, Drinking

may be true, but parents and peers are known to have a very big influence, and this study did not control for those influences
conclusion does not necessarily imply causation
USA Today got it right. they quoted ABC’s Julie Hoover, who likened the conclusions to saying “gynecologists make women pregnant because... there are so many pregnant women in their offices”

CORRELATION

Other coefficients of correlation
- Nominal data ($\phi, C$ coefficient, Cramer’s $V, \lambda$
- Ordinal data (Rank-biserial, tetrachoric)
- Interval/ratio and nominal (point-biserial)
- Nonlinear ($\eta$)

other uses of correlation
- Inferential statistics (sampling theory)
- Predict scores (linear regression)

CONCLUSIONS

Pearson $r$
size
interpretation
coefficient of determination
NEXT TIME

probability
rules
significance

Why casinos make money.