CORRELATION

**Positive Relationship**

![Graph showing positive correlation between test score and ability.]

**Negative Relationship**

![Graph showing negative correlation between test score and ability.]

**No Relationship**

![Graph showing no relationship between test score and hair length.]

\( r = +1 \)

\( r = -1 \)

**Stronger Positive**

![Graph showing stronger positive correlation between test score and ability.]

**Weaker**

![Graph showing weaker positive correlation between test score and ability.]

**Stronger Negative**

![Graph showing stronger negative correlation between test score and ability.]

**Weaker**

![Graph showing weaker negative correlation between test score and ability.]
<table>
<thead>
<tr>
<th>$r$</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>STRONG</td>
<td>1.00</td>
</tr>
<tr>
<td>MODERATE</td>
<td>.64</td>
</tr>
<tr>
<td>WEAK</td>
<td>.09</td>
</tr>
</tbody>
</table>

$\text{NO} \rightarrow \text{YES}$

$r$ underestimates a curved relationship.

**Curvilinearity**

$r$ underestimates the relationship at one point and overestimates at another point.

**Heteroscedasticity**

$r$ overestimates $r$ underestimates

**Outliers**

$r$ underestimates

**Restriction of Range**

*Weight*

*Height*

$\text{Reading Ability}$

$\text{Months since beginning first grade}$
FORMULAS FOR CORRELATION

\[
r = \frac{1}{n} \sqrt{\frac{(\sum XY - (\sum X)(\sum Y))^2}{(\sum X^2 - (\sum X)^2/n)(\sum Y^2 - (\sum Y)^2/n)}}
\]

\[
r = \frac{(\sum SP_{XY})^2}{(SS_x)(SS_y)}
\]

\[
r = \frac{1}{\sqrt{(\sum (x-\bar{x})(y-\bar{y}))^2}}
\]

\[
r = \frac{\sum z_x z_y}{n}
\]
Review of Pearson $r$ as a measure of correlation

1. $r$ is an index of the **linear** relationship between two variables.

2. The sign of $r$ indicates the direction of the relationship, the magnitude of $r$ indicates its strength.

3. Scattergrams (also called scatter diagrams) are graphic portrayals of $r$. If the trend of the points is /, then $r$ is positive. If the trend is \, then $r$ is negative. The spread of the points indicates the strength of $r$. If the points are close to a straight line, then $r$ is stronger; if they are more spread out, then $r$ is weaker.

4. $r$ can be calculated using the $z$ score formula, the raw score (computational) formula, or the deviation (definitional) formula.

5. $r$ does not imply causation. $r$ is not equal to the percent of relationship between the variables.

6. Regardless of how much of a relationship actually does or does not exist, $r$ can be misleading if there is curvilinearity, heteroscedasticity, outliers, restriction of range, or unreliable measurement.