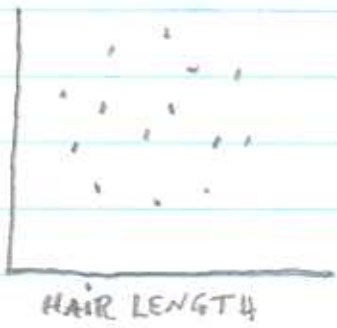
CORRELATION

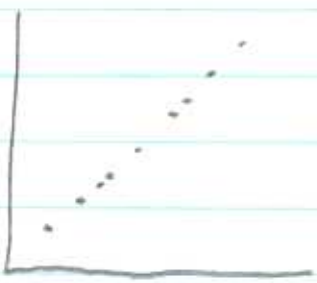
Positive Relationship

Negative Relationship

No Relationship



$r = +1$



$r = -1$

Positive

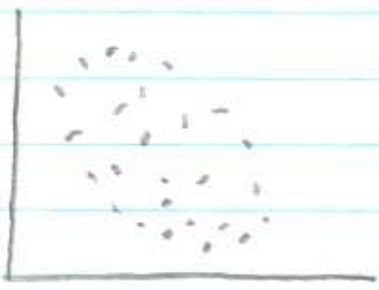
Stronger

Weaker

Negative

Stronger

Weaker



r		r^2
1.00	STRONG	1.00
.90		.81
.80	MODERATE	.64
.70		.49
.60		.36
.50	WEAK	.25
.40		.16
.30		.09
.20		.04
.10	NEAR ZERO	.01
0		0

r	
1.00	
.90	STRONG?
.80	
.70	
.60	MODERATE?
.50	
.40	
.30	
.20	WEAK?
.10	
0	

← NO YES →

r underestimates a curved relationship

Curvilinearity

Heteroscedasticity (different spread)

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

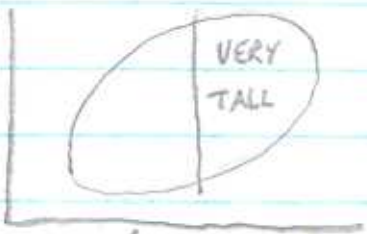
r overestimates

r underestimates

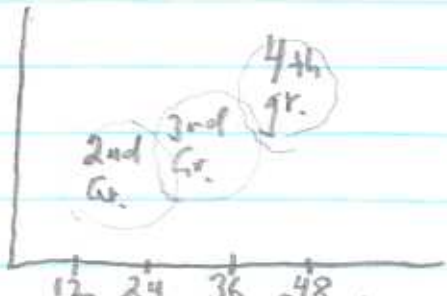
Outliers

r underestimates

Restriction of Range



Reading Ability



FORMULAS FOR CORRELATION

$$r = \sqrt{\frac{\left(\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}\right)^2}{\left(\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right)\left(\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}\right)}}$$

Raw score or computational formula

$$r = \sqrt{\frac{(SP_{XY})^2}{(SS_X)(SS_Y)}}$$

$$r = \sqrt{\frac{\left(\Sigma (X-\bar{X})(Y-\bar{Y})\right)^2}{\left(\Sigma (X-\bar{X})(X-\bar{X})\right)\left(\Sigma (Y-\bar{Y})(Y-\bar{Y})\right)}}$$

Deviation or definitional formula

$$r = \frac{\Sigma Z_X Z_Y}{n}$$

z score formula

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 alone imply causation necessarily. *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.