

Correlation

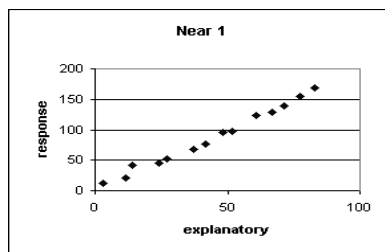
"A statistician is someone who loves to work with numbers but doesn't have the personality to be an accountant."

50% of the citizens of this country have a below average understanding of statistics.

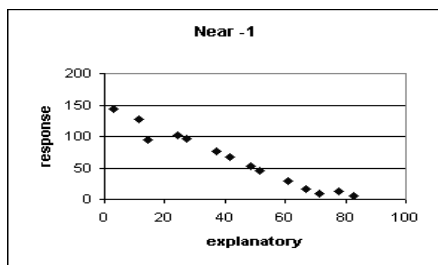
Correlation Coefficients

- Pearson product-moment correlation (a.k.a. Pearson's r)
 - The relationship between two variables of degree.
 - Positive: As one variable increases (or decreases) so does the other.
 - Negative: As one variable increases the other decreases.
 - Magnitude or strength of relationship
 - -1.00 to +1.00
 - Correlation does not equate to causation

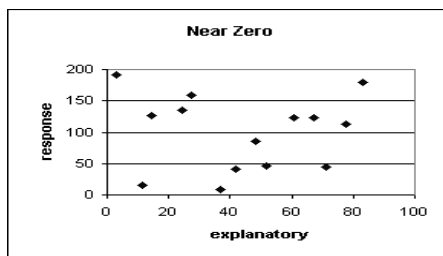
Positive Correlation



Negative Correlation



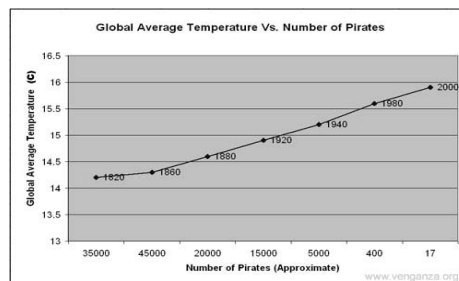
No Correlation



Correlations

- Thickness of scatter plot determines strength of correlation, not slope of line.
 - For example see:
 - <http://noppa5.pc.helsinki.fi/koe/corr/cor7.html>
- Remember correlation does not equate to causation.

Negative Correlation



Size of the correlation (Cohen, 1988)

	Negative	Positive
Small	-.10 to -.29	.10 to .29
Medium	-.30 to -.49	.30 to .49
Large	-.50 to -1.00	.50 to 1.00

An example from the literature (Cooper, et al., 1998)

- Correlations between homework behavior and achievement (Grades 6 through 10).
- Homework Behaviors
 - Amount teacher assigned
 - Portion completed
 - Time spent
- Achievement
 - Tennessee Comprehensive Assessment Program (TCAP)
- Independent variable(s)?
- Dependent variable(s)?

What do you expect?

- Predict the relationships between the amount assigned, portion completed, time spent with achievement.
 - Positive relationships?
 - Negative relationships?
 - No relationships?
- Based on Table 3
 - Explain the reported correlations

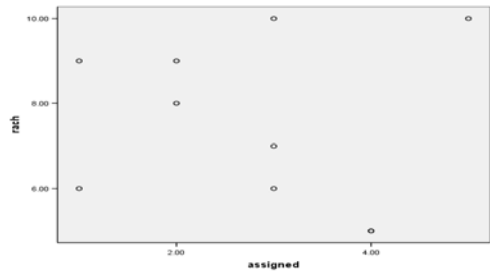
Calculate the correlation of teacher assigned and student achievement

- IV - "How much homework does your teacher assign each night?"
 - None (scored 1), 0 -14 minutes (scored 2), 15-29 minutes (scored 3), 30 -60 minutes (scored 4), and more than hour (scored 5).
- DV - TCAP raw scores on math portion of the test.

The Results

Assigned	Achievement
1.00	6.00
1.00	9.00
2.00	8.00
2.00	9.00
3.00	10.00
3.00	7.00
3.00	6.00
4.00	5.00
4.00	5.00
5.00	10.00

Draw a Scatterplot



Calculate Pearson's r using z scores.

- What is this formula telling us to do?
- The text uses N in the denominator.
 - This is related to using n-1 when calculating variance (population vs. sample).
 - If you want to get the same result as SPSS use n-1.

$$r = \frac{\sum Z_X Z_Y}{n - 1}$$

Steps

- First calculate descriptive statistics
 - Assigned
 - xbar = 2.8 and s = 1.36
 - Achievement
 - ybar = 7.5 and s = 1.95
- Now what do I need to do?

$$r = \frac{\sum Z_X Z_Y}{n - 1}$$

Calculate z scores

- Remember

$$z = \frac{x_i - \bar{x}}{s_x}$$

Z scores for X and Y

X (assigned)	Y (Achievement)
-1.36720	-.76613
-1.36720	.76613
-.60764	.25538
-.60764	.76613
.15191	1.27688
.15191	-.25538
.15191	-.76613
.91147	-1.27688
.91147	-1.27688
1.67102	1.27688

Now calculate the cross products of the z scores

X (assigned)	Y (Achievement)	ZxZy
-1.36720	-.76613	1.05
-1.36720	.76613	-1.05
-.60764	.25538	-.16
-.60764	.76613	-.47
.15191	1.27688	.19
.15191	-.25538	-.04
.15191	-.76613	-.12
.91147	-1.27688	-1.16
.91147	-1.27688	-1.16
1.67102	1.27688	2.13

Now what?

ZxZy
1.05
-1.05
-.16
-.47
.19
-.04
-.12
-1.16
-1.16
2.13

$$r = \frac{\sum Z_X Z_Y}{n - 1}$$

SPSS Output

Correlations

		assigned	rach
assigned	Pearson Correlation	1	-.086
	Sig. (2-tailed)		.813
	N	10	10
rach	Pearson Correlation	-.086	1
	Sig. (2-tailed)	.813	
	N	10	10

Definitional formula for Pearson's r

$$r_{XY} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$

Computational formula for Pearson's r.

$$r_{XY} = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{N}}{\sqrt{[\sum X^2 - \frac{(\sum X)^2}{N}][\sum Y^2 - \frac{(\sum Y)^2}{N}]}}$$

Covariance

- An index to the degree that two variables share variance (i.e., vary together).
 - By itself has no meaning.
 - Much like variance.
 - Needs to be standardized.
 - Text shows the total of cross products (the numerator)
- Definitional formula below

$$S_{xy} = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{n - 1}$$

Computational Formula for Covariance

- Which formula does this formula look similar to?

$$S_{xy}^2 = \frac{\sum XY - \frac{\sum X \sum Y}{n}}{n-1}$$

Let's calculate the covariance of assign and achieve

Descriptive Statistics

	N	Sum
assigned	10	28.00
rach	10	75.00
assignsquare	10	94.00
achsquare	10	597.00
assach	10	208.00
Valid N (listwise)	10	

Calculating Correlation from Covariance

$$r_{XY} = \frac{S_{XY}^2}{S_X S_Y}$$

Back to magnitude of effect

- Coefficient of determination
 - Also known as
 - Shared variance
 - The proportion of variance accounted for
 - Systematic variance
 - Percentage of variance accounted for
- Coefficient of nondetermination
 - Proportion of variance not accounted for

$$r^2$$

$$1-r^2$$

Problems associated with Pearson's r

- Lack of linear relationship
 - (e.g., anxiety and test performance)
- Restricted (truncated) range
 - Can reduce the magnitude of the correlation
- Sample size
- Outliers
- Two populations
 - It appears there is not a correlation (or the correlations is low), but when you stratify there is a correlation.
- Extreme scores
 - Selection bias
- Causal arguments
 - Correlation does not equate causation.