Notes 6: Correlation

1. Correlation

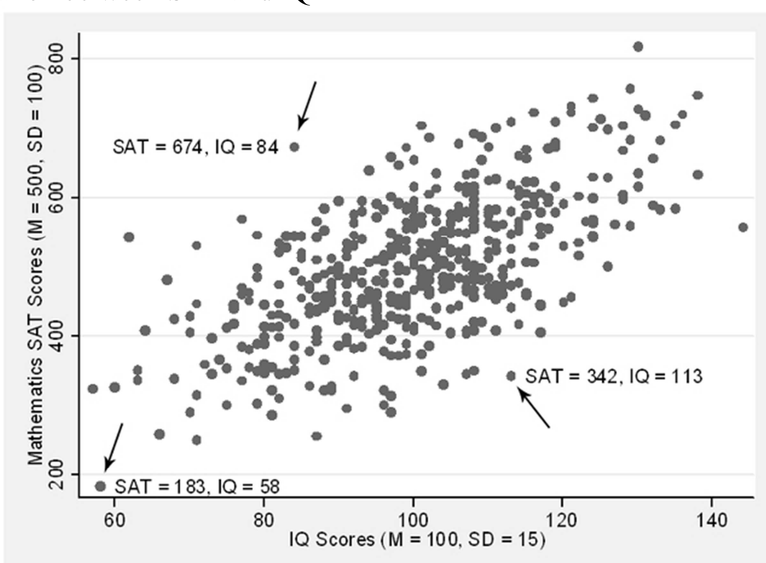
- correlation: this term usually refers to the degree of relationship or association between two quantitative variables, such as IQ and GPA, or GPA and SAT, or HEIGHT and WEIGHT, etc.
- positive relationship (↑↑): two variables vary in the same direction, i.e., they covary together; as one variable increases, the other variable also increases; e.g., higher GPAs correspond to higher SATs, and lower GPAs correspond to lower SATs;
- negative (inverse) relationship (↑↓): as one variables increases, the other decreases; e.g., higher GPAs correspond with lower SATs

(What type of relationship is it if both variables covary like ↓↓?)

- scatter plots, scatter grams: graphs that illustrate the relationship between two variables; each point of the scatter represents scores on two variables for one case or individual

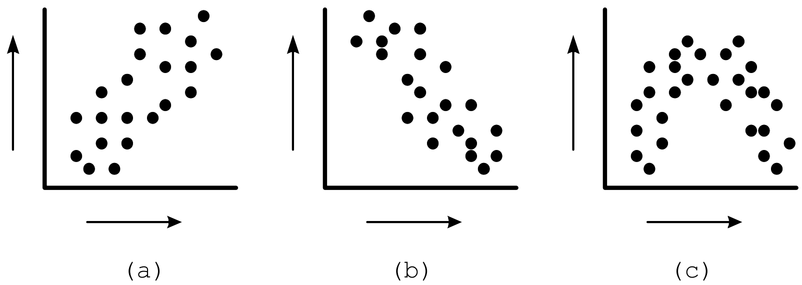
The figure below shows a positive and relatively strong correlation between SAT and IQ. Three points are identified in Figure 1 with arrows. One individual scored 674 on SAT and 84 on IQ, one scored 183 on SAT and 58 on IQ, and another scored 342 on SAT and 113 on IQ. As these three points illustrate, each dot represents the combination of two variables for one individual or case.

Figure 1: Correlation between SAT and IQ




```
***Begin Stata commands, Ignore these marks***
.corr2data SAT IQ, n(500) means(500 100) corr(1.00 .6321 1.00) sds(100 15) cstorage(lower)
.replace SAT = round(SAT)
.replace IQ = round(IQ)
.twoway (scatter SAT IQ, msymbol(circle)), ytitle(Mathematics SAT Scores (M = 500, SD = 100)) xtitle(IQ
Scores (M = 100, SD = 15)) legend(symplacement(north)) scheme(s2mono) text(342 113 "  SAT = 342, IQ =
113", placement(e))  text(674 84 "SAT = 674, IQ = 84  ", placement(w)) text(183 58 "  SAT = 183, IQ = 58",
placement(e))
***End Stata commands, Ignore these marks***
```

Figure 2: Three Scatterplots

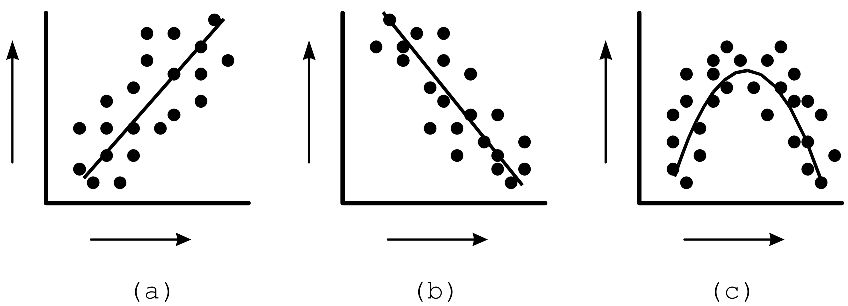


In the figure above are three scatterplots: (a) shows a positive relationship, figure (b) shows a negative relationship, and figure (c) depicts a curvilinear or non-linear relationship.

- linear representation: can a single, straight line be drawn for figure (a) that best represents the relationship between the two variables? what about (b), (c)? (draw the lines on the scatterplots)

In many cases the nature of relationship between two variables can be represented by a line that fits among the scatter. Examples are presented in the figure below.

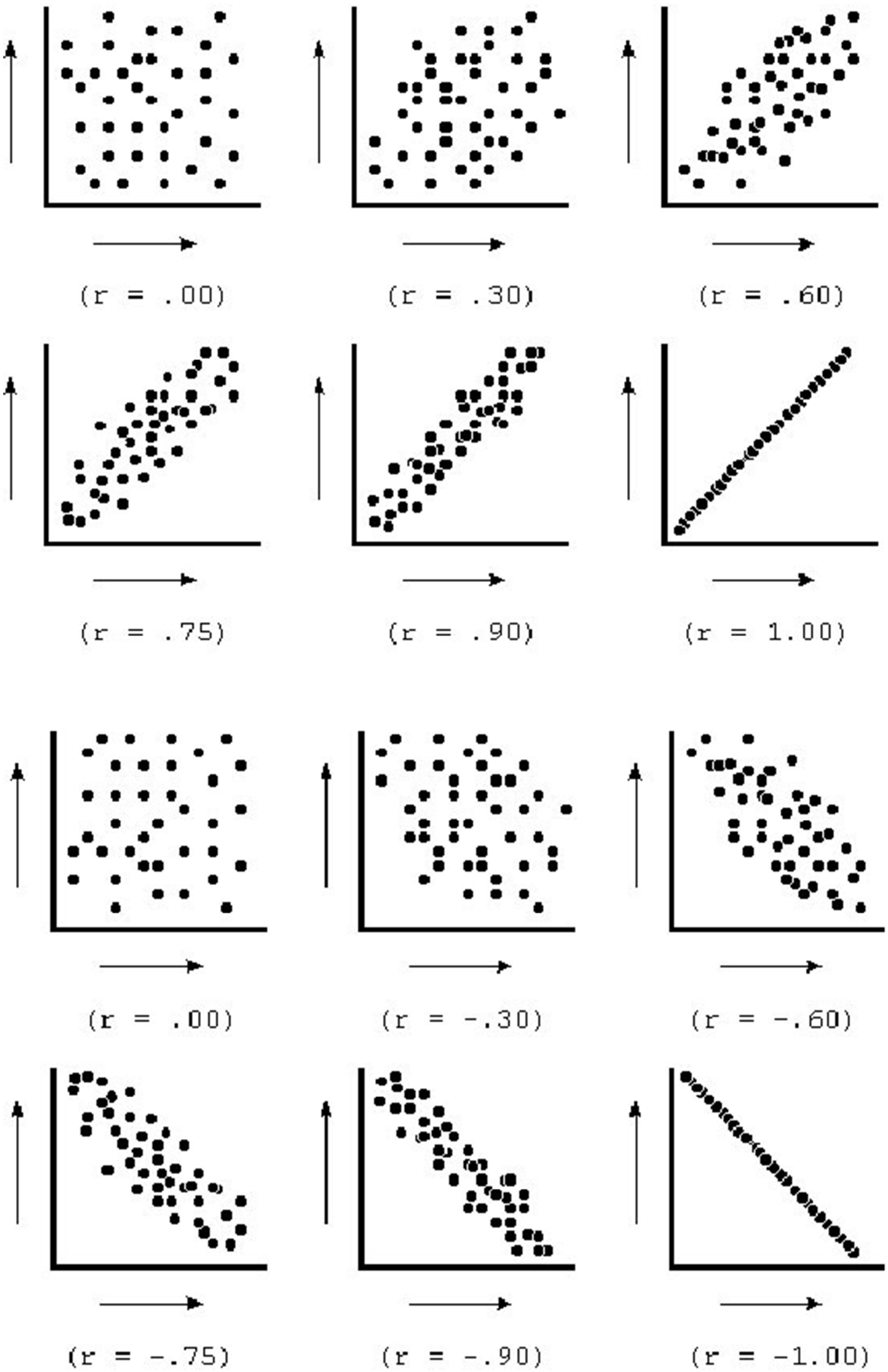
Figure 3: Scatterplots with lines representing the general trend of the relationship



2. Properties of Pearson's r

- linear: r can only measure linear relationships; note that it is possible for curvilinear relationships to exist, e.g., anxiety and performance
- range of values: r is bounded by +1.00 and -1.00; a perfect positive relationship is represented by 1.00; a perfect negative relationship is represented by -1.00; the closer r is to 0.00, the weaker the linear relationship between the two variables (see figure below for examples of various correlation values)
- r = 0.00: zero correlation indicates the weakest possible relationship, that is, no linear relationship between two variables; an r of 0.00 does not rule out all possible relationships since there is the possibility of a non-linear or curvilinear relationship.
- no variance: if either variable has zero variance ($s^2 = 0.00$), then there is no relationship between the two variables and Pearson's r is undefined (but r = 0.00); if a variable has a variance of 0.00, then it does not vary, and is therefore not a variable—it is a constant
- change of scale: r will remain the same between two variables even if the scale of one (or both) of the variables is changed; e.g., converting X to z or T does not affect the value of r

Figure 4: Scatterplots for Various Degrees of Correlation



Change of Scale

As noted above, the Pearson correlation coefficient is scale independent for linear transformations of a variable. This means correlation variable A and B will produce the same correlation as A and (B / X), A and (B × X), or any other linear transformation of B.

Table 1: Examples of Correlations between A and variations of B

Variable A	Variable B	Variable B × 10	Variable B / 3	Log_{10}(Variable B)
1	2	20	0.6666	0.30103
2	3	30	1.0000	0.47712
3	5	50	1.6666	0.69897
4	6	60	2.0000	0.77815
5	7	70	2.3333	0.84509
6	4	40	1.3333	0.60206
7	9	90	3.0000	0.95424

The logarithm to base 10 is a non-linear transformation, so it will change the correlation between A and B.

(*Side Note*: The Log_{10} transformation provides a power value that is used to find B. For example, $Log_{10}(2) = 0.30103$, so if 10 is raised to this value it will produce = 10^0.30103 ≈ 2.)

The following correlation values are produced for each variable combination:

A and B: r = .80023
A and (B × 10): r = .80023
A and (B / 3): r = .80023
A and Log(B): r ≠ .80023 since Log does not produce a linear transformation.

3. Factors That May Alter Pearson's r

The following may inflate r, deflate r, change the sign of r, or have no effect:

- variability or restriction of range (SAT → GPA; SAT restricted)
- extreme scores
- combined data (i.e., grouping data that should not be grouped)

Examples for each are provided below.

(a) Range Restriction

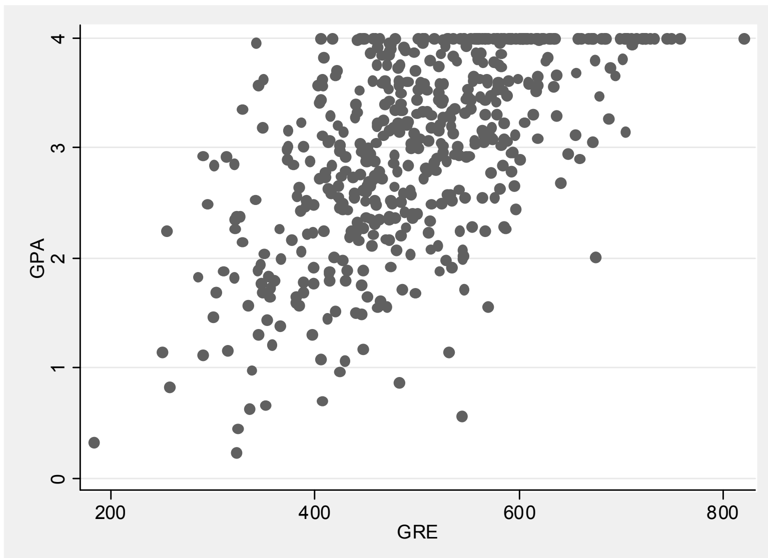
Range restriction does not always affect relationships between variables or correlation values, but sometimes range restriction can affect relationship estimates. For example: Universities frequently make use standardized tests in an effort to screen applicants. Below are descriptive statistics for GRE mathematics scores and first year graduate GPA from 500 students. Note from the statistical output that the correlation is .59.

```
. correlate GRE GPA, means
(obs=500)
    Variable |         Mean    Std. Dev.          Min          Max
-------------+----------------------------------------------------
         GRE |          500          100     183.0000     800.0000
         GPA |     3.000795     .8455727     .2300000            4

             |      GRE      GPA
-------------+------------------
         GRE |   1.0000
         GPA |   0.5932   1.0000
```

As the figure below shows, there is a positive relation between GRE and GPA. Note the ceiling effect of GPA with a number of students earning 4.00 during their first year.

Figure 5: Scatterplot for GRE and GPA



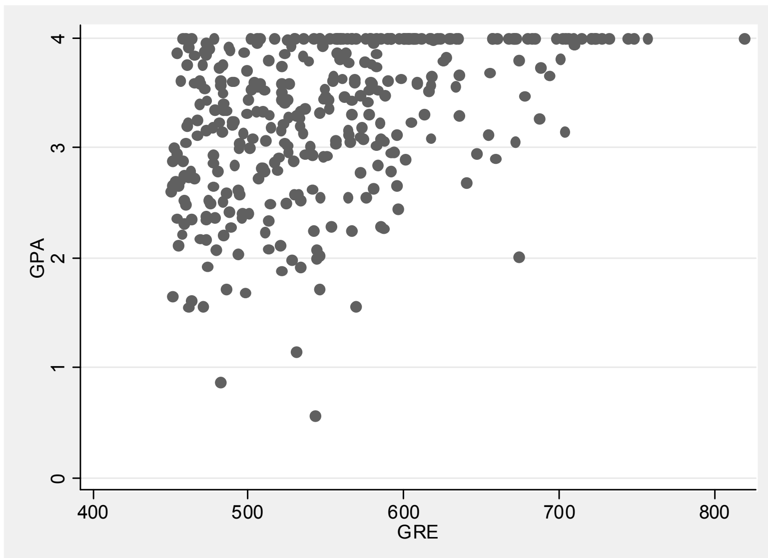
What would happen if students with GRE mathematics scores of only 450 or better are admitted? How might that change the scatterplot and correlation? These data used to generate descriptive statistics and scatterplot above are used again, below, but this time restricting observations to only students with a score of 450 or better on GRE mathematics sub-section.

As results presented in the descriptive statistics and scatterplot show, the correlation between GRE and GPA is reduced as a result of the range restriction placed on GPA, from .59 to .41. Range restriction in a situation like this falsely implies that GRE scores provide less predictive power than is actually the case.

```
. correlate GRE GPA if GRE>=450, means
    Variable |         Mean    Std. Dev.          Min          Max
-------------+----------------------------------------------------
         GRE |     549.8239     72.44123     450.0053     818.9229
         GPA |     3.277684     .6831325      .564881            4

             |      GRE      GPA
-------------+------------------
         GRE |   1.0000
         GPA |   0.4122   1.0000
```

Figure 6: Scatterplot for GRE and GPA with only GRE scores 450 or better



(b) Extreme Scores

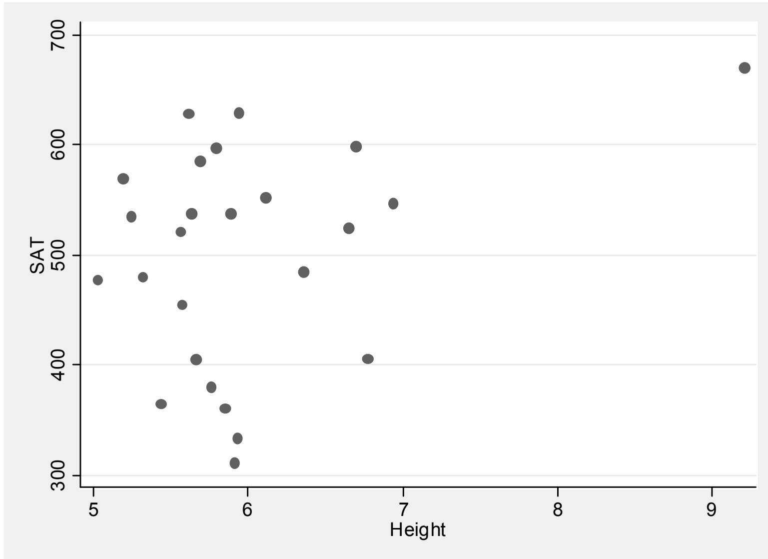
Example 1: SAT and Height

Extreme scores often have some effect on relationship estimates Here's one example: Normally one would think that there should be no relationship between a male's height and his SAT verbal score. However, the data below from 25 men show a moderate correlation of r = .31 height, measured in feet, and SAT verbal scores. How can this be? The scatterplot below provides the answer. Note in figure the extreme score, the outlier, showing one individual with a height of over 9 feet. This is a data entry error.

```
. correlate SAT Height, means
    Variable |         Mean    Std. Dev.          Min          Max
-------------+----------------------------------------------------
         SAT |          500          100     311.5515     670.9754
      Height |     5.988138     .8350335     5.026478          9.2

             |      SAT   Height
-------------+------------------
         SAT |   1.0000
      Height |   0.3138   1.0000
```

Figure 7: Correlation between Height and SAT verbal scores (note extreme score, outlier to right of scatterplot)



Removing the data entry error (9+ foot tall man) provides a correction to the correlation estimate, as the correlation statistic below reveals. The correlation drops from an unlikely .31 to one closer to .00 at .05.

```
. correlate SAT Height if Height<8, means
    Variable |         Mean    Std. Dev.          Min          Max
-------------+----------------------------------------------------
         SAT |      492.876     95.45074     311.5515     629.8821
      Height |     5.854311     .5102792     5.026478     6.934643

             |      SAT   Height
-------------+------------------
         SAT |   1.0000
      Height |   0.0508   1.0000
```

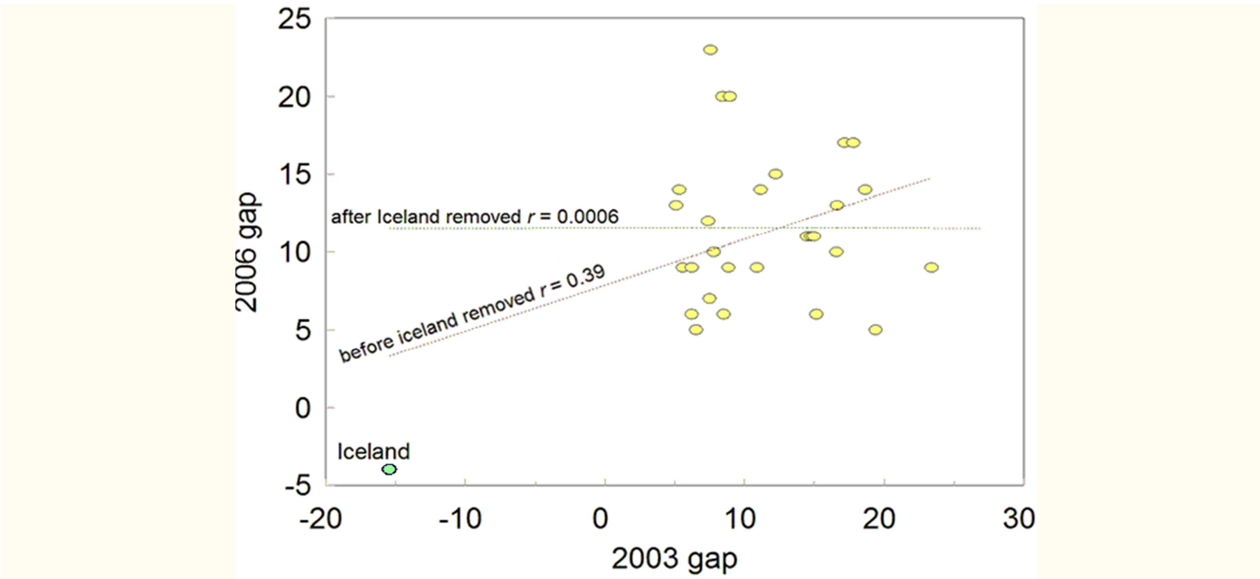
Example 2: Sex Gap in Mathematics Performance

The Program for International Student Assessment (PISA) is an assessment in reading, science, and mathematics administered triennially to students in multiple countries. PISA is designed to enable cross-country comparisons.

Of interest to some researchers is the gap, or achievement difference, in mean mathematics scores between females and males. On average males tend to score higher in mathematics, and this pattern exists across most countries with one exception. One question is whether the size of this gap remains constant over time or varies over time for each country. Below is a scatter plot showing the mathematics gap between males and females for two testing years, 2003 and 2006.

Iceland's gap is curious because females scored higher in mathematics than males, on average, so Iceland produced a negative gap for both 2003 and 2006 while other countries produced a positive gap in both years. Including Iceland in the calculation of the scatterplot suggests a positive association between gap scores – the larger the gap in 2003, the larger the gap in 2006 on average. However, removing Iceland's outlying gap reveals that there is no association between gap scores – the size of the gap in 2003 does not predict the size of the mathematics difference between males and females in 2006.

Figure 8: "A scatter plot of 2006 gaps vs. 2003 gaps. Each point represents the gaps obtained by a single country in successive PISA years. Except for the anomalous Iceland, there is no relation, whatsoever, between gaps observed in different years (r = 0.0006)." Scatter plot and Figure description quoted from http://www.lagriffedulion.f2s.com/math2.htm



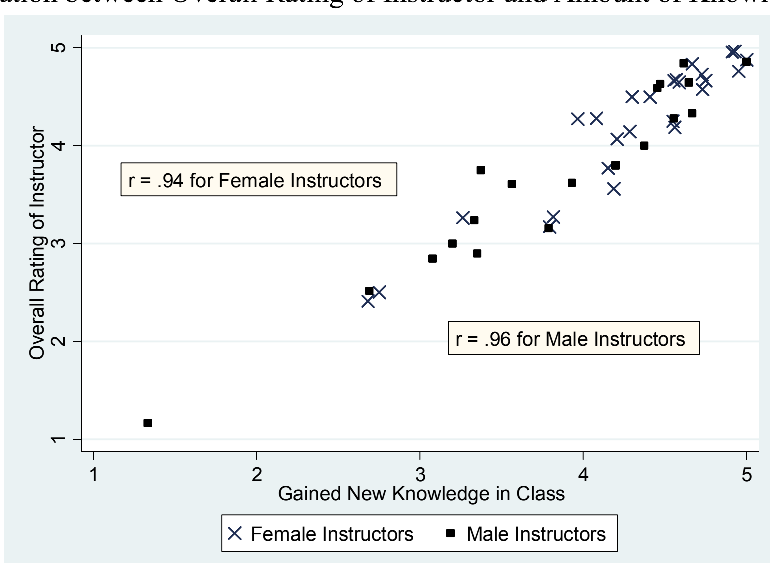
(c) Combined Data

Combining Group Data May Not Affect Estimates

Often when one combines data from different groups the results are similar as the example below illustrates.

For instructors in the College of Education at GSU, the correlations between Overall Rating of the Instructor (1 = Poor to 5 = Excellent) and Amount of Knowledge Gained (1 = Little to None to 5 = Large Amount) is .94 for female and .96 for male instructors. Given the similarity of the scatter plot and correlations, instructor sex can be ignored in this analysis.

Figure 9: Correlation between Overall Rating of Instructor and Amount of Knowledge Gained in Course

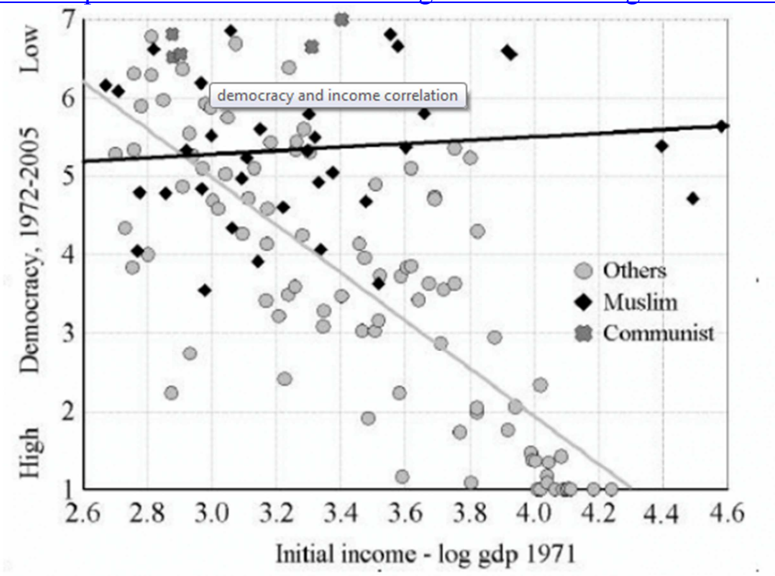


Combined Group Data May Affect Estimates

Example 1: Democracy and Income (by Muslim vs. others)

Sometimes, but not always, combined data can produce relationship estimates that are incorrect. Below is a scatter plot that shows the association between country income (GDP 1971) and democracy scores (1972 to 2005). Three groups are plotted, Muslim countries, Communist countries, and others. If this grouping were ignored, the correlation would be weak. However, when the groupings are considered, one can see two pronounced associations are revealed. For Muslim and communist counties, there is little association between income and democracy. For other countries, there is a strong positive association (it appears to be negative in the scatterplot, but democracy is reverse scored so the association is actually positive).

Figure 10: Average Democracy Score (1972 to 2005) and Country Income (1971), Source: http://filipspagnoli.wordpress.com/stats-on-human-rights/statistics-on-gross-domestic-product-correlations/



Example 2: Class Size and Mean Grade Discrepancy

Students in a number of College of Education classes at GSU were asked to complete a student ratings of instruction questionnaire. Two questions about grades were asked, and both are listed below:

40. What grade do you think the instructor will assign you in this course? (Circle one)

F	D-	D	D+	C-	C	C+	B-	B	B+	A-	A	A+
1	2	3	4	5	6	7	8	9	10	11	12	13

41. What grade do you think you deserve in this course? (Circle one)

F	D-	D	D+	C-	C	C+	B-	B	B+	A-	A	A+
1	2	3	4	5	6	7	8	9	10	11	12	13

These grades were converted to a 4.0 scale, and then a grade discrepancy was calculated:

$$Grade\ Assigned\ (or\ Expected) - Grade\ Deserved = Grade\ Discrepancy$$

Thus a Grade Discrepancy of -1 suggests students believe they deserve one letter grader higher than they will be assigned; a Grade Discrepancy of -.5 suggest a deserved grade half a letter grade lower than will be assigned; a Grade Discrepancy of 0.0 indicates a deserved grade that is equal to the assigned grade.

Below is the scatter plot for all classes sampled. Note there appears to be no relation between Grade Discrepancy and Class Size since r = -.01.

Figure 11: Scatter plot of Grade Discrepancy and Class Size

A curious feature of these data results when separate scatter plots are developed based upon the class instructor's sex. See the figure below for Grade Discrepancy by Class Size scatter plots for female and male instructors separately.

As the figure below shows, the relationship between Grade Discrepancy and Class Size demonstrates completely reversed associations depending upon whether the instructor is female or male. For female instructors larger class sizes are associated with a less grade discrepancy – as class size increases the grade discrepancy moves from -1 to about 0.00 (r = .30). For male instructors the relationship is reversed: larger class sizes are associated with larger negative discrepancies – as class size increases the grade discrepancy moves from about 0.0 to about -1.0.

Figure 12: Scatter plot of Grade Discrepancy and Class Size

(d) Moral of the Story: Visually Inspect Your Data

One should always visually inspect data to ensure a basic understanding of the nature of those data. Without visual inspection, one may mistakenly interpret and present results that are in error. This recommendation – to visually inspect one's data – holds for all types of data; it applies to correlations and scatter plots, and also to any type of data collected and any relevant graphical display. Although not a graphical display, a frequency distribution is an efficient means of quickly spotting data entry errors, for example.

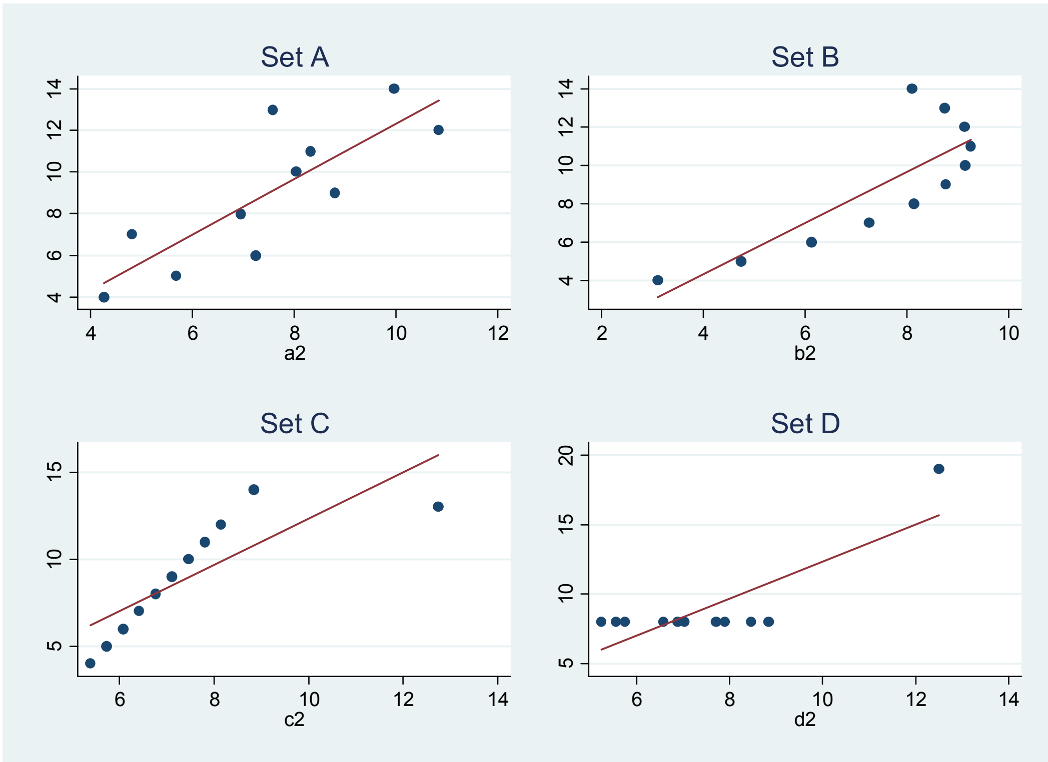
As a final example of the importance of visual inspection, consider Anscombe's (1973) quartet (source: http://en.wikipedia.org/wiki/Anscombe's_quartet). He provided four data sets which are listed below.

Table 2: Anscombe's Quartet, Raw Scores and Descriptive Statistics

	Set A		Set B		Set C		Set D	
	10	8.04	10	9.14	10	7.46	8	6.58
	8	6.95	8	8.14	8	6.77	8	5.76
	13	7.58	13	8.74	13	12.74	8	7.71
	9	8.81	9	8.77	9	7.11	8	8.84
	11	8.33	11	9.26	11	7.81	8	8.47
	14	9.96	14	8.1	14	8.84	8	7.04
	6	7.24	6	6.13	6	6.08	8	5.25
	4	4.26	4	3.1	4	5.39	19	12.5
	12	10.84	12	9.13	12	8.15	8	5.56
	7	4.82	7	7.26	7	6.42	8	7.91
	5	5.68	5	4.74	5	5.73	8	6.89
Mean =	9.00	7.50	9.00	7.50	9.00	7.50	9.00	7.50
SD =	3.32	2.03	3.32	2.03	3.32	2.03	3.32	2.03
r =	.816		.816		.816		.816	

Note that each set in Anscombe's quartet has the same M (9.00 and 7.50), SD (3.32 and 2.03), and correlation coefficient value (r = .816). Without visual aids to help examine these data one may be tempted to claim these data demonstrate similar relationships. Below, however, are the scatterplots to show the differences among the four sets of data.

Figure 11: Scatter plots of Anscombe's Quartet




```
****begin Stata commands****
. twoway (scatter a1 a2) (lfit a1 a2) , title(Set A) legend(off)  name(a)
. twoway (scatter b1 b2) (lfit b1 b2) , title(Set B) legend(off)  name(b)
. twoway (scatter c1 c2) (lfit c1 c2) , title(Set C) legend(off)  name(c)
. twoway (scatter d1 d2) (lfit d1 d2) , title(Set D) legend(off)  name(d)
. graph combine a b c d
****end Stata commands****
```

4. Correlation and Causation

A correlation between variables does not imply the existence of causation, i.e., X → Y or Y → X. A strong correlation does not imply causation (e.g., r = .98; fire trucks and damage in urban areas), neither does a weak correlation imply the lack of causation (e.g., r = .04). Causation can only be established via experimental research and replications. Correlational research (i.e., non-experimental research) cannot be used to establish the existence of causal relationships, although correlational research can provide evidence that a causal relation exists.

5. Pearson's r Formulas and Calculation Examples

Pearson r Formula

Pearson's r, or the *Product Moment Coefficient of Correlation*, is a measure of the degree of linear relationship or association between two (usually quantitative) variables; the population correlation is denoted as ρ (Greek rho), and r refers to correlation obtained from a sample.

Below are three formulas for calculating Pearson's r.

Formula A: $r = \frac{\sum Z_x Z_y}{n-1}$

where n - 1 is the sample size minus 1, z_x are the z scores on variable X, z_y are the z scores on variable Y, and r is the Pearson's correlation coefficient:

Formula B: $r = \frac{s_{xy}}{s_x s_y}$

where s_x is the standard deviation of variable X, s_y is the standard deviation of variable y, and s_{xy} is the *covariance* of variables X and Y; s_{xy} is computed as:

$$s_{xy} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{n-1} = \frac{\sum X_i Y_i - n(\overline{X})(\overline{Y})}{n-1}$$

Covariance is a measure of the degree to which two variables share common variance and vary together or change together (i.e., variables tend to move together).

Formula C: $r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$

Calculation of Pearson r

What is the correlation between IQ and SAT scores? The data are provided below.

Table 3: Fictional SAT and IQ Scores

Student	SAT	IQ
Bill	1010	100
Beth	1085	101
Bryan	1080	102
Bertha	990	95
Barry	970	87
Betty	1100	120
Bret	990	99

For formula A,

Formula A: $r = \frac{\sum Z_x Z_y}{n-1}$

first calculate Z scores for both variables (see two tables below). Recall that the formula for a Z score is

Z = (X – M)/SD

Table 4: Z Scores for IQ

Student	IQ	$(\text{IQ} - \overline{\text{IQ}})$	$(\text{IQ} - \overline{\text{IQ}})^2$	Z_{IQ}
Bill	100	-0.571	0.326	-0.057
Beth	101	0.429	0.184	0.043
Bryan	102	1.429	2.042	0.143
Bertha	95	-5.571	31.036	-0.558
Barry	87	-13.571	184.172	-1.360
Betty	120	19.429	377.486	1.947
Bret	99	-1.571	2.468	-0.157

$\overline{IQ}$ = 100.571; SS = 597.714; s^2_{IQ} = 99.619; s_{IQ} = 9.981

Table 5: Z Scores for SAT

Student	SAT	$(\text{SAT} - \overline{\text{SAT}})$	$(\text{SAT} - \overline{\text{SAT}})^2$	Z_{SAT}
Bill	1010	-22.143	490.312	-0.409
Beth	1085	52.857	2793.862	0.976
Bryan	1080	47.857	2290.292	0.884
Bertha	990	-42.143	1776.032	-0.778
Barry	970	-62.143	3861.752	-1.148
Betty	1100	67.857	4604.572	1.253
Bret	990	-42.143	1776.032	-0.778

$\overline{SAT}$ = 1032.143; SS = 17592.854; s^2_{SAT} = 2932.142; s_{SAT} = 54.149

Next, find the sum of the product of the Z scores.

Table 6: Product of Z Scores

Student	SAT	IQ	Z_{SAT}	Z_{IQ}	$Z_{IQ} \times Z_{SAT}$
Bill	1010	100	-0.409	-0.057	0.023
Beth	1085	101	0.976	0.043	0.042
Bryan	1080	102	0.884	0.143	0.126
Bertha	990	95	-0.778	-0.558	0.434
Barry	970	87	-1.148	-1.360	1.561
Betty	1100	120	1.253	1.947	2.440
Bret	990	99	-0.778	-0.157	0.122

$\sum Z_{IQ} Z_{SAT} = 4.748$

Formula A: $r = \frac{\sum Z_{IQ} Z_{SAT}}{n-1} = \frac{4.748}{6} = .791$

For formula B, one must first calculate the covariance, s_{xy}, between both variables.

Formula B: r = $\frac{s_{XY}}{s_X s_Y}$, where $s_{xy} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{n-1} = \frac{\sum X_i Y_i - n(\overline{X})(\overline{Y})}{n-1}$

Table 7: Two Variables Multiplied

Student	SAT	IQ	$SAT \times IQ$
Bill	1010	100	101000
Beth	1085	101	109585
Bryan	1080	102	110160
Bertha	990	95	94050
Barry	970	87	84390
Betty	1100	120	132000
Bret	990	99	98010

$\overline{IQ}$ = 100.571; SS_{IQ} = 597.714; s^2 = 99.619; s_{IQ} = 9.981

$\overline{SAT}$ = 1032.143; SS_{SAT} = 17592.854;, s^2 = 2932.142; s_{SAT} = 54.149

$\sum SAT \times IQ =$ 729195

$$s_{xy} = s_{IQ\ SAT} = \frac{s_{IQSAT}}{s_{IQ} s_{SAT}} = \frac{\sum (X_i - \overline{X})(Y_i - \overline{Y})}{n-1} = \frac{\sum X_i Y_i - n(\overline{X})(\overline{Y})}{n-1} = \frac{729195 - (7 \times 100.571 \times 1032.143)}{7-1}$$

$$= \frac{729195 - (726625.576)}{6} = \frac{2569.424}{6} = 428.237$$

Formula B: r = $\frac{s_{XY}}{s_X s_Y} = \frac{s_{IQSAT}}{s_{IQ} s_{SAT}} = \frac{428.237}{9.981 \times 54.149} = \frac{428.237}{540.461} = 0.792$

Formula C was used before the computers gained in popularity due to its ease of calculation (despite that this formula looks more complex than the other two!).

Formula C: r = $\frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}}$

Table 8: Formula C Begins

Student	SAT	IQ	IQxSAT	SAT^2	IQ^2
Bill	1010	100	101000	1020100	10000
Beth	1085	101	109585	1177225	10201
Bryan	1080	102	110160	1166400	10404
Bertha	990	95	94050	980100	9025
Barry	970	87	84390	940900	7569
Betty	1100	120	132000	1210000	14400
Bret	990	99	98010	980100	9801

$\sum SAT = 7{,}225$; $\sum(SAT^2) = 7{,}474{,}825$; $\sum$□IQ = 704; $\sum(IQ^2) = 71{,}400$; and $\sum(SAT*IQ) = 729{,}195$

Formula C: $r = \frac{n\sum XY - (\sum X)(\sum Y)}{\sqrt{[n\sum X^2 - (\sum X)^2][n\sum Y^2 - (\sum Y)^2]}} = \frac{n\sum(IQ \times SAT) - (\sum IQ)(\sum SAT)}{\sqrt{[n\sum IQ^2 - (\sum IQ)^2][n\sum SAT^2 - (\sum SAT)^2]}}$

$$= \frac{(7 \times 729195) - (7225 \times 704)}{\sqrt{[(7 \times 7474825) - 52200625][(7 \times 71400) - 495616]}}$$

$$= \frac{5104365 - 5086400}{\sqrt{123150 \times 4184}} = \frac{17965}{\sqrt{515259600}} = \frac{17965}{22699.330} = 0.791$$

As the above results show, all three formulas produce the save value for r within rounding error.

6. Statistical Inference with the Correlation Coefficient r

(a) Calculating a t-ratio for r

When testing a correlation coefficient, one wishes to know whether the correlation coefficient is statistically different from a value of 0.00 (i.e., is calculated correlation statistically different from no linear relationship, r = 0.00). The formula for obtaining a calculated t-value for the correlation is:

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

and the df (or υ) are

df = n - 2

where n is the *number of pairs of scores* (or the number of subjects, the sample size).

(b) Hypotheses for r

Non-directional

H_1: $\rho \neq 0.00$
H_0: $\rho = 0.00$

Directional (one-tail tests):

(a) Lower-tail (r is negative)

H_1: $\rho < 0.00$
H_0: $\rho = 0.00$

(b) Upper-tail (r is positive)

H_1: $\rho > 0.00$
H_0: $\rho = 0.00$

(c) Statistical Significance Level and Decision Rules for t-values

The statistical significance level, alpha (α), is usually set at the conventional .10, .05, or .01 level. Critical values one uses for testing the correlation are the same t-values used above for the one sample t-test, or critical correlation values.

Decision rules

The *decision rules* for the test of the correlation follow:

(a) Two-tailed tests

If $t \leq -t_{crit}$ or $t \geq t_{crit}$, then reject H_0; otherwise, fail to reject H_0

or, alternatively

If |t calculated| ≥ |t critical| then reject H_0; otherwise, fail to reject H_0

(b) One-tailed test (upper-tailed, a hypothesized positive r)

If $t \geq t_{crit}$, then reject H_0; otherwise, fail to reject H_0

(c) One-tailed test (lower-tailed, a hypothesized negative r)

If $t \leq - t_{crit}$, then reject H_0; otherwise, fail to reject H_0

Note that t_{crit} symbolizes the critical t-value.

(d) An Example with t-values

Is there a relationship between the number of hours spent studying and performance on a statistics test? Surveying 17 students in a statistics class, the correlation found between hours studied and test grade was r = .42. Test this correlation at the .01 level of significance to determine whether this sample of students comes from a population with $\rho = 0$ (i.e., no linear relationship between the variables).

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{.42\sqrt{17-2}}{\sqrt{1-.42^2}} = \frac{.42\sqrt{17-2}}{\sqrt{1-.1764}} = \frac{1.627}{.908} = 1.792$$

A two-tailed test with $\alpha = .01$ and df = 17 - 2 = 15, has a critical value of $\pm$ 2.947.

Decision rule: If |1.792| ≥ |2.947| reject Ho, otherwise *fail to reject Ho*

Since 1.792 is less than 2.947 one would fail to reject Ho.

(e) Second Example with t-values

Test whether there is sufficient evidence to reject Ho: $\rho = 0.00$ for the relationship between academic self-efficacy and test anxiety. For a sample (n = 197) of college students, the correlation obtained is r = -.52. Set the probability of a Type 1 error to .01.

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-.52\sqrt{197-2}}{\sqrt{1-(-.52)^2}} = \frac{-.52\sqrt{195}}{\sqrt{1-(.2704)}} = \frac{-7.2614}{0.8542} = -8.50$$

With $\alpha = .01$ and df = 197 - 2 = 195, the critical t-ratio would be about $t_{crit} \approx \pm 2.61$.

Decision rule: If |8.50| ≥ |2.61| *reject Ho*

Since -8.50 is greater than 2.61 in absolute value the null will be rejected and one would conclude there is evidence of negative association between academic self-efficacy and test anxiety.

(f) Critical r values, r_{crit}

Note that for statistical tests of correlation coefficients, there is an easier procedure than calculating t-ratios from correlation coefficients. Once you have:

(1) identified the correct H_0 and H_1,
(2) set the significance level (α),
(3) calculated the correlation,
(4) and calculated the df,
(5) use table of critical correlation values (r_{crit}) for r – see linked table on course web page under the notes for correlation,
(6) apply the appropriate decision rule.

As practice, find the critical r value for the following (assume non-directional alternatives for each):

- r = -.33, n = 17, α = .05
- r = .17, n = 165, α = .05
- r = -.83, n = 10, α = .01
- r =.47, n = 55, α = .01
- r = .21, n = 86, α = .10

Decision rules using r_{crit}

(a) Two-tailed tests

If $r \leq -r_{crit}$ or $r \geq r_{crit}$, then reject H_0; otherwise, fail to reject H_0

or, alternatively,

If |r calculated| ≥ |r critical| then reject H_0; otherwise, fail to reject H_0

(b) One-tailed test (upper-tailed, a hypothesized positive r)

If $r \geq r_{crit}$, then reject H_0; otherwise, fail to reject H_0

(c) One-tailed test (lower-tailed, a hypothesized negative r)

If $r \leq -r_{crit}$, then reject H_0; otherwise, fail to reject H_0

Recall the two examples presented above:

Example 1: Hours studying and statistics test scores

- r = .42
- alpha = .01
- Ho: ρ = 0.00
- n = 17

Find the critical r and apply the decision rule.

Since n = 17 the df = n -2 so df = 15, therefore $r_{critical} = \pm 0.606$

Decision Rule: If |.42| ≥ |.606| reject Ho otherwise *fail to reject Ho*

Since the obtained r of .42 is less than the critical r of .606 one would fail to reject Ho.

Example 2: Academic self-efficacy and test anxiety

- r = -.52
- alpha = .01
- Ho: ρ = 0.00
- n = 197

Find the critical r and apply the decision rule.

Since n = 197 the df = n -2 so df = 195, therefore $r_{critical} = \pm 0.208$ (use df = 150 since table does not provide df = 195).

Decision Rule: If |-.52| ≥ |.208| *reject Ho*

Since the absolute value of r is .52 and since .52 is larger than .208 (the critical r), one would reject Ho.

(g) Hypothesis Testing with p-values

If using statistical software to perform hypothesis testing, one simply compares the obtained p-value for the correlation, r, against α to determine statistical significance. Note that most software reports, by default, p-values for two-tailed tests. The decision rule is:

If p $\leq$ α, then reject H_0; otherwise, fail to reject H_0

Example 1: Hours studying and statistics test scores

- r = .42
- alpha = .01
- Ho: ρ = 0.00
- n = 17

For r = .42 and n = 17 is the corresponding p = 0.093.

Decision Rule: If .093 ≤ .01 reject Ho otherwise *fail to reject Ho*

Since the p is larger than α, one would fail to reject Ho.

Example 2: Academic self-efficacy and test anxiety

- r = -.52
- alpha = .01
- Ho: ρ = 0.00
- n = 197

The p-value for r = -.52 and n = 197 is less than .0001.

Decision Rule: If .0001 ≤ .01 *reject Ho*

Since p is less than α one would reject Ho.

(h) Exercises

(1) A researcher finds the following correlation between GPA and SAT scores: r = .56, n = 19. Test for the statistical significance of this correlation with α = .01 for Ho: ρ = 0.00. Use both the critical t and critical r methods for testing the correlation.

(2) A researcher finds a correlation of r = .139 between academic self-efficacy and academic performance. Is this correlation statistically different from zero? (Note: n = 30, α = .05, and use a two-tailed test.)

(3) What is the smallest sample size is needed in (2) to reject Ho?

7. Correlation Matrices (and SPSS worked example)

Often researchers calculate correlation coefficients among several variables. A convenient method for displaying these correlations is via a correlation matrix. Below are the correlations among IQ, SAT, GRE, and GPA. Usually such tables include a footnote such as this: * p < .05. The asterisk denotes that a particular correlation is statistically different from 0.00 at the .05 level. If the asterisk is not next to a particular correlation, that means the null hypothesis was not rejected for that correlation. The dashed lines, ---, denote perfect correlations (r = 1.00). The correlation between a variable and itself is always equal to 1.00.

	IQ	SAT	GRE	GPA
IQ	---			
SAT	.75*	---		
GRE	.81*	.82*	---	
GPA	.45*	.36	.42	---

* p < .05

Which correlations are statistically significant?

Example: College GPA, Pretest, and Posttest

Data were collected from a number of students at GSU for a classroom experiment. Students reported their current GPA, completed a pretest to measure initial knowledge of the content, experienced an instructional treatment for several weeks, and then completed a posttest of content knowledge. Below are scores from a sample of 13 students who completed the study.

Produce a correlation matrix of these variables using software such as SPSS.

Table 9: Pre-test, Posttest, and GPA Scores

GPA	Pretest	Posttest
2.8	21	80
3.4	55	89
2.2	25	60
3.6	42	91
2.9	54	82
3.2	38	79
2.6	50	88
2.4	41	50
2.7	54	79
3.2	44	79

8. APA Style Presentation of Results

(a) Table of Correlations – The table below provides an example correlation matrix of results. The data represent Ed.D. students reported levels of anxiety and efficacy toward doctoral study, their graduate GPA, and sex.

Table 10: Correlations and Descriptive Statistics for Anxiety and Efficacy toward Doctoral Study, Graduate GPA, and Sex of Student

	1	2	3	4
1. Doctoral Anxiety	---			
2. Doctoral Efficacy	-.43*	---		
3. Graduate GPA	-.24*	.31*	---	
4. Sex	-.11	.19*	-.02	---
M	3.20	4.12	3.92	0.40
SD	1.12	1.31	0.24	0.51
Scale Min/Max Values	1 to 5	1 to 5	0 to 4	0, 1
Cronbach's α	.83	.76	---	---

Note. Sex coded Male = 1, Female = 0; n = 235.
* p < .05.

(b) Interpretation of Results – For inferential statistical tests, one should provide discussion of inferential findings (was null hypothesis rejected; are results statistically significant), and follow this with interpretation of results. The focus of this study was to determine whether anxiety and efficacy toward doctoral study are related, and whether any sex differences for doctoral students are present for anxiety and efficacy.

> Statistical analysis reveals that efficacy toward doctoral study was negatively and statistically related, at the .05 level of significance, to students' reported level of anxiety toward doctoral study, and positively related with students' sex. There was not a statistically significant relationship between student sex and doctoral study anxiety. These results indicate that students' who have higher levels of anxiety about doctoral study also tend to demonstrate lower levels of efficacy toward doctoral work. The positive correlation between sex and efficacy must be interpreted within the context of the coding scheme adopted for the variable sex where 1 = males and 0 = females. Since the correlation is positive, this means that males hold higher average efficacy scores than do females. Lastly, there is no evidence in this sample that anxiety toward doctoral study differs between males and females; both sexes appear to display similar levels of anxiety when thinking about doctoral work.

9. Proportional Reduction in Error (PRE); Predictable Variance; Variance Explained

(a) Interpretation of r

Recall that the coefficient r indicates direction of linear association and, to a lesser extent, strength of association; the closer the value r to + 1.00 (or -1.00), the stronger the linear relationship, while the closer to 0.00 the weaker the linear relationship.

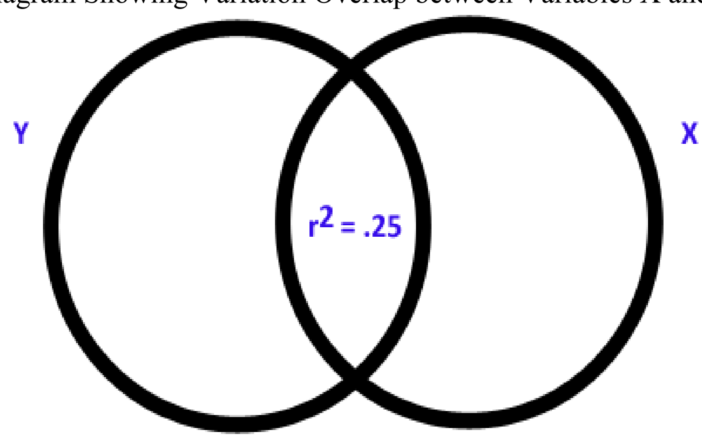
(b) Interpretation of r^2

The correlation squared – the *coefficient of determination* (r^2) – represents the:

- proportional reduction in error in predicting Y;
- proportion of predictable variance accounted for in Y given X;
- proportion of variance overlap between two variables; or the
- extent of co-variation in percentage terms.

Simply put, r^2 is a measure of the amount of variability, in proportions, that overlaps between two variables, and can be illustrated graphically as denoted below.

Figure 12: Venn Diagram Showing Variation Overlap between Variables X and Y



The coefficient of determination, r^2, is a measure of the strength of the relationship between two variables; the larger r^2, the stronger the relationship. The value r^2 ranges from 0.00 to 1.00; the closer to 1.00, the stronger the relationship.

Viewing r^2 as the Proportional Reduction in Error (PRE) is perhaps easiest to understand. Assume that one is trying to predict freshmen college GPA. Suppose one knows, from previous years, that freshmen GPA ranges from 2.00 to 4.00 for students in a given university, with a mean of 3.00. Without know anything else about students, one's best prediction for the likely values of GPA for a group of freshmen is the mean of 3.00, with a range of 2.00 to 4.00. If the mean for a given group of students is 2.50, then our predicted mean of 3.00 is in error. If the mean for another group of students is 3.1, then our predicted mean of 3.00 is again in error. Is there any way to reduce the amount of error one has in making predictions for GPA? The answer is yes if one has access to additional information about each student.

Now suppose additional information about each student is available, such as their high school class rank and their rank based upon SAT scores. By using this information, it is possible to reduce errors in prediction. While correlation coefficients are not designed to provide prediction equations (regression is used for that purpose), squared correlation coefficients can provide information about the extent to which prediction error will be minimized. The squared correlation coefficient, r^2, indicates the proportional reduction in error that will result for knowing additional information, such as SAT rank.

(c) r^2, Proportional Reduction in Error Illustrated

Below is a correlation matrix showing college GPA correlates .116 with High School Class Rank (HS_rank) and .697 with a student's SAT score rank (SAT_rank). Since both correlations are positive, they indicate that as rank, either HS or SAT, increases, GPA also increases.

```
. correlate GPA HS_rank SAT_rank, means
    Variable |         Mean    Std. Dev.          Min          Max
-------------+----------------------------------------------------
         GPA |          3.1           .3     2.162229     4.016737
     HS_rank |           50     28.93181            0          100
    SAT_rank |           50     28.93181            0          100

             |      GPA  HS_rank SAT_rank
-------------+---------------------------
         GPA |   1.0000
     HS_rank |   0.1166   1.0000
    SAT_rank |   0.6970   0.2829   1.0000
```

If one created a regression equation using highs school class rank, the error in predicting college GPA would be reduced by $r^2 = .116^2 = .013$ or 1.3%. If, however, one were to use SAT score rank, the PRE (proportional reduction in error) would be $r^2 = .697^2 = .486$ or 48.6%, a big improvement over using just HS rank.

This reduction in error is loosely illustrated in Figures 8 and 9. Figure 8 shows the relationship between SAT rank and college GPA, while Figure 9 shows the scatterplot for HS class rank and college GPA. Both scatterplots have a gray band behind the scatter. This gray band represents a interval of predicted values. Note that for students with an SAT Rank of 0.00, the gray band ranges from a GPA of 2.20 to about 3.30, while for students with a HS Class Rank equal to 0.00, the range of predicted GPA falls between about 2.25 to 3.80. Note that the band of predicted values is tighter, narrower, for SAT rank than for HS class rank thus reflecting the better prediction capabilities for SAT rank.

Figure 13: Scatterplot with College GPA with SAT Rank

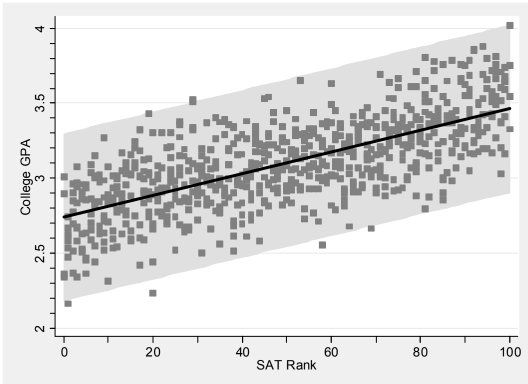
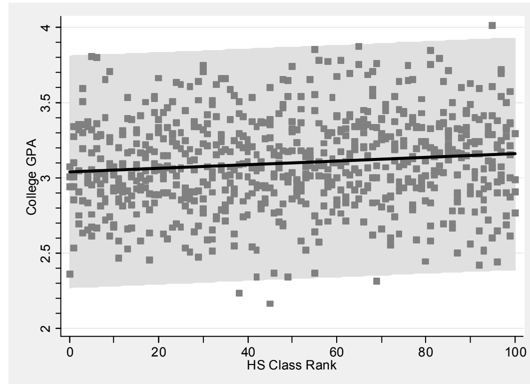


Figure 14: Scatterplot with College GPA with HS Class Rank



Ignore marks below

```
. use "C:\Documents\GSU\COURSES\EDUR 8131\Data and Graphs\GPA HS_rank SAT_rank.dta", clear
. twoway  (lfitci GPA SAT_rank, stdf level(99)) (scatter GPA SAT_rank) (lfit GPA SAT_rank, clpattern(solid) clwidth(thick)), ytitle(College
GPA) xtitle(SAT Rank) legend(off) ytick(2(.1)4)  scheme(s2mono) xtick(0(10)100)
. twoway  (lfitci GPA HS_rank, stdf level(99)) (scatter GPA HS_rank) (lfit GPA HS_rank, clpattern(solid) clwidth(thick)), ytitle(College GPA)
xtitle(HS Class Rank) legend(off) ytick(2(.1)4)  scheme(s2mono) xtick(0(10)100)
```

Ignore marks above

10. Alternative Correlations

Pearson's r assumes both variables are continuous with an interval or ratio scale of measurement. Alternative measures of association exist that do not make this assumption.

- Spearman Rho (Rank Order Correlation), r_{ranks}: appropriate for two ordinal variables which are converted to ranks; once variables are converted to ranks, simply apply the Pearson's r formula to the ranks to obtain r_{ranks}; if untied ranks exist (i.e., no ties exist), then the following formula will simplify calculation of r_{ranks}:

$$r_{ranks} = 1 - \frac{6\sum D^2}{n(n^2-1)}$$

where D refers to the difference between the ranks on the two variables.

- Phi Coefficient, φ or r_φ: appropriate for two dichotomous variables (a nominal variable with only two categories is referred to as dichotomous); one may apply Pearson's r to the two dichotomous variables to obtain r_φ, although simplified formulas exist.
- Point-Biserial Coefficient, r_{pb}: this correlation is appropriate when one variable is dichotomous and the other is continuous with either an interval or ratio scale of measurement; as with the other two, r_{pb} is simply Pearson's r, although a simplified formula exists.
- Gamma: Like Pearson’s r, gamma ranges from -1.0 to 1.0 and a value of 0.00 suggests no association. Gamma us well suited for measures of association among ordinal variables, but can also be used for two dichotomous variables.
- Somer’s D: Like gamma, Somer’s D is also suited for ordinal variables and produces an index that ranges from -1.0 to 1.0. .

Table 11: Summary of Correlation Coefficients

Correlation	Coefficient	Variable X	Variable Y
Pearson's r	r	interval, ratio (some ordinal appropriate too)	interval, ratio (some ordinal appropriate too)
Spearman's Rank	r_{ranks}	ordinal (ranked)	ordinal (ranked)
gamma	γ	ordinal (ranked)	ordinal (ranked)
Somer's D	D	ordinal (ranked)	ordinal (ranked)
Phi	r_{φ}	dichotomous	dichotomous
Point-Biserial	r_{pb}	dichotomous	interval, ratio

Note: ranked refers to ordering the original data from highest to lowest and then assigning ordinal ranks, e.g., 1, 2, 3, etc.

11. Exercises

See Course Index, Exercises for additional worked examples.

12. Partitioned Variance for SAT

IGNORE THIS SECTION – NOT COVERED IN EDUR 8131

$r^2 = .7912 = .626$

total variance in y	=	variance predicted	+	variance not predicted
s^2_{SAT}	=	$(r^2)(s^2_{SAT})$	+	$(1 - r^2)(s^2_{SAT})$
2932.142	=	(.626)(2932.142)	+	(1 - .626)(2932.142)
2932.142	=	1835.521	+	1096.621

Proportion explained = 0.626 (in percent, 62.6%)
Proportion not explained = 0.374 (in percent, 37.4%)
Variance explained = 1835.521
Variance not explained = 1096.621