More About Correlations

Different combinations of the slope, variance of X and Y, and error variance can yield the same correlation coefficient. That is, the correlation coefficient depends not only on the nature of the linear relationship between X and Y, but also on the variability of X and Y in the sample.

Correlation

$$r = b_1 \frac{S_X}{S_Y} = b_1 \sqrt{\frac{SS_X}{SS_Y}}$$

Squared Correlation

$$r^2 = \left(b_1 \frac{S_X}{S_Y} \right)^2 = \frac{b_1^2 SS_X}{b_1^2 SS_X + SS_{residual}} = \frac{(N-1) b_1^2 S_X^2}{(N-1) b_1^2 S_X^2 + (N-2) MSE}$$

$$r^2 = 1 - \frac{MSE}{S_Y^2} \left[\frac{N-2}{N-1} \right]$$

Mean Squared Error

$$MSE = \frac{1}{N-2} \sum \left(Y - \hat{Y} \right)^2 \qquad MSE = \frac{\sum e_i^2}{N-2} \qquad MSE = \frac{\left(1 - r^2 \right) SS_Y}{N-2}$$

Standard Error of the Estimate

Standard Error of the Estimate, $S_e = \sqrt{MSE}$

---

1. When all the data points fall on the regression line so that the *MSE* and Standard Error of the Estimate are zero, then r^2 is 1, and r must be +1 or –1.
2. The larger the standard error of the estimate, the smaller the magnitude of the correlation coefficient.
3. This is a restatement of (2) using the *MSE* instead: The larger the *MSE*, the smaller the magnitude of the correlation coefficient.
4. The larger the magnitude of the slope, the larger the magnitude of the correlation coefficient.
5. The more variability there is in *X*, the predictor, the larger the magnitude of the correlation coefficient.
6. The more variability there is in *Y*, the dependent variable, the larger the magnitude of the correlation coefficient.

RESTRICTION IN RANGE, Example from textbook (pp. 482-483)

Total Sample: All Women
The correlation between total cholesterol level and age for females is .506.
The standard deviation for age is 11.69.

Subsample: Women 50 and over
The correlation between total cholesterol level and age for these women is .148.
The standard deviation for age is 6.26.

Subsample: Women under 50
The correlation between total cholesterol level and age for these women is .264.
The standard deviation for age is 6.58.

- Comparisons of correlations between groups may be ambiguous.
- A finding that the correlation is larger in one group than in another might occur because the nature of the linear relationship between X and Y is different in the two groups.
- However, we might also find the same result even if the linear relation is the same in both groups, but the variability of *X (age)* is different.
- If we want to compare the rate at which *Y* changes with *X* in the two groups, we should compare their regression coefficients, not their correlations.

Restriction in Range (i.e., screening on one or more variables) can lead to biased correlation estimates.

Restriction in range is often thought to make the magnitude of the correlation smaller, but it is possible for restriction in range to artificially increase the magnitude of the correlation.

- Small correlation between GRE and success in graduate school is example where restriction in range reduces the magnitude of the correlation coefficient.

- The artificial increase in the magnitude of the correlation might occur when a researcher is comparing the extremes within a population (i.e., excluding the moderate values). For example, the *Bell Curve* authors did this.

- **It is possible to mathematically correct for restriction in range** if you know what the variability is for the entire population**.** r' and $S_{X'}$ refer to the range restricted values.

$$\hat{r} = \frac{r' \dfrac{S_X}{S_{X'}}}{\sqrt{1 + r'^2 \left[\dfrac{S_X^2}{S_{X'}^2} - 1 \right]}}$$

MEASUREMENT ERROR

Two Variable Case

If *X* and *Y* are measured with error, the obtained correlation will underestimate the 'true' correlation that would be obtained if *X* and *Y* could be measured without error.

$$ObservedCorrelation = TrueCorrelation(\mathrm{Re}liability_X)(\mathrm{Re}liability_Y)$$

We can correct for measurement error in *X* and *Y*. Correcting for measurement error is referred to as 'correcting for attenuation'.

To obtain an **estimate** of the relationship between *X* and *Y* that is not limited by the reliability estimates of *X* and *Y*, we use the following formula.

$$CorrectedCorrelation = \frac{ObservedCorrelation}{\mathrm{Re}liability_X * \mathrm{Re}liability_Y}$$

- Correcting for attenuation does not change the significance of the correlation.
- Interpreting the corrected correlation has some difficulties since in the 'real world' it is difficult to measure things without error.

Three or More Variables

Measurement error can overestimate or underestimate the 'true' correlations among the variables. Thus, it is more difficult to correct for such unreliability.

THE SHAPES OF THE *X* AND *Y* DISTRIBUTIONS

The marginal distributions of *X* and *Y* place constraints on the possible values of the correlation between *X* and *Y*.

- If *X* and *Y* have identical, symmetrical distributions, it is possible for any given value of Z_X to be paired with any of the values of Z_Y, and so it is possible for *r* to range between –1 and 1.
- If *X* and *Y* have distributions that are different from one another, or if they are asymmetric, the full range of correlations from –1 to +1 cannot occur, no matter how the values of *X* and *Y* are paired.

COMBINING DATA ACROSS GROUPS

When the data from different groups are combined, the correlation in the resultant data set may not characterize the relation between X and Y in any of the groups.

The problem occurs because the 'aggregate' correlation reflects not only the relations between X and Y, but also the differences among the group means.

The same problem can occur with the slope.

CORRELATIONS BASED ON RATES OR AVERAGES

Correlations based on the *averages* of groups, may not tell us anything useful about the correlations based on the *individuals* within the groups.

Robinson (1950) illustrated the dangers of generalizing from correlations based on means to correlations based on individuals by showing that when measures such as race, national origin, and illiteracy were correlated, the results could differ dramatically depending on the unit of analysis.

Correlation (race, illiteracy) = .203 for individuals
Correlation (race, illiteracy) = .773 for the means of the states
Correlation (race, illiteracy) = .946 for the census tracts

TESTING THE NULL HYPOTHESIS THAT THE CORRELATION IS ZERO

Example 1

Null Hypothesis: There is not a linear relationship between GPA and IQ, $\rho = 0$.

Alternative Hypothesis: There is a linear relationship between GPA and IQ, $\rho \neq 0$.

$$t = r\sqrt{\frac{N-2}{1-r^2}} \quad \text{with } df = N-2 .$$

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
GPA	2.5830	.84833	30
IQ	97.2667	13.27932	30

Correlations

		GPA	IQ
GPA	Pearson Correlation	1	.702**
	Sig. (2-tailed)	.	.000
	N	30	30
IQ	Pearson Correlation	.702**	1
	Sig. (2-tailed)	.000	.
	N	30	30

**. Correlation is significant at the 0.01 level

Example 2

Null Hypothesis: There is not a linear relationship between children's aggression and parents' aggression, $\rho = 0$.

Alternative Hypothesis: There is a linear relationship between children's aggression and parents' aggression, $\rho \neq 0$.

$$t = r\sqrt{\frac{N-2}{1-r^2}} \quad \text{with } df = N-2 .$$

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
PARENT	16.0000	4.24264	10
CHILD	13.0000	3.68179	10

Correlations

		PARENT	CHILD
PARENT	Pearson Correlation	1	.683*
	Sig. (2-tailed)	.	.030
	N	10	10
CHILD	Pearson Correlation	.683*	1
	Sig. (2-tailed)	.030	.
	N	10	10

*. Correlation is significant at the 0.05 level (2-tailed).

What if the assumption of bivariate normality is violated?

- The type I error rates are close to their nominal values, even for skewed distributions unless from a composite distribution. Then, the type I error rates may be 2 or 3 times the nominal alpha.

- Power may be drastically reduced if the bivariate normality assumption is severely violated. In this case, it might be better to use an adusted bootstrapping method to estimate the confidence interval (pp. 495-496).

Total Sample Size Needed to Have a Specified Power with Alpha = .05 for a Test that $\rho = 0$. Table is copied from Cohen (1988, p. 102).

	Predicted Correlation								
Power	**.10**	**.20**	**.30**	**.40**	**.50**	**.60**	**.70**	**.80**	**.90**
.25	167	42	20	12	8	6	5	4	3
.50	385	96	42	24	15	10	7	6	4
.60	490	122	53	29	18	12	9	6	5
2/3	570	142	63	34	21	14	10	7	5
.70	616	153	67	37	23	15	10	7	5
.75	692	172	75	41	25	17	11	8	6
.80	783	194	85	46	28	18	12	9	6
.85	895	221	97	52	32	21	14	10	6
.90	1047	259	113	62	37	24	16	11	7
.95	1294	319	139	75	46	30	19	13	8
.99	1828	450	195	105	64	40	27	18	11

Abbreviated Demographic Questionnaire (Administered at Time 1 only)

1. What is your gender? (circle one) Male Female
2. How many math classes have you successfully completed while in college?________________
3. Have you successfully completed any computer programming classes while in college?_________________
4. What is your major? (circle one) Psychology Nursing Business Other______________

Abbreviated Personality Questionnaire (Administered at Time 1, Time 2, and Time 3)

5. How anxious do you feel when you think about listening to a lecture on statistics?

Not Anxious						Very Anxious
1	2	3	4	5	6	7

6. How anxious do you feel when you think about taking notes from a lecture on statistics?

Not Anxious						Very Anxious
1	2	3	4	5	6	7

7. How anxious do you feel when you think about using the computer to perform statistics?

Not Anxious						Very Anxious
1	2	3	4	5	6	7

8. How anxious do you feel when you think about conducting statistical analyses by hand?

Not Anxious						Very Anxious
1	2	3	4	5	6	7

9. How anxious do you feel when you think about taking a statistics exam?

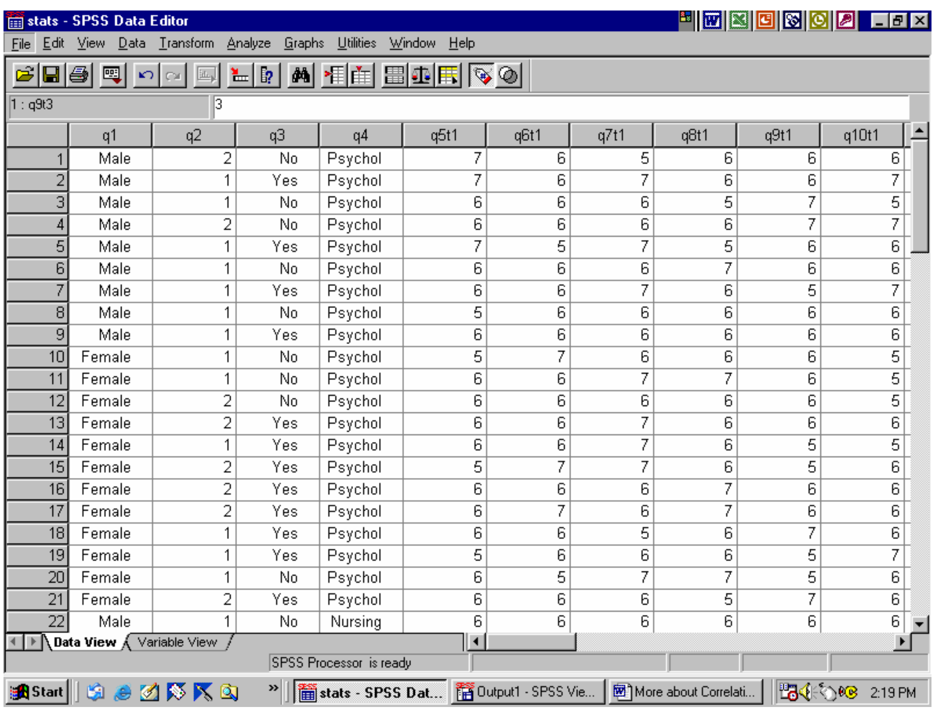
Not Anxious						Very Anxious
1	2	3	4	5	6	7

10. How would you describe your attitude about taking the statistics class?

Very Negative						Very Positive
1	2	3	4	5	6	7

11. How relevant is statistics in your field of study?

Irrelevant						Very Relevant
1	2	3	4	5	6	7



stats - SPSS Data Editor

1 : q9t3 3

	q1	q2	q3	q4	q5t1	q6t1	q7t1	q8t1	q9t1	q10t1
1	Male	2	No	Psychol	7	6	5	6	6	6
2	Male	1	Yes	Psychol	7	6	7	6	6	7
3	Male	1	No	Psychol	6	6	6	5	7	5
4	Male	2	No	Psychol	6	6	6	6	7	7
5	Male	1	Yes	Psychol	7	5	7	5	6	6
6	Male	1	No	Psychol	6	6	6	7	6	6
7	Male	1	Yes	Psychol	6	6	7	6	5	7
8	Male	1	No	Psychol	5	6	6	6	6	6
9	Male	1	Yes	Psychol	6	6	6	6	6	6
10	Female	1	No	Psychol	5	7	6	6	6	5
11	Female	1	No	Psychol	6	6	7	7	6	5
12	Female	2	No	Psychol	6	6	6	6	6	5
13	Female	2	Yes	Psychol	6	6	7	6	6	6
14	Female	1	Yes	Psychol	6	6	7	6	5	5
15	Female	2	Yes	Psychol	5	7	7	6	5	6
16	Female	2	Yes	Psychol	6	6	6	7	6	6
17	Female	2	Yes	Psychol	6	7	6	7	6	6
18	Female	1	Yes	Psychol	6	6	5	6	7	6
19	Female	1	Yes	Psychol	5	6	6	6	5	7
20	Female	1	No	Psychol	6	5	7	7	5	6
21	Female	2	Yes	Psychol	6	6	6	5	7	6
22	Male	1	No	Nursing	6	6	6	6	6	6

Confidence Intervals for a Particular Correlation

First, apply the Fisher Z transform to the sample correlation,

- $Z_r = \frac{1}{2}\ln\left[\frac{1+r}{1-r}\right] =$
- ***Z_r* = arctanh *r* = inverse hyperbolic tangent *r***
- ***Z_r*** can be read directly from the table on page 684 (Table C.11) of the textbook.

Second, calculate the standard error associated with the Fisher Z transform, $\sigma_r = \frac{1}{\sqrt{N-3}}$

Third, calculate the confidence interval for Z_r

- the 95% confidence interval for a particular correlation is $Z_r \pm 1.96\sigma_r$.
- the 99% confidence interval for a particular correlation is $Z_r \pm 2.57\sigma_r$

Convert the confidence interval for Z_r back to r values

- ***r = tanh Z = hyperbolic tangent Z***
- ***r*** can be read from the table on page 684 (C.11) in the textbook.

Interpret the confidence intervals for the correlation

Correlations

		Q7T1	Q8T1
Q7T1	Pearson Correlation	1	.489**
	Sig. (2-tailed)	.	.000
	N	48	48
Q8T1	Pearson Correlation	.489**	1
	Sig. (2-tailed)	.000	.
	N	48	48

**. Correlation is significant at the 0.01 level

Testing the Hypothesis that the Correlation Is a Specific Value Other Than Zero, $\rho = \rho_{hyp}$

First, apply the Fisher Z transform to the sample correlation,

- $Z_r = \frac{1}{2}\ln\left[\frac{1+r}{1-r}\right] =$
- $\boldsymbol{Z_r}$ **= arctanh *r* = inverse hyperbolic tangent *r***
- $\boldsymbol{Z_r}$ can be read directly from the table on page 684 (Table C.11) of the textbook.

Second, apply the Fisher Z transform to the hypothesized correlation,

- $Z_{\text{hyp}} = \frac{1}{2}\ln\left[\frac{1+\rho}{1-\rho}\right] =$
- $\boldsymbol{Z_{hyp}}$ **= arctanh ρ = inverse hyperbolic tangent ρ**
- $\boldsymbol{Z_{hyp}}$ can be read directly from the table on page 684 (Table C.11) of the textbook.

Third, calculate the standard error associated with the Fisher Z transform, $\sigma_r = \frac{1}{\sqrt{N-3}}$

Fourth, calculate the Z test, $Z = \frac{Z_r - Z_{\rho hyp}}{\sigma_r}$.

Compare the Z test results to the appropriate critical values.

- ±1.96 for an alpha of .05.
- ± 2.57 for an alpha of .01.

Interpret the results.

Testing Whether Two Independent Correlations Are Significantly Different, $\rho_1 = \rho_2$

Apply the Fisher Z transform to the sample correlation from **group 1**.

- $Z_{r1} = \frac{1}{2}\ln\left[\frac{1+r_1}{1-r_1}\right] =$
- ***Z_{r1}* = arctanh *r_1* = inverse hyperbolic tangent *r_1***
- ***Z_{r1}*** can be read directly from the table on page 684 (Table C.11) of the textbook.

Apply the Fisher Z transform to the sample correlation from **group 2.**

- $Z_{r2} = \frac{1}{2}\ln\left[\frac{1+r_2}{1-r_2}\right] =$
- ***Z_{r2}* = arctanh *r_2* = inverse hyperbolic tangent *r_2***
- ***Z_{r2}*** can be read directly from the table on page 684 (Table C.11) of the textbook.

Calculate the standard error associated with the Fisher Z transform,

$$\sigma_r = \sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}$$

Calculate the Z test, $Z = \frac{Z_{r1} - Z_{r2}}{\sigma_r}$ **.**

Compare the Z test results to the appropriate critical values.

- ±1.96 for an alpha of .05.
- ± 2.57 for an alpha of .01.

Interpret the results.

Correlations
gender = Male

Correlations[a]

		Q7T1	Q8T1
Q7T1	Pearson Correlation	1	.483*
	Sig. (2-tailed)	.	.036
	N	19	19
Q8T1	Pearson Correlation	.483*	1
	Sig. (2-tailed)	.036	.
	N	19	19

*. Correlation is significant at the 0.05 level (2-tailed).

a. gender = Male

gender = Female

Correlations[a]

		Q7T1	Q8T1
Q7T1	Pearson Correlation	1	.516**
	Sig. (2-tailed)	.	.004
	N	29	29
Q8T1	Pearson Correlation	.516**	1
	Sig. (2-tailed)	.004	.
	N	29	29

**. Correlation is significant at the 0.01 level

a. gender = Female

Sample Size Needed PER GROUP to Have a Specified Power with Alpha = .05 for a Test that Two Independent Correlations Are Different from Each Other, $\rho_1 - \rho_2 = 0$. Table is copied from Cohen (1988, p. 135).

	$q = Z_{r1} - Z_{r2}$										
Power	**.10**	**.20**	**.30**	**.40**	**.50**	**.60**	**.70**	**.80**	**1.00**	**1.20**	**1.40**
.25	333	86	40	24	16	12	10	8	6	5	5
.50	771	195	88	51	34	24	19	15	11	8	7
.60	983	248	112	64	42	30	23	18	13	10	8
2/3	1146	289	130	74	49	35	26	21	14	11	9
.70	1237	312	140	80	52	37	28	22	15	12	9
.75	1391	350	157	90	59	42	31	25	17	13	10
.80	1573	395	177	101	66	47	35	28	19	14	11
.85	1799	452	203	115	75	53	40	31	21	15	12
.90	2104	528	236	134	87	61	46	36	24	18	14
.95	2602	653	292	165	107	75	56	44	29	21	16
.99	3677	922	411	233	150	105	78	60	40	29	22

Testing for Homogeneity of J Independent Correlation Coefficients

Apply the Fisher Z transform to **each** of the sample correlations.

- $Z_j = \frac{1}{2}\ln\left[\frac{1+r_j}{1-r_j}\right]$
- **Z_j = arctanh r_j = inverse hyperbolic tangent r_j**
- ***Z_j*** can be read directly from the table on page 684 (Table C.11) of the textbook.

Group	N_j	r_j	$(N_j - 3)$	Z_j	$(N_j - 3)Z_j$	Z_j^2	$(N_j - 3)Z_j^2$
			Total A		Total B		Total C

$\chi^2 = C - \frac{B^2}{A}$ with $df = J - 1$.

If the χ^2 is equal to or greater than the appropriate Chi-square Critical Value (Table C.4, p. 663), then we would conclude the correlations differ among the groups.

Interpretation

Correlations

major = Psychology

Correlations[a]

		Q7T1	Q8T1
Q7T1	Pearson Correlation	1	.053
	Sig. (2-tailed)	.	.819
	N	21	21
Q8T1	Pearson Correlation	.053	1
	Sig. (2-tailed)	.819	.
	N	21	21

a. major = Psychology

major = Nursing

Correlations[a]

		Q7T1	Q8T1
Q7T1	Pearson Correlation	1	.750*
	Sig. (2-tailed)	.	.020
	N	9	9
Q8T1	Pearson Correlation	.750*	1
	Sig. (2-tailed)	.020	.
	N	9	9

*. Correlation is significant at the 0.05 level (2-tailed).

a. major = Nursing

major = Business

Correlations[a]

		Q7T1	Q8T1
Q7T1	Pearson Correlation	1	-.287
	Sig. (2-tailed)	.	.248
	N	18	18
Q8T1	Pearson Correlation	-.287	1
	Sig. (2-tailed)	.248	.
	N	18	18

a. major = Business

Testing Whether Two $\boxed{\text{Dependent}}$ Correlations Are Significantly Different, $\rho_1 = \rho_2$

Suppose I wanted to know whether the correlation between anxiety about using the computer to perform statistics and anxiety about taking a statistics exam at time 1 (Q7T1, Q9T1) **is the same as** the correlation between anxiety about conducting statistical analyses by hand and anxiety about taking a statistics exam at time 1 (Q8T1, Q9T1).

Y = Q9T1 $\quad\quad$ $X1$ = Q7T1 $\quad\quad$ $X2$ = Q8T1

Calculate the average r^2 that $X1$ and $X2$ have with Y:

r^2(X1,Y) =
r^2(X2,Y) =

$\bar{r}^2 =$

Calculate $f = \dfrac{1 - r_{X1X2}}{2(1-\bar{r}^2)}$ **but cannot be larger than 1.**

Calculate $h = \dfrac{1 - f\bar{r}^2}{1-\bar{r}^2}$.

Calculate Z_{rYX1}.

Calculate Z_{rYX2}.

Calculate the TEST, $z = \left(Z_{rYX1} - Z_{rYX2}\right)\sqrt{\dfrac{N-3}{2\left(1 - r_{X1X2}\right)h}}$.

Compare the Z test results to the appropriate critical values.

- ±1.96 for an alpha of .05.
- ± 2.57 for an alpha of .01.

Interpret the results.

Correlations

Correlations

		Q7T1	Q8T1	Q9T1
Q7T1	Pearson Correlation	1	.489**	.295*
	Sig. (2-tailed)	.	.000	.042
	N	48	48	48
Q8T1	Pearson Correlation	.489**	1	.332*
	Sig. (2-tailed)	.000	.	.021
	N	48	48	48
Q9T1	Pearson Correlation	.295*	.332*	1
	Sig. (2-tailed)	.042	.021	.
	N	48	48	48

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Partial and Part Correlations

Correlations

		age	# alcohol drinks	driving mistakes	cell phone time	# passengers
age	Pearson Correlation	1	-.107	-.492**	-.176	-.158
	Sig. (2-tailed)	.	.302	.000	.088	.126
	N	95	95	95	95	95
# alcohol drinks	Pearson Correlation	-.107	1	.248*	.074	-.061
	Sig. (2-tailed)	.302	.	.016	.477	.558
	N	95	95	95	95	95
driving mistakes	Pearson Correlation	-.492**	.248*	1	.361**	.173
	Sig. (2-tailed)	.000	.016	.	.000	.093
	N	95	95	95	95	95
cell phone time	Pearson Correlation	-.176	.074	.361**	1	.237*
	Sig. (2-tailed)	.088	.477	.000	.	.021
	N	95	95	95	95	95
# passengers	Pearson Correlation	-.158	-.061	.173	.237*	1
	Sig. (2-tailed)	.126	.558	.093	.021	.
	N	95	95	95	95	95

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Zero-order Correlation/Pearson Correlation/Pairwise Correlation

$r_{XY} = corr(X,Y)$ represents the correlation between *X* and *Y*.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Correlations Partial	Correlations Part
1	(Constant)	13.648	1.920		7.107	.000			
	cell phone time	.061	.019	.283	3.241	.002	.361	.320	.279
	age	-.228	.045	-.442	-5.060	.000	-.492	-.467	-.435

a. Dependent Variable: driving mistakes

First-order partial correlation

$r_{XY|W} = corr(X \mid W, Y \mid W)$ represents the correlation between *X* and *Y* after the effects of *W* have been removed from *X* and *Y*.

$$r_{XY|W} = \frac{r_{XY} - r_{XW}r_{YW}}{\sqrt{\left(1 - r_{XW}^2\right)\left(1 - r_{YW}^2\right)}}$$

First-order part correlation

$r_{Y(X|W)} = corr(Y, X \mid W)$ represents the correlation between *X* and *Y* after the effects of *W* have been removed from *X*.

$$r_{Y(X|W)} = \frac{r_{XY} - r_{XW}r_{YW}}{\sqrt{\left(1 - r_{XW}^2\right)}}$$

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Partial	Part
1	(Constant)	12.819	1.928		6.650	.000			
	cell phone time	.056	.019	.260	2.946	.004	.361	.297	.249
	age	-.215	.045	-.417	-4.787	.000	-.492	-.450	-.405
	# alcohol drinks	.513	.235	.187	2.187	.031	.248	.225	.185
	# passengers	.130	.200	.057	.647	.519	.173	.068	.055

a. Dependent Variable: driving mistakes

Coefficients[a]

Model		Unstandardized Coefficients B	Std. Error	Standardized Coefficients Beta	t	Sig.	Correlations Zero-order	Partial	Part
1	(Constant)	41.100	11.649		3.528	.001			
	age	.047	.266	.020	.177	.860	-.176	.019	.017
	# alcohol drinks	.043	1.271	.003	.033	.973	.074	.004	.003
	# passengers	1.918	1.041	.182	1.841	.069	.237	.191	.178
	driving mistakes	1.566	.531	.338	2.946	.004	.361	.297	.284

a. Dependent Variable: cell phone time

Second-order partial correlation

$r_{XY|WQ} = corr(X \mid WQ, Y \mid WQ)$ represents the correlation between *X* and *Y* after the effects of *W* and *Q* have been removed from *X* and *Y*.

Second-order part correlation

$r_{Y(X|WQ)} = corr(Y, X \mid WQ)$ represents the correlation between *X* and *Y* after the effects of *W* and *Q* have been removed from *X*.

The examples shown above contain third-order partial and part correlations.

Significance Tests for Zero-order Correlations

- SPSS: Determine significance by inspecting the correlation matrix.

- By hand: Conduct the test for determining whether $\rho = 0$ that was mentioned earlier in this chapter.

Significance Tests for Partial and Part Correlations

If the partial correlation is significant, so are the part correlation, the unstandardized regression coefficient, and the standardized regression coefficient.

To test whether the partial correlation is significantly <u>different from zero</u>:

- SPSS: Determine significance by inspecting the significance of the t test associated with the predictor.

- By Hand:
  - Calculate $t = r_{partial}\sqrt{\frac{N-2-p}{1-r_{partial}^2}}$

  - Calculate $df = N - 2 - p$ where p is the number of variables partialed out.

  - Find the t critical values from Table C.3 on p. 662 of the textbook.

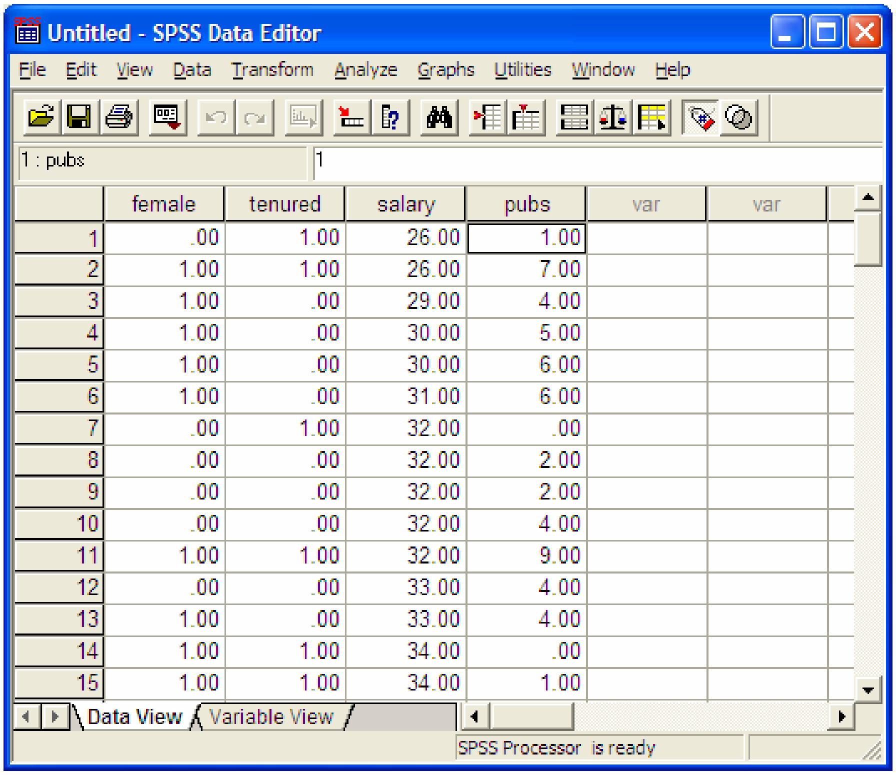
  - The partial correlation is significantly different from zero if the calculated t is more extreme than the critical values.

To test whether the partial correlation is significantly <u>different from a specific value</u>:

- Convert the partial correlation to a Z score using Table C.11 on p. 684 of the text.
- Convert the hypothesized partial correlation to a Z score using Table C.11 on p. 684 of the text.
- Calculate $Z = \left(Z_{Partial} - Z_{hypothesizedpartial}\right)\sqrt{N - 3 - p}$.
- The partial correlation is significantly different than the hypothesized partial correlation if the Z is more extreme than the critical values (± 1.96 for alpha = .05, ± 2.57 for alpha = .01).

To calculate a confidence interval for a partial correlation,

- Convert the partial correlation to a Z score using Table C.11 on p. 684 of the textbook.
- Calculate the upper and lower limits for the Z score:
  - $Z_{partial} \pm 1.96\sqrt{\dfrac{1}{N-3-p}}$ for a 95% confidence interval.
  - $Z_{partial} \pm 2.57\sqrt{\dfrac{1}{N-3-p}}$ for a 99% confidence interval.
- Convert the lower limit of the Z score back to a correlation using Table C.11 on p. 684 of the textbook.
- Convert the upper limit of the Z score back to a correlation using Table C.11 on p. 684 of the textbook.
- Interpret the confidence interval.



	female	tenured	salary	pubs	var	var
1	.00	1.00	26.00	1.00		
2	1.00	1.00	26.00	7.00		
3	1.00	.00	29.00	4.00		
4	1.00	.00	30.00	5.00		
5	1.00	.00	30.00	6.00		
6	1.00	.00	31.00	6.00		
7	.00	1.00	32.00	.00		
8	.00	.00	32.00	2.00		
9	.00	.00	32.00	2.00		
10	.00	.00	32.00	4.00		
11	1.00	1.00	32.00	9.00		
12	.00	.00	33.00	4.00		
13	1.00	.00	33.00	4.00		
14	1.00	1.00	34.00	.00		
15	1.00	1.00	34.00	1.00		

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
FEMALE	.5059	.50293	85
TENURED	.3529	.48072	85
SALARY	38.6118	5.81069	85
PUBS	3.1412	2.21549	85

Correlations

		FEMALE	TENURED	SALARY	PUBS
FEMALE	Pearson Correlation	1	-.304**	-.189	.042
	Sig. (2-tailed)	.	.005	.084	.703
	N	85	85	85	85
TENURED	Pearson Correlation	-.304**	1	.024	.143
	Sig. (2-tailed)	.005	.	.827	.193
	N	85	85	85	85
SALARY	Pearson Correlation	-.189	.024	1	-.193
	Sig. (2-tailed)	.084	.827	.	.077
	N	85	85	85	85
PUBS	Pearson Correlation	.042	.143	-.193	1
	Sig. (2-tailed)	.703	.193	.077	.
	N	85	85	85	85

**. Correlation is significant at the 0.01 level (2-tailed).

Correlations

Descriptive Statistics

	Mean	Std. Deviation	N
math score	49.8941	12.38857	85
verbal score	50.1765	5.83335	85
job performance	20.1176	2.95768	85
VERBPASS	.4353	.49874	85
MATHPASS	.4471	.50014	85

Correlations

		math score	verbal score	job performance	VERBPASS	MATHPASS
math score	Pearson Correlation	1	.563**	.641**	.420**	.807**
	Sig. (2-tailed)	.	.000	.000	.000	.000
	N	85	85	85	85	85
verbal score	Pearson Correlation	.563**	1	.407**	.820**	.499**
	Sig. (2-tailed)	.000	.	.000	.000	.000
	N	85	85	85	85	85
job performance	Pearson Correlation	.641**	.407**	1	.312**	.552**
	Sig. (2-tailed)	.000	.000	.	.004	.000
	N	85	85	85	85	85
VERBPASS	Pearson Correlation	.420**	.820**	.312**	1	.404**
	Sig. (2-tailed)	.000	.000	.004	.	.000
	N	85	85	85	85	85
MATHPASS	Pearson Correlation	.807**	.499**	.552**	.404**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.
	N	85	85	85	85	85

**. Correlation is significant at the 0.01 level (2-tailed).

Nonparametric Correlations

Correlations

			math score	verbal score	job performance	VERBPASS	MATHPASS
Kendall's tau_b	math score	Correlation Coefficient	1.000	.380**	.476**	.344**	.716**
		Sig. (2-tailed)	.	.000	.000	.000	.000
		N	85	85	85	85	85
	verbal score	Correlation Coefficient	.380**	1.000	.280**	.723**	.413**
		Sig. (2-tailed)	.000	.	.000	.000	.000
		N	85	85	85	85	85
	job performance	Correlation Coefficient	.476**	.280**	1.000	.267**	.474**
		Sig. (2-tailed)	.000	.000	.	.004	.000
		N	85	85	85	85	85
	VERBPASS	Correlation Coefficient	.344**	.723**	.267**	1.000	.404**
		Sig. (2-tailed)	.000	.000	.004	.	.000
		N	85	85	85	85	85
	MATHPASS	Correlation Coefficient	.716**	.413**	.474**	.404**	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.
		N	85	85	85	85	85
Spearman's rho	math score	Correlation Coefficient	1.000	.524**	.637**	.414**	.862**
		Sig. (2-tailed)	.	.000	.000	.000	.000
		N	85	85	85	85	85
	verbal score	Correlation Coefficient	.524**	1.000	.381**	.861**	.492**
		Sig. (2-tailed)	.000	.	.000	.000	.000
		N	85	85	85	85	85
	job performance	Correlation Coefficient	.637**	.381**	1.000	.311**	.553**
		Sig. (2-tailed)	.000	.000	.	.004	.000
		N	85	85	85	85	85
	VERBPASS	Correlation Coefficient	.414**	.861**	.311**	1.000	.404**
		Sig. (2-tailed)	.000	.000	.004	.	.000
		N	85	85	85	85	85
	MATHPASS	Correlation Coefficient	.862**	.492**	.553**	.404**	1.000
		Sig. (2-tailed)	.000	.000	.000	.000	.
		N	85	85	85	85	85

**. Correlation is significant at the .01 level (2-tailed).

Pearson's Correlation Coefficient

- Possible values are -1 to 1 if distributions are symmetric and identical.

Point-biserial correlation coefficient

- Pearson's correlation where one variable is continuous and one is dichotomous;
- Identical to the independent groups *t* test.

Phi correlation coefficient-Pearson's correlation where both variables are dichotomous.

- Possible values are –1 to 1 for a 2x2 table. Otherwise, the possible values are 0 to 1. The upper bound may be less than 1, depending on the marginal distributions.

Cramer's V- a modification of the Phi coefficient so the attainable upper bound is always 1.

- Possible values are –1 to 1.

Biserial Correlation Coefficient

- One dichotomous variable and one continuous variable;
- Estimate of what Pearson's correlation coefficient would be if, instead of a dichotomous variable, we actually had normally distributed scores;
- Always at least 25% larger than the point-biserial correlation;
- Under some circumstances, may take on values greater than 1;
- Has typically been discussed in the context of test development theory;
- Should be used only with caution.

Tetrachoric Correlation Coefficient

- Both variables are dichotomous;
- Estimate of what Pearson's correlation coefficient would be if, scores on the two normally distributed underlying variables were available;
- Has typically been discussed in the context of test development theory;
- Should be used only with caution.

Spearman's Correlation Coefficient-Pearson's correlation applied to ranked data.
- Possible values range between –1 and 1.

Kendall's Tau
- It is calculated using ranked data for X and Y. See page 509 in textbook.
- It is a measure of monotonicity, the tendency for the underlying measures to increase or decrease together.
- It is more resistant to outliers than is Pearson's correlation coefficient.
- Possible values range between –1 and 1.

$$\tau = \frac{\text{number of agreements in order - number of disagreements in order}}{\text{total number of pairs}} \quad \text{if there are no ties.}$$

The formula is more complicated when ties are present.

Goodman-Kruskal Gamma Coefficient for Ranked Data
- It is calculated using ranked data for X and Y. See page 509 in textbook.
- It is a measure of monotonicity, the tendency for the underlying measures to increase or decrease together.
- It is more resistant to outliers than is Pearson's correlation coefficient.
- Possible values range between –1 and 1.

$$\gamma = \frac{\text{number of agreements in order - number of disagreements in order}}{\text{number of agreements in order + number of disagreements in order}}$$

Ties are excluded from the calculation for Goodman-Kruskal coefficient.