Correlation

Lecture 05

DISCOVERING STATISTICS USING IBM SPSS STATISTICS

4TH EDITION

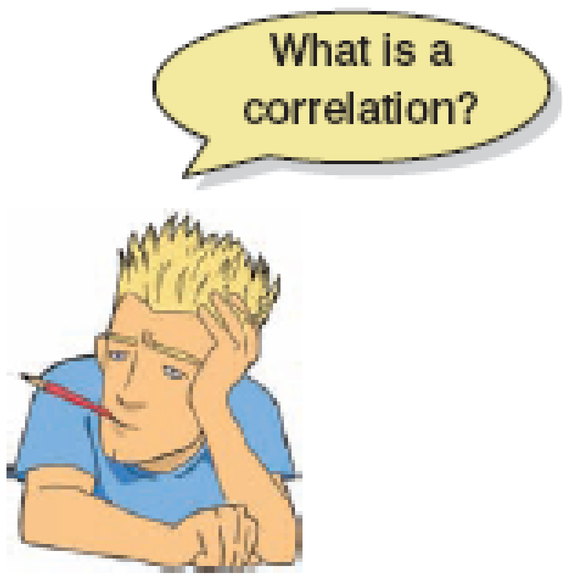
ANDY FIELD

Aims

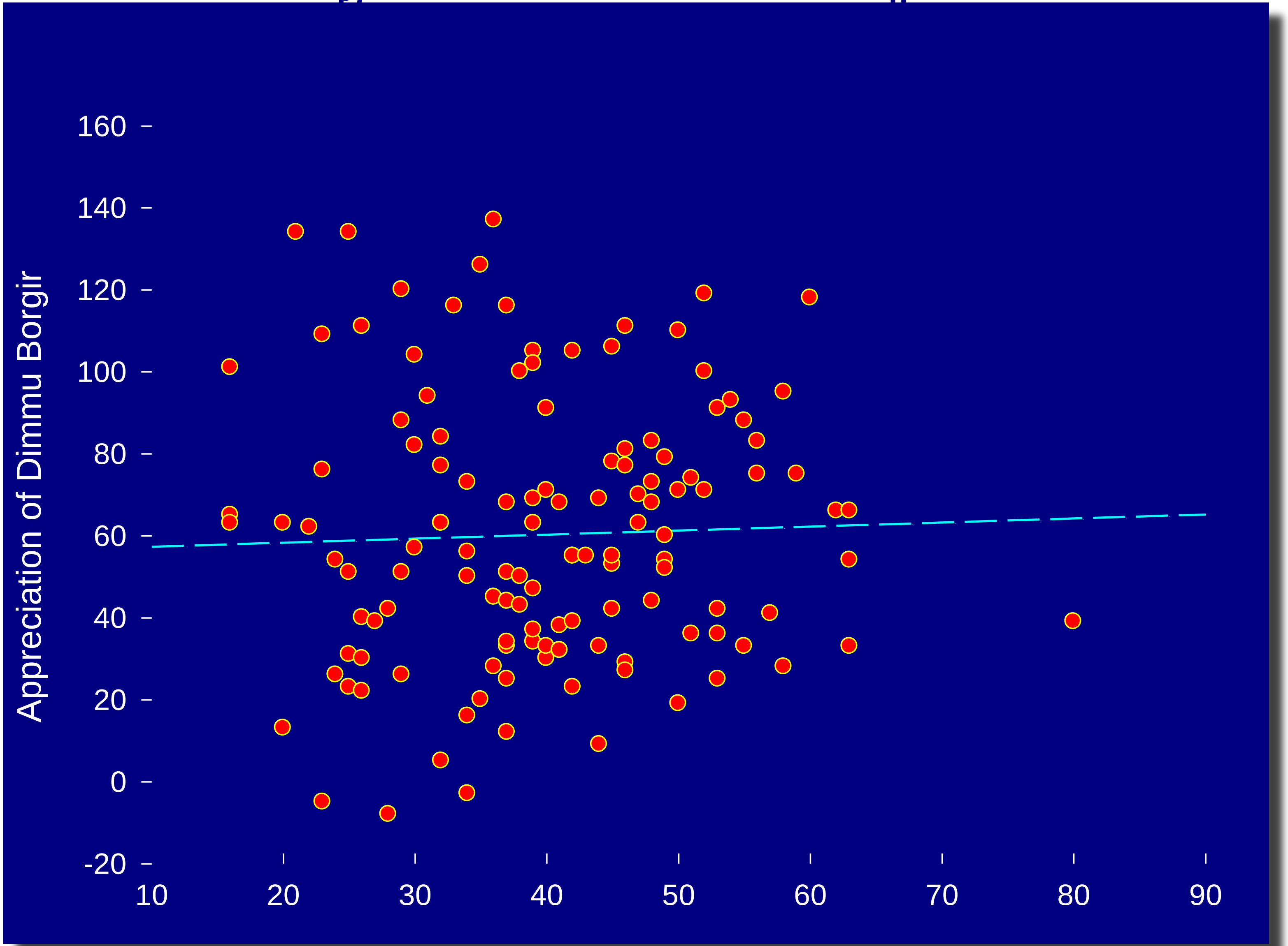
- Measuring Relationships
  - Scatterplots
  - Covariance
  - Pearson's Correlation Coefficient
- Nonparametric measures
  - Spearman's Rho
  - Kendall's Tau
- Interpreting Correlations
  - Causality
- Partial Correlations

What is a Correlation?

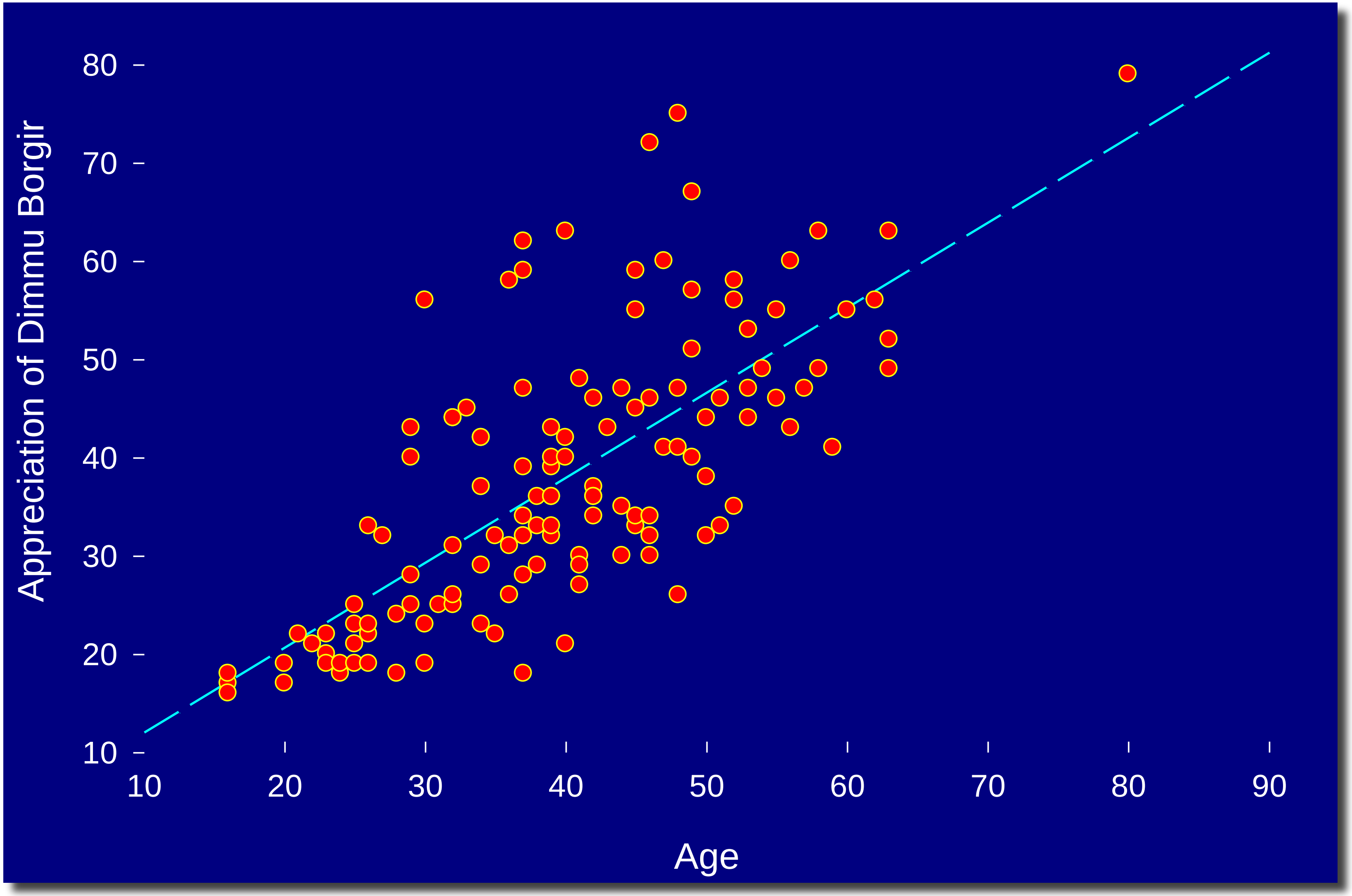
- It is a way of measuring the extent to which two variables are related.
- It measures the pattern of responses across variables.



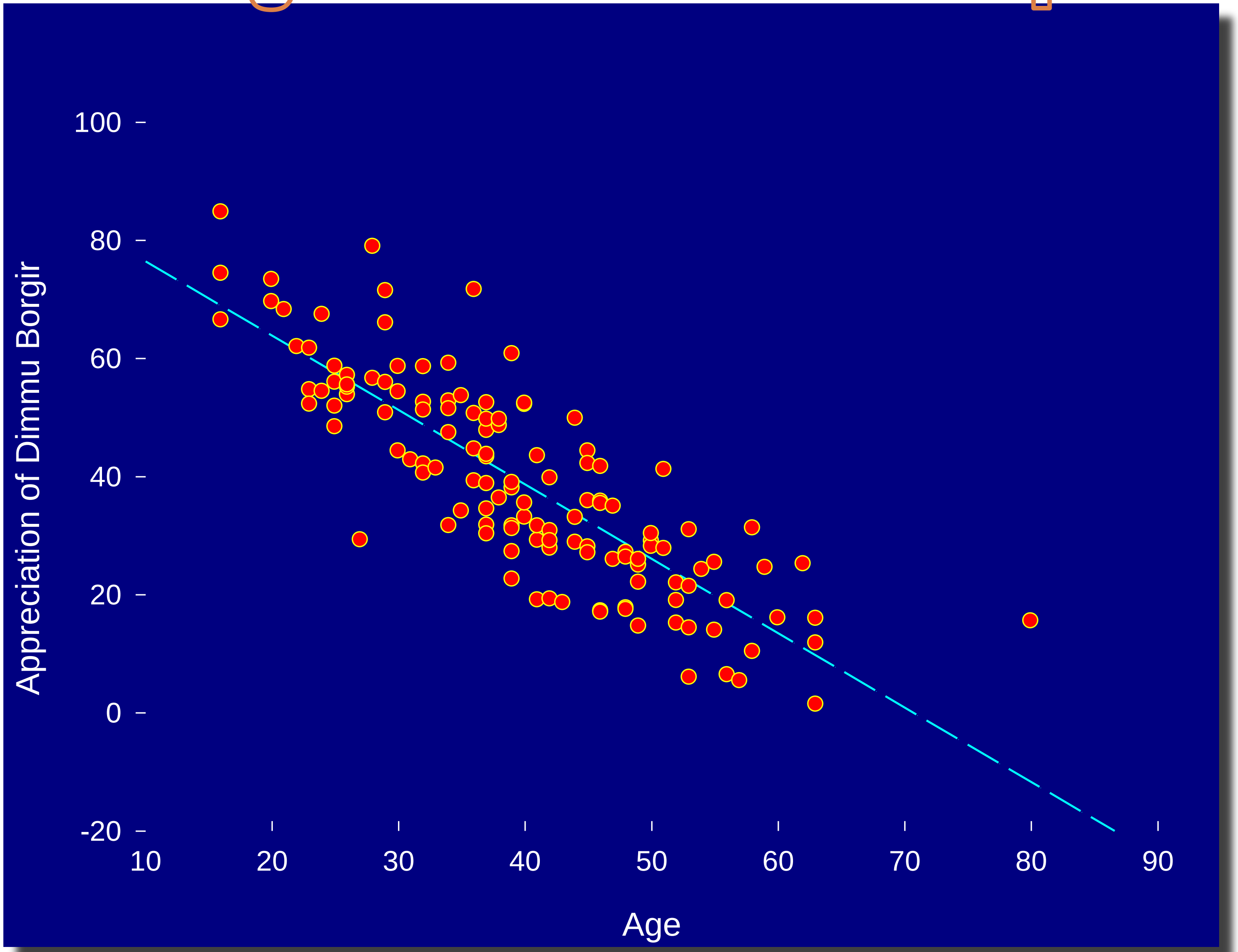
Very Small Relationship



Positive Relationship



Negative Relationship



Measuring Relationships

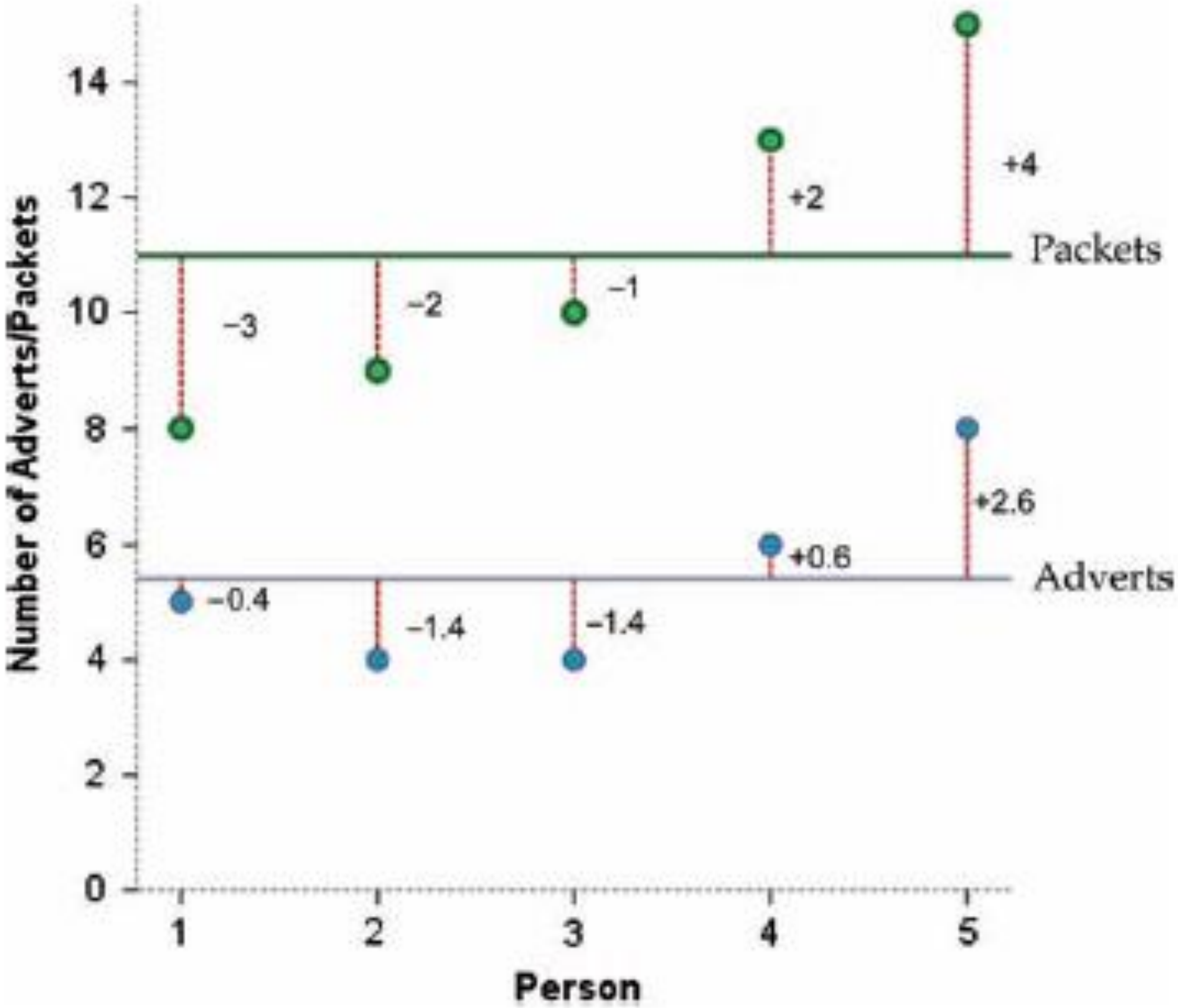
- We need to see whether as one variable increases, the other increases, decreases or stays the same.
- This can be done by calculating the Covariance.
  – We look at how much each score deviates from the mean.
  – If both variables deviate from the mean by the same amount, they are likely to be related.

Modeling Relationships

- First, look at some scatterplots of the variables that have been measured.
- $\text{Outcome}_i = (\text{model}) + \text{error}_i$
- $\text{Outcome}_i = (bX_i) + \text{error}_i$

TABLE 7.1

Participant:	***1***	***2***	***3***	***4***	***5***	***Mean***	***s***
Adverts Watched	5	4	4	6	8	5.4	1.67
Packets Bought	8	9	10	13	15	11.0	2.92


Number of Adverts/Packets
Person
Packets
Adverts
+4
+2
−3
−2
−1
+2.6
+0.6
−0.4
−1.4
−1.4

Revision of Variance

- The variance tells us by how much scores deviate from the mean for a single variable.
- It is closely linked to the sum of squares.
- Covariance is similar – it tells is by how much scores on two variables differ from their respective means.

$$\text{Variance} = \frac{\sum\left(x_i - \bar{x}\right)^2}{N-1} = \frac{\sum\left(x_i - \bar{x}\right)\left(x_i - \bar{x}\right)}{N-1}$$

Covariance

- Calculate the error between the mean and each subject's score for the first variable (x).
- Calculate the error between the mean and their score for the second variable (y).
- Multiply these error values.
- Add these values and you get the cross product deviations.
- The covariance is the average cross-product deviations:

DISCOVERING STATISTICS USING IBM SPSS STATISTICS

4TH EDITION

$$Cov(x, y) = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{N - 1}$$

ANDY FIELD

$$
\begin{aligned}
\operatorname{cov}(x,y) &= \frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{N-1}\\
&= \frac{(-0.4)(-3)+(-1.4)(-2)+(-1.4)(-1)+(0.6)(2)+(2.6)(4)}{4}\\
&= \frac{1.2+2.8+1.4+1.2+10.4}{4}\\
&= \frac{17}{4}\\
&= 4.25
\end{aligned}
$$

Problems with Covariance

- It depends upon the units of measurement.
  - E.g. The Covariance of two variables measured in Miles might be 4.25, but if the same scores are converted to Km, the Covariance is 11.
- One solution: standardise it!
  - Divide by the standard deviations of both variables.
- The standardised version of Covariance is known as the Correlation coefficient.
  - It is relatively affected by units of measurement.

The Correlation Coefficient

$$r = \frac{Cov_{xy}}{s_x s_y}$$

$$= \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{(N-1) s_x s_y}$$

The Correlation Coefficient

$$r = \frac{Cov_{xy}}{s_x s_y}$$

$$= \frac{4.25}{1.67 \times 2.92}$$

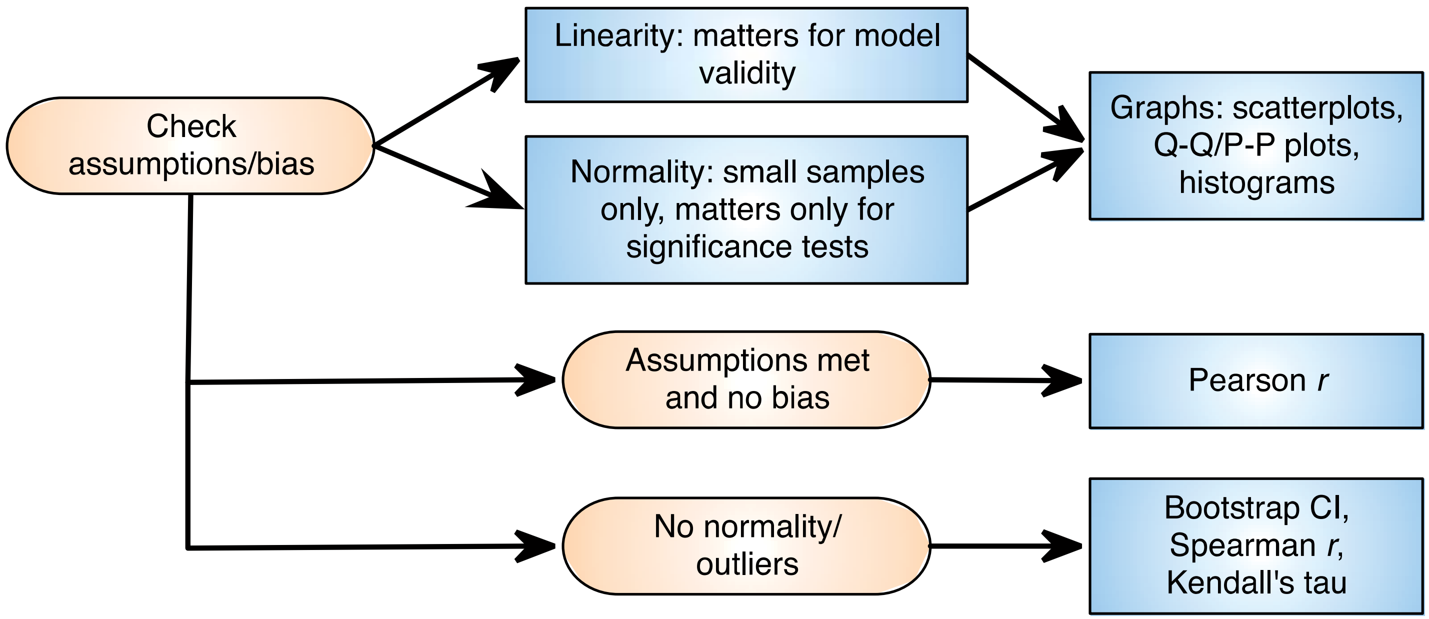
$$= .87$$

Correlation: Example

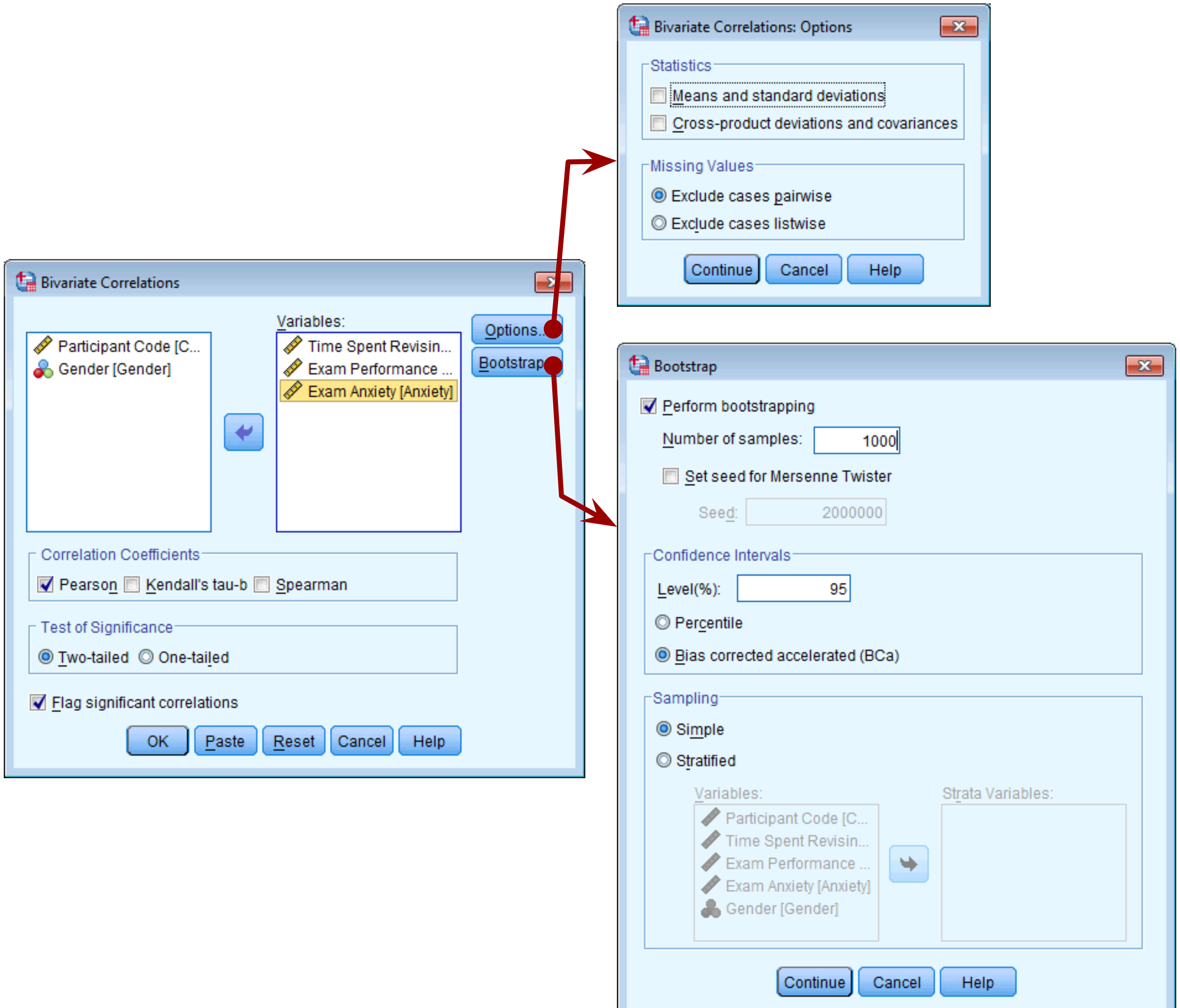
- Anxiety and Exam Performance
- Participants:
  - 103 students
- Measures
  - Time spent revising (hours)
  - Exam performance (%)
  - Exam Anxiety (the EAQ, score out of 100)
  - Gender

Conducting Correlation Analysis

FIGURE 7.5
The general process for conducting correlation analysis



Doing a Correlation



Correlation Output

DISCOVERING STATISTICS USING IBM SPSS STATISTICS

4TH EDITION

OUTPUT 7.1 Output for a Pearson's correlation

Correlation coefficients

Confidence intervals

Correlations

				Time Spent Revising	Exam Performance (%)	Exam Anxiety
Time Spent Revising	Pearson Correlation			1	.397**	-.709**
	Sig. (2-tailed)				.000	.000
	N			103	103	103
	Bootstrap[c]	Bias		0	-.002	-.004
		Std. Error		0	.070	.112
		BCa 95% Confidence Interval	Lower	.	.245	-.863
			Upper	.	.524	-.492
Exam Performance (%)	Pearson Correlation			.397**	1	-.441**
	Sig. (2-tailed)			.000		.000
	N			103	103	103
	Bootstrap[c]	Bias		-.002	0	.004
		Std. Error		.070	0	.065
		BCa 95% Confidence Interval	Lower	.245	.	-.564
			Upper	.524	.	-.301
Exam Anxiety	Pearson Correlation			-.709**	-.441**	1
	Sig. (2-tailed)			.000	.000	
	N			103	103	103
	Bootstrap[c]	Bias		-.004	.004	0
		Std. Error		.112	.065	0
		BCa 95% Confidence Interval	Lower	-.863	-.564	.
			Upper	-.492	-.301	.

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

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

c. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

ANDY FIELD

Reporting the Results

TABLE 7.2 An example of reporting a table of correlations

	Exam Performance	*Exam Anxiety*	*Revision Time*
Exam Performance	1	−.44*** [−.564, −.301]	.40*** [.245,.524]
Exam Anxiety	103	1	−.71*** [−.863, −.492]
Revision Time	103	103	1

ns = not significant ($p > .05$), * $p < .05$, ** $p < .01$, *** $p < .001$. BCa bootstrap 95% CIs reported in brackets.

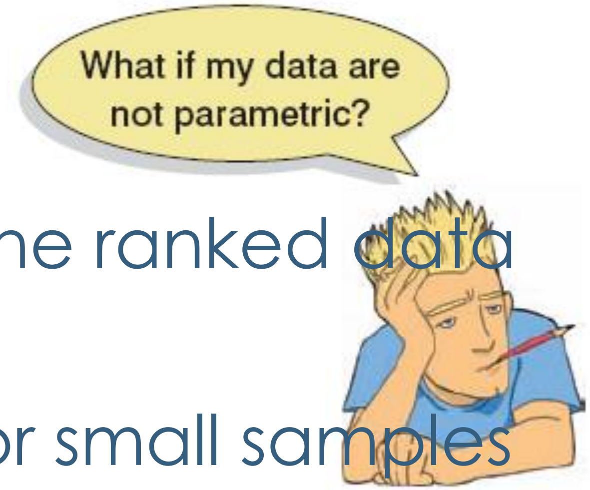
Things to know about the Correlation

- It varies between -1 and +1
  - 0 = no relationship
- It is an effect size
  - ±.1 = small effect
  - ±.3 = medium effect
  - ±.5 = large effect
- Coefficient of determination, r^2
  - By squaring the value of *r* you get the proportion of variance in one variable shared by the other.

Correlation and Causality

- The third-variable problem:
  - in any correlation, causality between two variables cannot be assumed because there may be other measured or unmeasured variables affecting the results.
- Direction of causality:
  - Correlation coefficients say nothing about which variable causes the other to change

Nonparametric Correlation



- Spearman's Rho
  - Pearson's correlation on the ranked data
- Kendall's Tau
  - Better than Spearman's for small samples
- World's best Liar Competition
  - 68 contestants
  - Measures
    - Where they were placed in the competition (first, second, third, etc.)
    - Creativity questionnaire (maximum score 60)

Correlation Output Spearman's rho

Correlations

						Creativity	Position in Best Liar Competition
Spearman's rho	Creativity	Correlation Coefficient				1.000	-.373**
		Sig. (2-tailed)				.	.002
		N				68	68
		Bootstrap[c]	Bias			.000	.007
			Std. Error			.000	.125
			BCa 95% Confidence Interval		Lower	.	-.604
					Upper	.	-.114
	Position in Best Liar Competition	Correlation Coefficient				-.373**	1.000
		Sig. (2-tailed)				.002	.
		N				68	68
		Bootstrap[c]	Bias			.007	.000
			Std. Error			.125	.000
			BCa 95% Confidence Interval		Lower	-.604	.
					Upper	-.114	.

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

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

c. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Correlation Output
Kendall's tau

Correlations

						Creativity	Position in Best Liar Competition
Kendall's tau_b	Creativity	Correlation Coefficient				1.000	-.300**
		Sig. (2-tailed)				.	.001
		N				68	68
		Bootstrap[c]	Bias			.000	.001
			Std. Error			.000	.098
			BCa 95% Confidence Interval	Lower		.	-.491
				Upper		.	-.100
	Position in Best Liar Competition	Correlation Coefficient				-.300**	1.000
		Sig. (2-tailed)				.001	.
		N				68	68
		Bootstrap[c]	Bias			.001	.000
			Std. Error			.098	.000
			BCa 95% Confidence Interval	Lower		-.491	.
				Upper		-.100	.

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

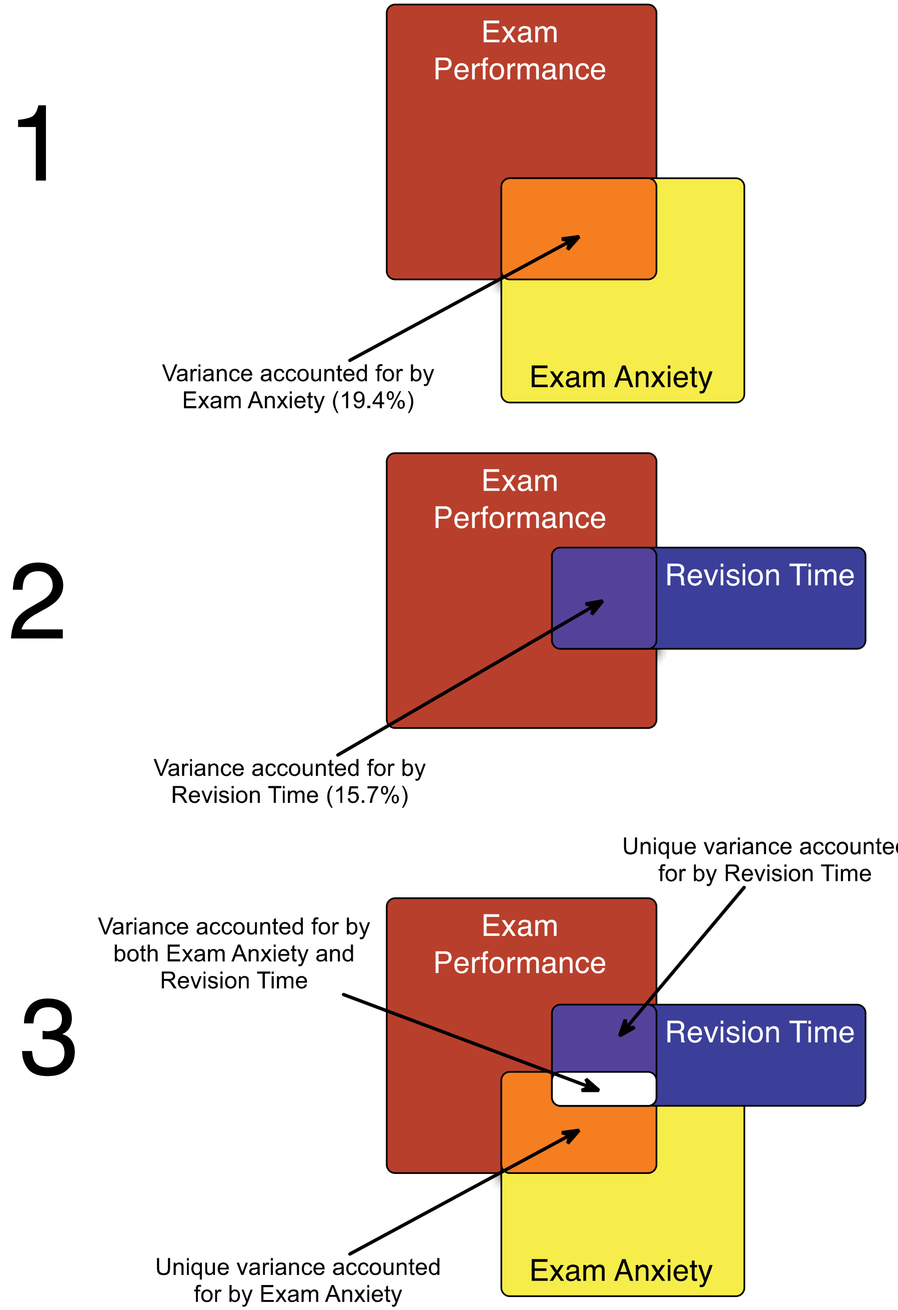
*. Correlation is significant at the 0.05 level (2-tailed).

c. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples

Partial and Semi-Partial Correlations

- Partial correlation:
  - Measures the relationship between two variables, controlling for the effect that a third variable has on them both.
- Semi-partial correlation:
  - Measures the relationship between two variables controlling for the effect that a third variable has on only one of the others.

FIGURE 7.9 Diagram showing the principle of partial correlation



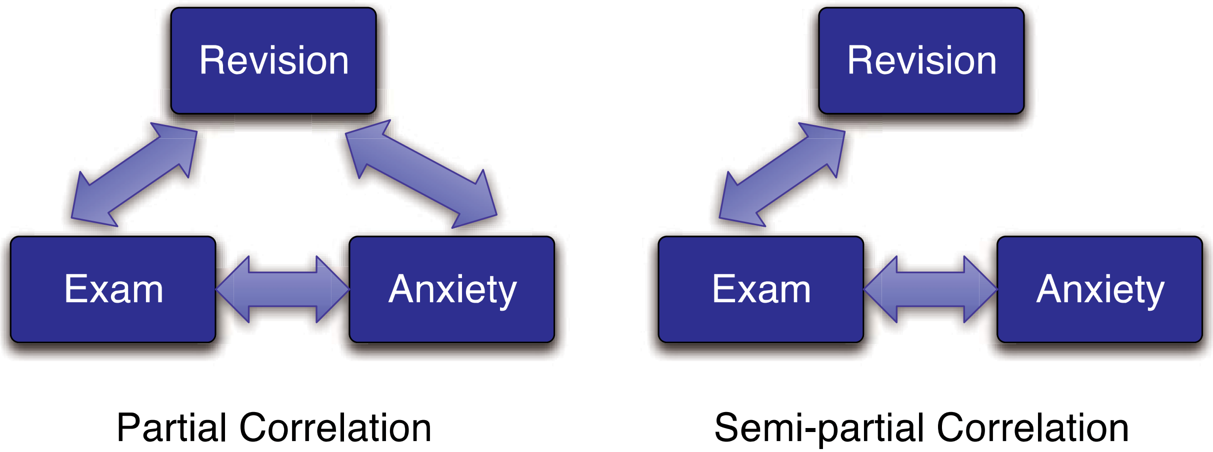
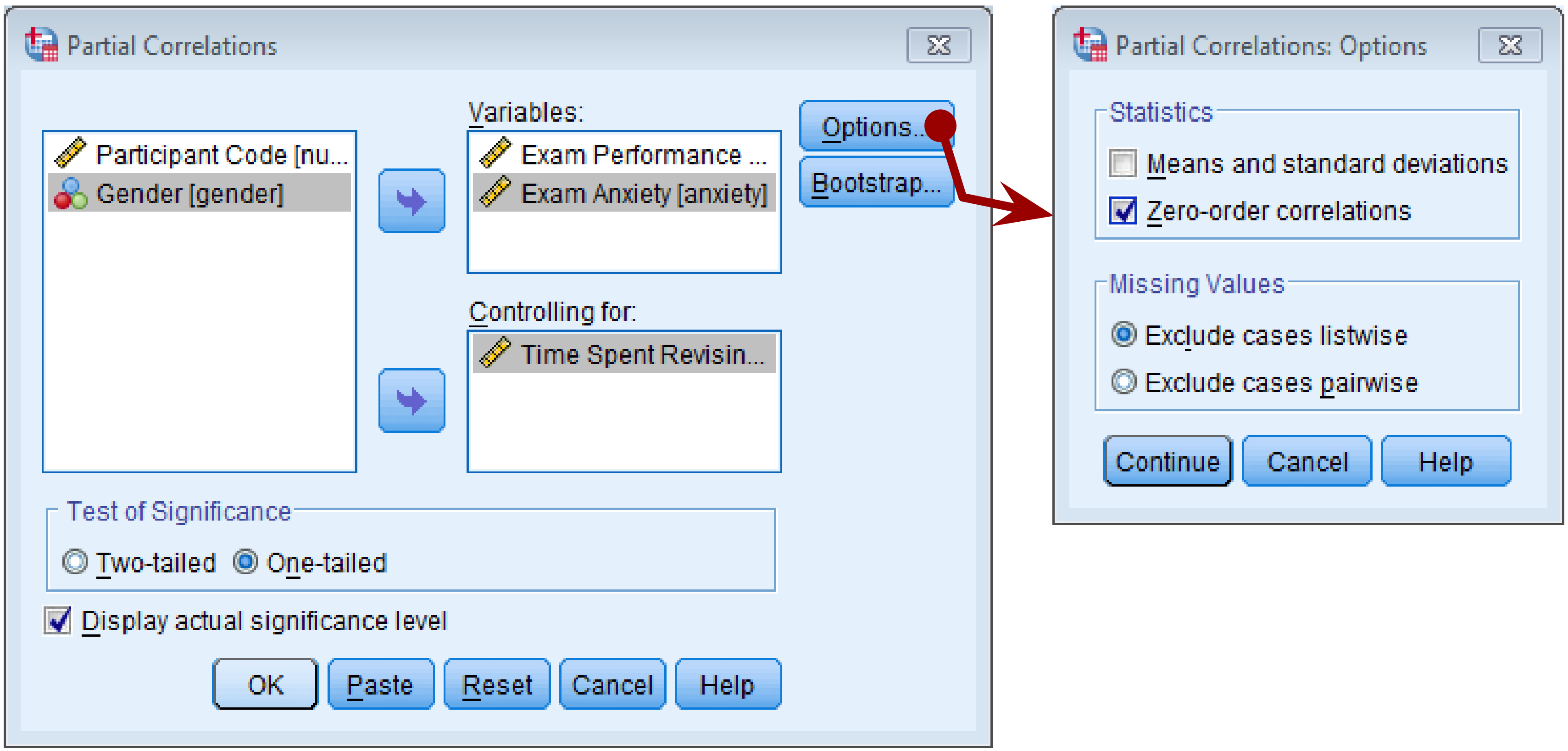


FIGURE 7.11
The difference between a partial and a semi-partial correlation

Doing Partial Correlation



Partial Correlation Output

Correlations

Control Variables					Exam Performance (%)	Exam Anxiety
Time Spent Revising	Exam Performance (%)	Correlation			1.000	-.247
		Significance (2-tailed)			.	.012
		df			0	100
		Bootstrap[a]	Bias		.000	.010
			Std. Error		.000	.102
			BCa 95% Confidence Interval	Lower	.	-.434
				Upper	.	-.005
	Exam Anxiety	Correlation			-.247	1.000
		Significance (2-tailed)			.012	.
		df			100	0
		Bootstrap[a]	Bias		.010	.000
			Std. Error		.102	.000
			BCa 95% Confidence Interval	Lower	-.434	.
				Upper	-.005	.

a. Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples