

Outline for Ch. 12 - Correlation

1

1. Overview of when to use correlation
2. What are correlations, conceptually?
 - null and alternative hypotheses
 - calculation of Pearson's r , the *correlation coefficient*
 - interpretation of r , including how NOT to interpret r
3. The coefficient of determination, R^2 (r^2)
- 4. Range restriction**
- 5. Outliers**
- 6. Correlation matrices**

Range restriction

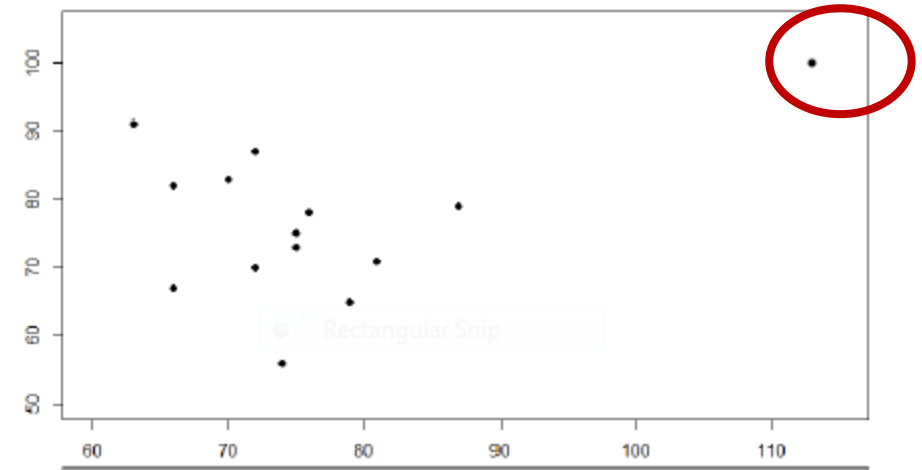
rho, the correlation in the population

- r may severely misrepresent ρ when your sample has a restricted range for one or both measured variables
- Ex. Suppose HS GPA and college GPA are correlated in the population at $\rho = .40$, and you collect data from 8 college students...

ID#	X (high school GPA)	Y (college GPA)
1	4.0	2.7
2	3.9	2.9
3	4.0	1.6
4	3.8	3.0
5	4.0	2.2
6	4.0	3.5
7	4.0	3.8
8	3.8	1.9

Outliers: scores very different from the rest of the data; extreme scores

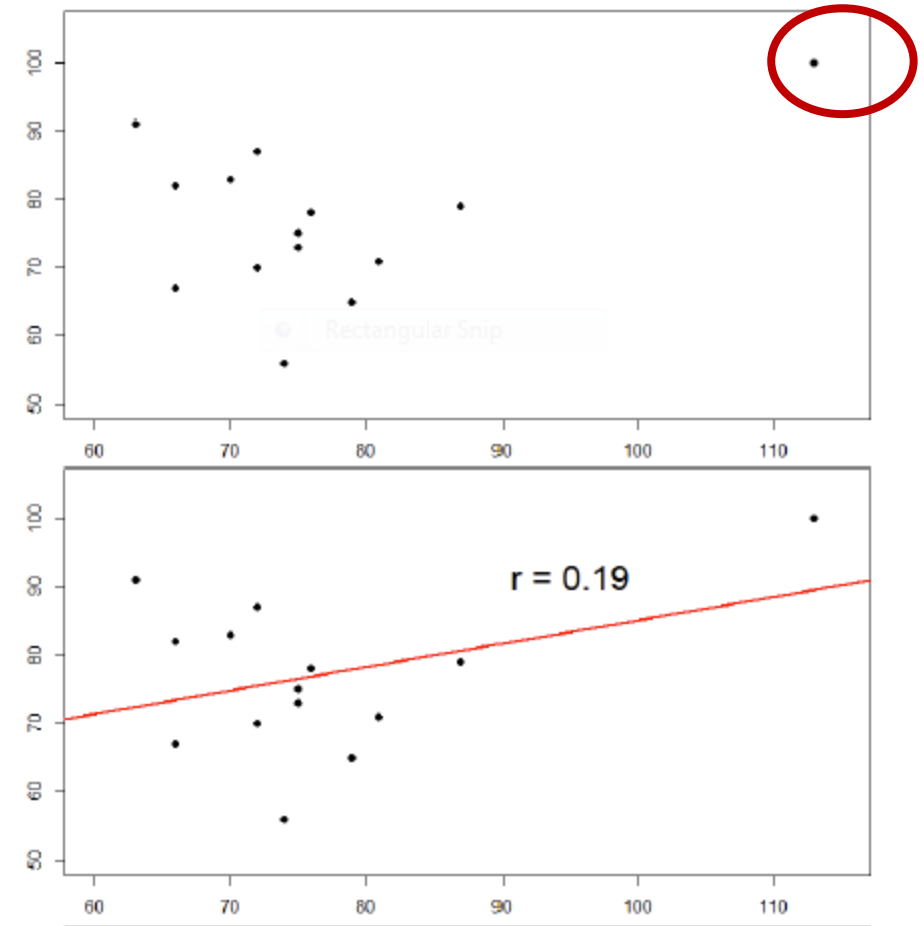
- Outliers can strongly influence results of correlational analyses, especially when sample is small.



Outliers: scores very different from the rest of the data

- Outliers can strongly influence results of correlational analyses, especially when sample is small.

Keeping outlier score in data set. Weak-to-moderate positive relationship.



Outliers: scores very different from the rest of the data

- Outliers can strongly influence results of correlational analyses, especially when sample is small.
- Okay to remove outlier scores if you provide justification in your research report.

Keeping outlier score in data set. Weak-to-moderate positive relationship.

Removing outlier from data set. Moderate negative relationship.

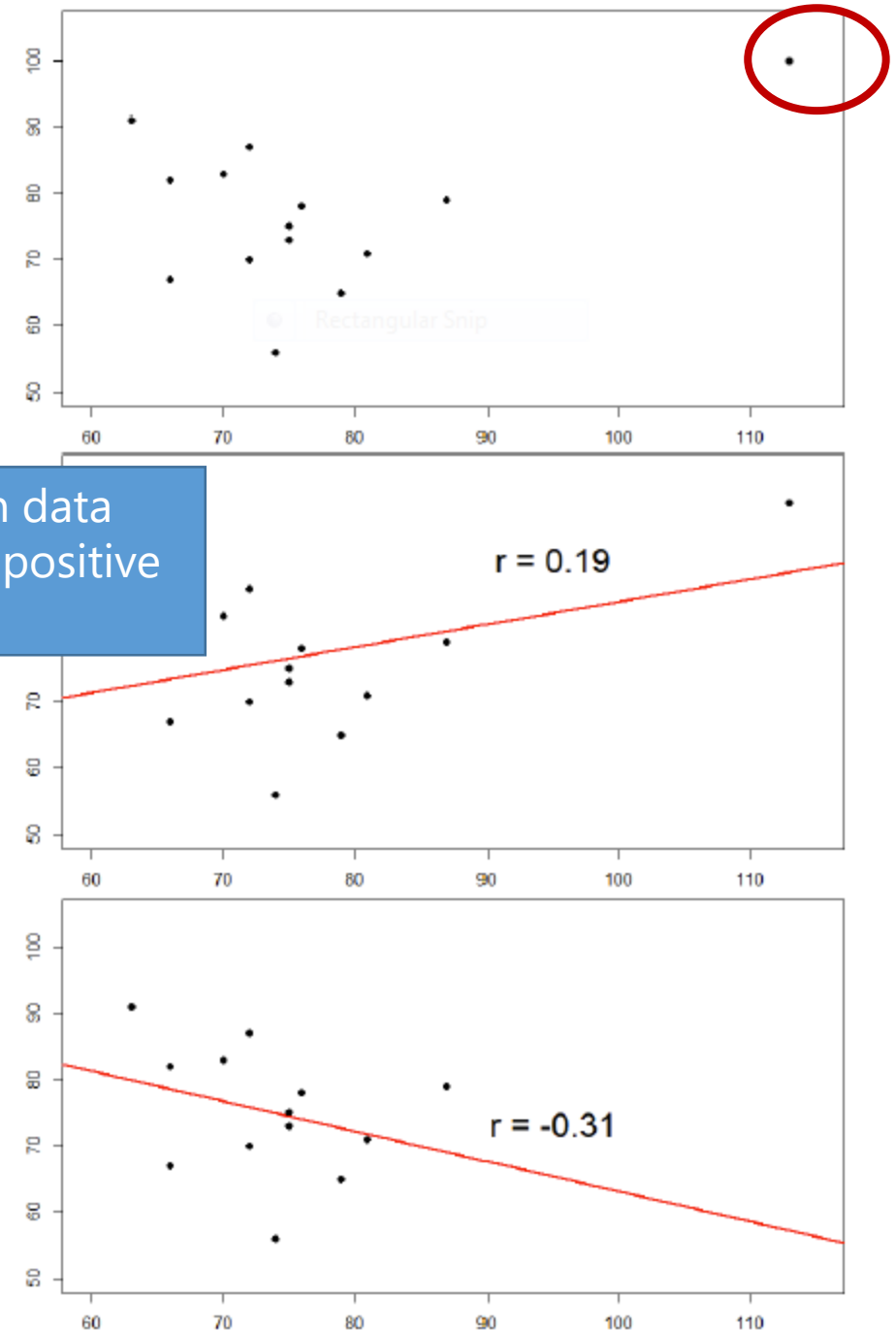


Figure 14. Three plots showing correlations with and without outliers.

Correlation matrices

- Correlations among multiple pairs of variables can be presented in a table called a ***correlation matrix***
 - Example from textbook
 - All of these values are values of r

	Satisfaction	Well-Being	Burnout	Performance
Satisfaction	1.00			
Well-Being	0.41	1.00		
Burnout	-0.54	-0.87	1.00	
Performance	0.08	0.21	-0.33	1.00

Correlation matrices – one more example

- Suppose a researcher measures three traits from the Big Five personality inventory . . .

Agreeableness:
how much you
are cooperative,
friendly, kind,
empathic

Conscientious-
ness: how much
you are
goal-directed,
hard-working,
organized

Neuroticism:
How moody
you are, how
much negative
emotion you
tend to feel

Correlation Matrix ▼

Pearson Correlations ▼

		Agreeableness	Conscientiousness	Neuroticism
Agreeableness	Pearson's r	—	0.159***	-0.134**
	p-value	—	< .001	0.003
Conscientiousness	Pearson's r	—	—	-0.368***
	p-value	—	—	< .001
Neuroticism	Pearson's r	—	—	—
	p-value	—	—	—

* $p < .05$, ** $p < .01$, *** $p < .001$

Practicing Correlation Concepts

A correlation of $r = -.53$ was found between number of hours slept last night and current sleepiness levels in a sample of 50 participants. Which of the following conclusions do you **know** to be **true** from this finding? *You may choose more than one if you find more than one that is correct.*

- a. Amount of time slept last night accounted for 53% of the variance in current sleepiness levels.
- b. Amount of current sleepiness accounted for 47% of the variance in hours slept last night.
- c. Spending more hours sleeping last night caused participants to have less current sleepiness.
- d. There was a negative relationship between the number of hours slept last night and current sleepiness.
- e. You could estimate that the effect size is large (there is a strong relationship).
- f. Given the information in the paragraph, you could infer that the correlation is statistically significant.
- g. The covariance between the two variables is $-.53$
- h. 28% of the variance in current sleepiness levels is shared by number of hours slept last night