

V37 - Simple Linear Regression - Part 2

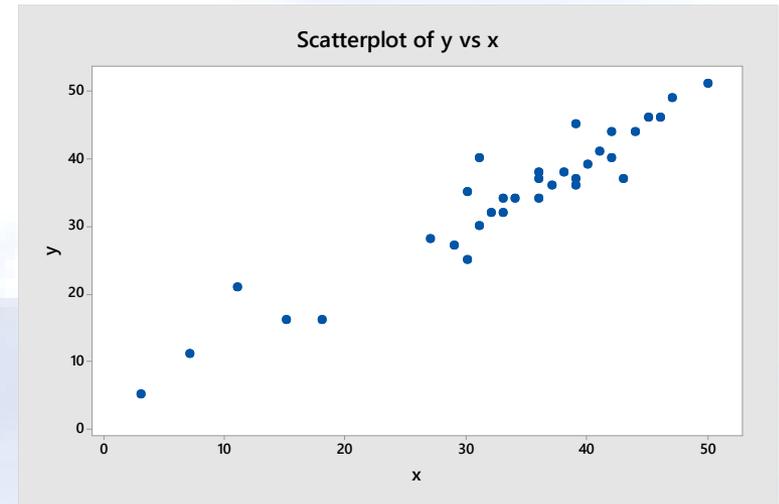
- ❑ Correlation
- ❑ Properties of correlation coefficient
- ❑ Impact of outliers on the correlation coefficient

Course: Statistical Testing & Regression
Dr. Renee Clark
Swanson School of Engineering
Industrial Engineering
University of Pittsburgh



Correlation

- ❑ Correlation measures strength of **linear relationship** between two variables
- ❑ **Plot** data **first** to assess reasonableness of linearity
- ❑ Sample correlation coefficient denoted by _____
- ❑ r aka “Pearson Correlation Coefficient”



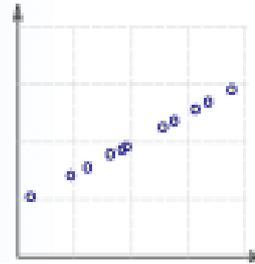
Properties of Correlation Coefficient r

$$-1 \leq r \leq 1$$

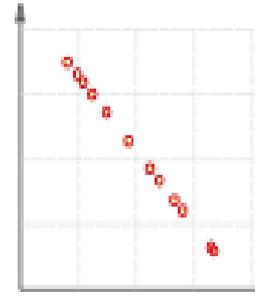


Properties of Correlation Coefficient r

If $r = 1 \rightarrow$ all (x, y) pairs lie on straight line (with positive slope)



If $r = -1 \rightarrow$ all (x, y) pairs lie on a straight line (with negative slope)



If $|r| = 1 \rightarrow$ *perfect linear relationship between x and y*

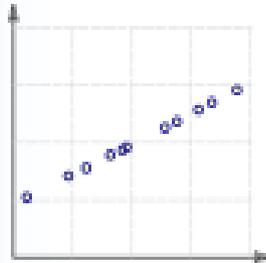
If $r = 0 \rightarrow$ *no correlation between variables*

If $r \approx 0 \rightarrow$ *little correlation between variables*

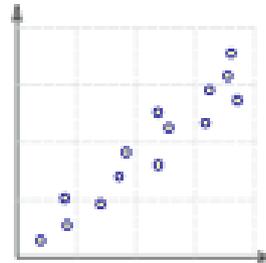


Correlation Cases

*Perfect
Positive
Correlation*



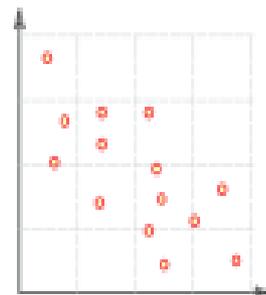
*High
Positive
Correlation*



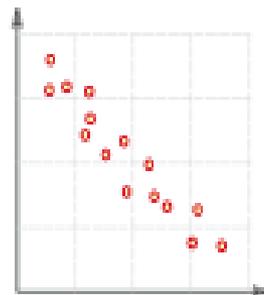
*Lower
Positive
Correlation*



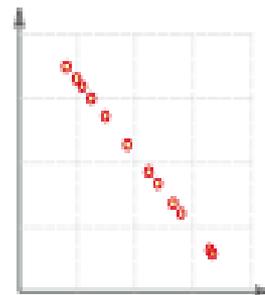
*Lower
Negative
Correlation*



*High
Negative
Correlation*

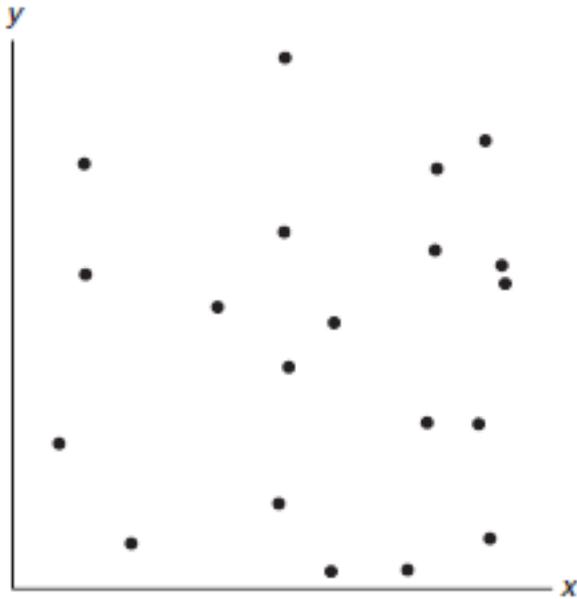


*Perfect
Negative
Correlation*

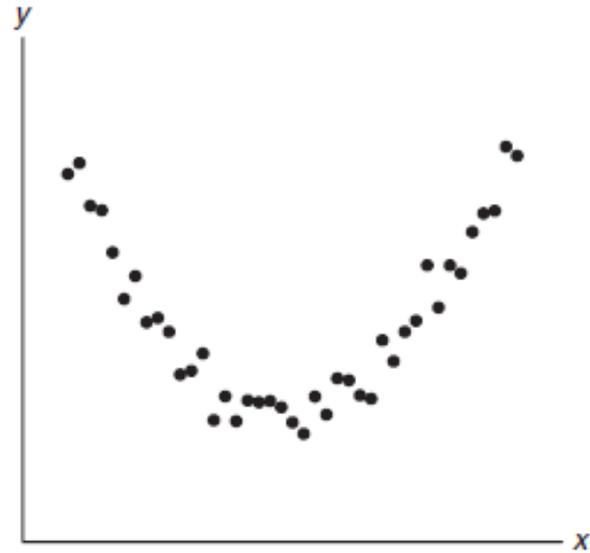


Zero Correlation Cases

*Plot
first!*



Random scatter of points; no association

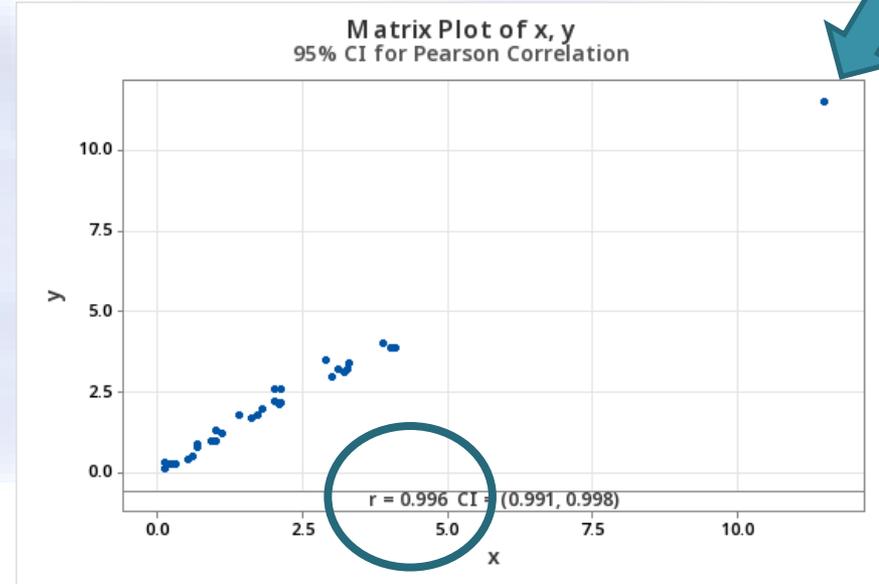
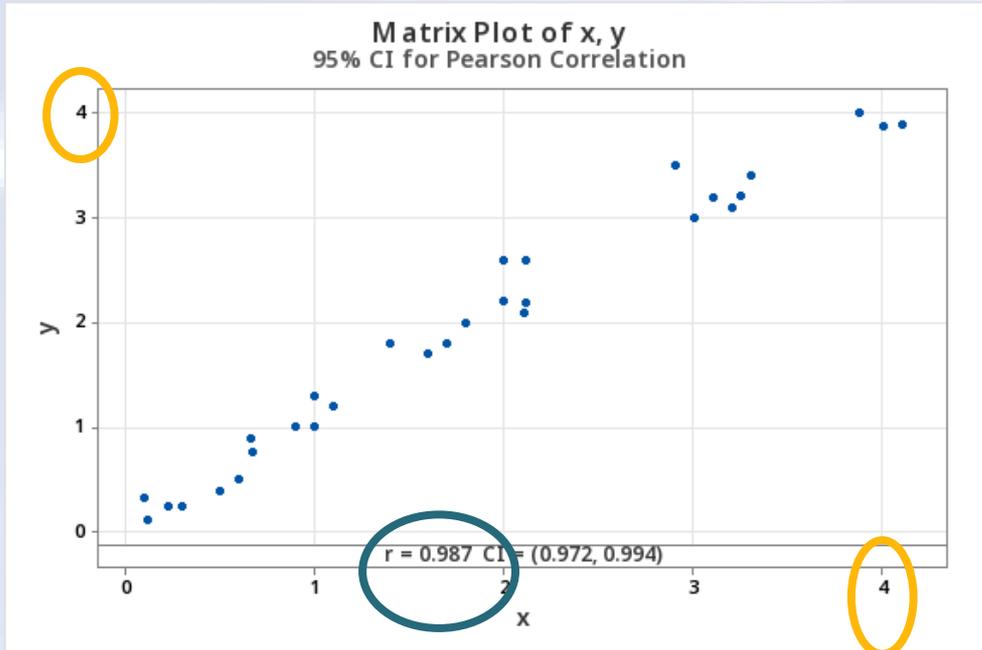


Quadratic relationship; lack of linear relationship



Correlation & Outliers

Same data
but with
outlier



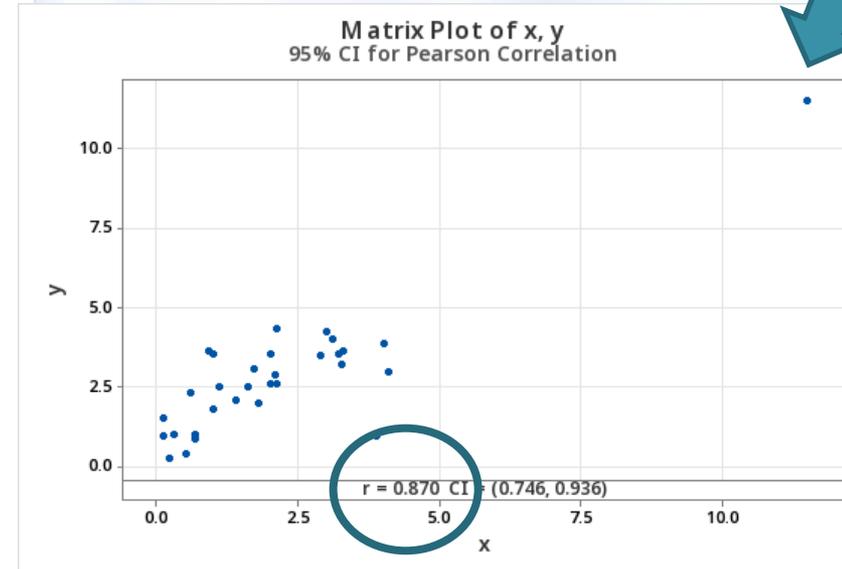
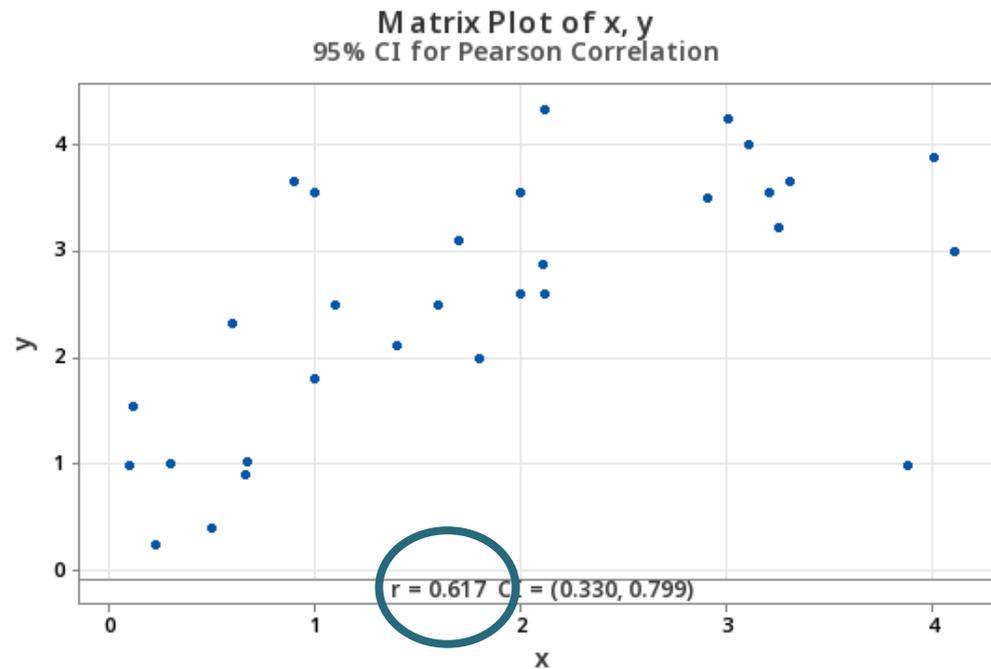
Little-to-no change in r .
Outlier had little influence on r .

Outlier: _____

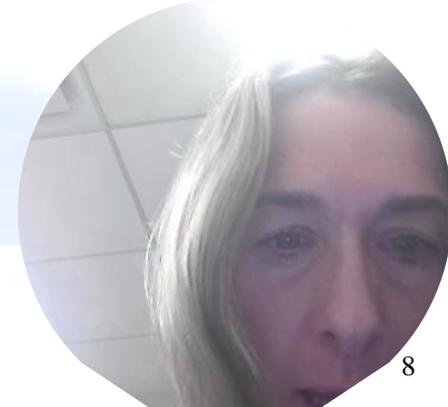


Correlation & Outliers

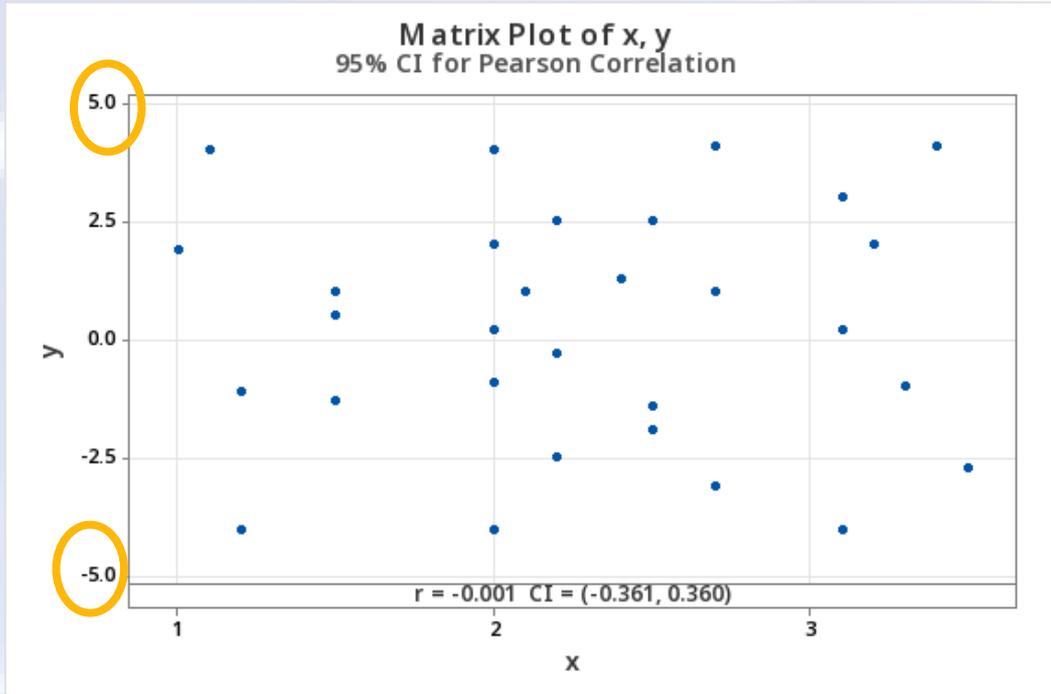
Same data
but with
outlier



Outlier was more influential. Increase in r
from _____ to _____



Correlation & Outliers



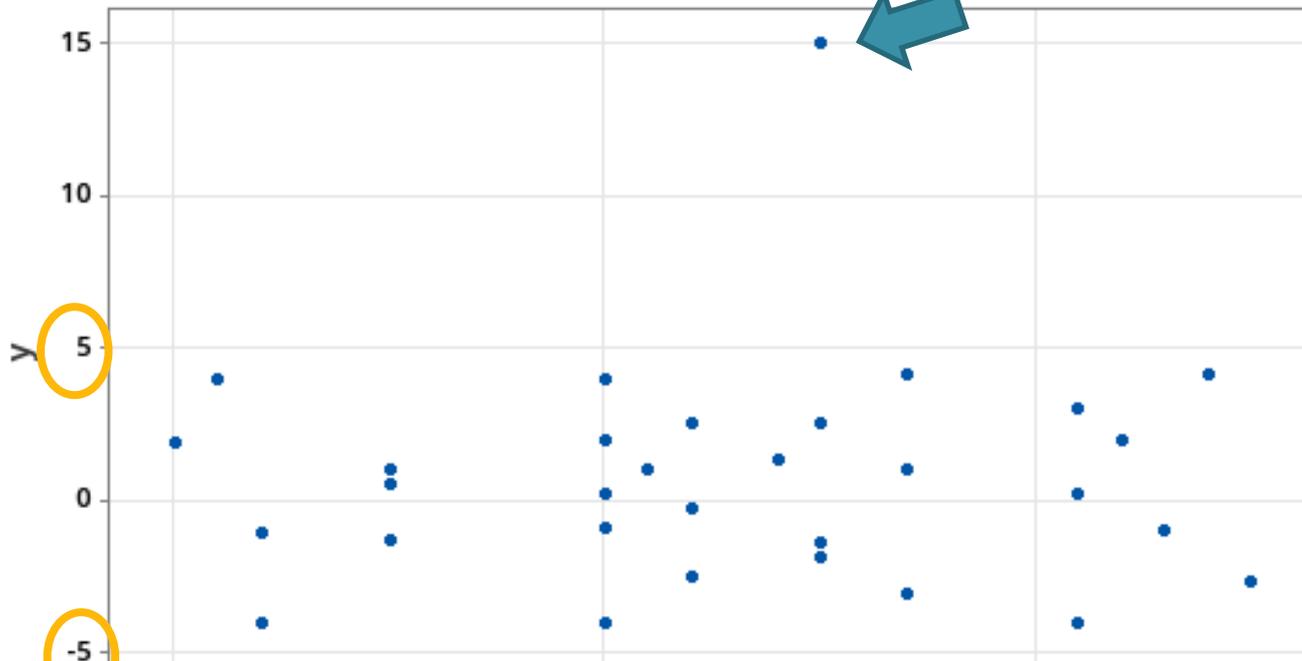
Correlations	
	x
y	-0.001



Correlation & Outliers

Same data
but with
outlier

Matrix Plot of x, y
95% CI for Pearson Correlation



Correlations

	x
y	0.04

Outlier very influential –
increase in r by a factor of
40

Correlations

	x
y	-0.001





Acknowledgement

This material is based upon work partially supported by the National Science Foundation under Grant# 2335802. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.





You are free

- to **Share** – to copy, distribute, display and perform the work
- to **Remix** – to make derivative works

Under the following conditions

- **Attribution** — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- **Noncommercial** — You may not use this work for commercial purposes.
- **Share Alike** — If you alter, transform, or build upon work, you may distribute the resulting work only under the same or similar license to this one.





THE END

