



Bookmarks

- ▶ Important Pre-Course Survey
- ▶ Contact Us
- ▶ How To Navigate the Course
- ▶ Discussion Board
- ▶ Office Hours
- ▶ Week 1: Introduction to Data
- ▶ Week 2: Univariate Descriptive Statistics

▼ **Week 3:**
Bivariate Distributions

Readings

Reading Check due
Mar 15, 2016 at 18:00
UTC

Lecture Videos

Comprehension Check
due Mar 15, 2016 at
18:00 UTC

R Tutorial Videos**Pre-Lab**

Pre-Lab due Mar 15,
2016 at 18:00 UTC

Lab

Week 3: Bivariate Distributions > Lecture Videos > Linearity and the Correlation Coefficient



Bookmark

Linearity and the Correlation Coefficient



SPEAKER: MICHAEL J.
MAHOMETA, Ph.D.

One of the main things
that we'll hear or read

is that the Pearson
correlation coefficient -
the r value - is only

appropriate to use
when there is a LINEAR
relationship to our
data.

▶ 0:00 / 4:55

▶ 1.0x



Download transcript

.srt

Comprehension Check

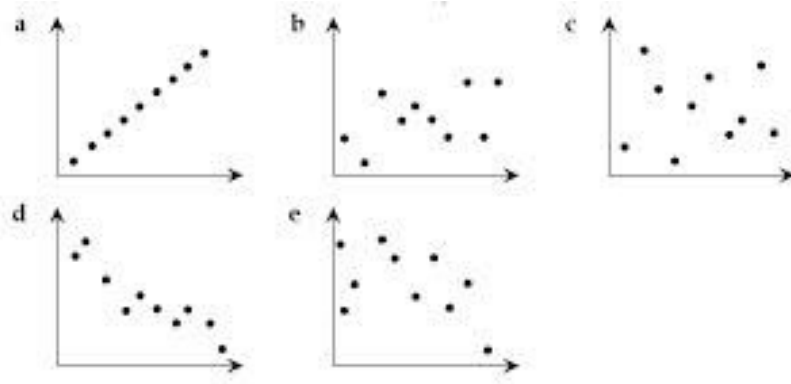
**Below are several scatterplots depicting relationships
between bivariate data:**

Lab due Mar 15, 2016
at 18:00 UTC

Problem Set

Problem Set due Mar
15, 2016 at 18:00 UTC

► Week 4:
Bivariate
Distributions
(Categorical
Data)



(5/5 points)

Please match each description with the most appropriate graph above. The five different descriptions are each most appropriate to a different graph; therefore, answers should not repeat.

Perfect, positive linear relationship

Graph A ▼



Answer: Graph A

Non-linear relationship that should not be measured with a correlation coefficient

Graph E ▼



Answer: Graph E

Strong negative relationship

Graph D ▼



Answer: Graph D

$r = 0.35$

Graph B ▼



Answer: Graph B

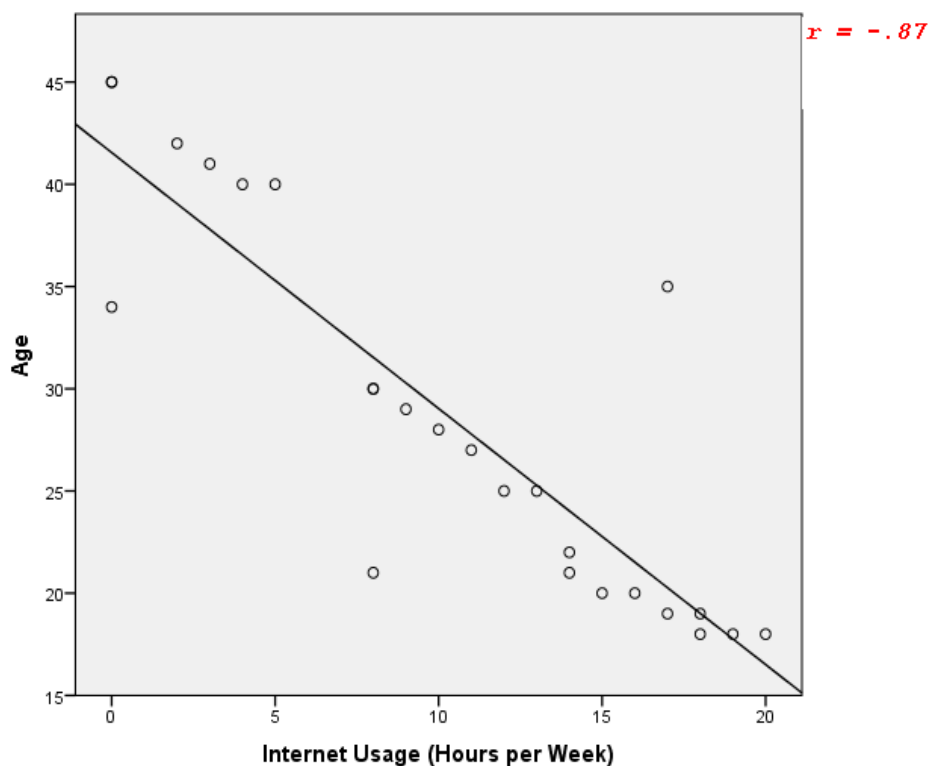
No relationship

Graph C ▼



Answer: Graph C

Is there a relationship between age and internet usage each week? Here is bivariate data collected to examine this question:



(3/3 points)

What makes the data in this scatterplot "bivariate data"?

- ☐ One variable is quantitative, and one variable is qualitative.
- ☐ The data is collected on two different populations of people: those that use the internet, and those that don't.
- ☒ For each subject, we know both their age and how often they use the internet each week. ✓
- ☐ There are multiple data points being graphed using the same two axes.

There are two ~35-year-olds in this dataset. One uses the internet not at all, and the other uses it more than 15 hours per week. Which individual is the stronger outlier, and why?

- ☐ The 35-year-old that doesn't use the internet at all, because very few people did not use the internet.
- ☒ The 35-year-old that uses the internet 15+ hours per week. His

data point is farther away from the linear trend. ✓

- ☐ They are both equally outliers because their data points are both at the same height on the graph.

Another researcher was only interested in individuals that use the internet 18+ hours per week. He calculated the correlation coefficient from the same dataset and got $r = 0.02$, showing no relationship. What happened?

- ☒ He created a restriction of range that made it look like there was no relationship. ✓

- ☐ He must have added or subtracted incorrectly; the same data should produce the same correlation coefficient.

- ☐ He should have selected the people that used the internet less than 18 hours per week.

- ☐ The strong non-linear relationship in that part of the graph cannot be described by r .

© All Rights Reserved



© edX Inc. All rights reserved except where noted. EdX, Open edX and the edX and Open EdX logos are registered trademarks or trademarks of edX Inc.

POWERED BY
OPENedX

