INTR 5057 Research Design & Methods

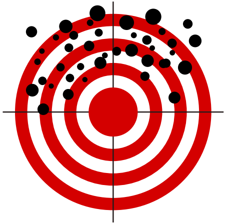
Juraj Medzihorsky

Day 10, 2016-11-25

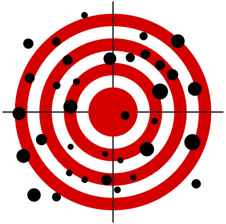
Homework #2

- Any questions?

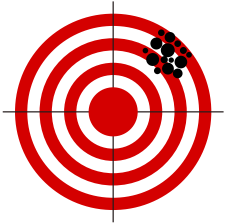
Reliability & Validity



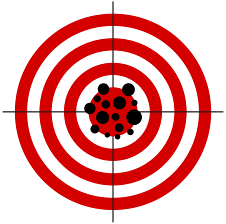
Unreliable & Invalid



Unreliable, But Valid

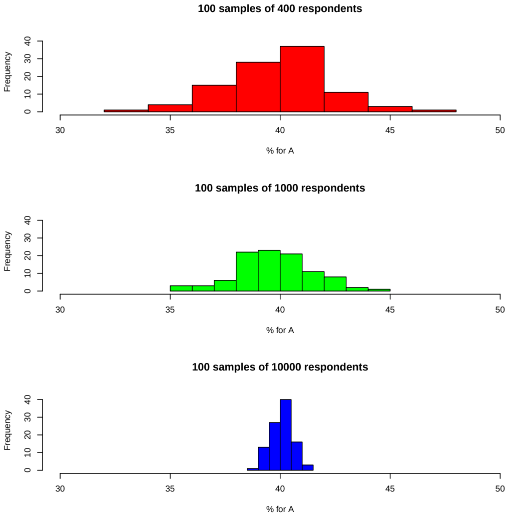


Reliable, Not Valid



Both Reliable & Valid

Random Sampling



100 samples of 400 respondents

100 samples of 1000 respondents

100 samples of 10000 respondents

Independence

Lack of association.

Independence & Probability

- Suppose X and Y that are *binary*, i.e. 0 or 1.
- Suppose we know that X is independent of Y.
- Probability of X is the same if $Y = 1$ and $Y = 0$.

$$P(X|\mathrm{Y}) = P(X|\text{not Y})$$

Independence & Drawing

- When drawing **with** replacement the draws are **independent**
- When drawing **without** replacement the draws are **dependent**

Ex.

- 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- At least how many silver coins are there?

Ex.

- 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- At least how many silver coins are there?
- 0. Why?

Ex.

- 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- At least how many silver coins are there?
- 0. Why?
- Both coins can be golden.

Ex.

- 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- At most how many silver coins are there?

Ex.

- 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- At most how many silver coins are there?
- 2, Why?

Ex.

- 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- At most how many silver coins are there?
- 2, Why?
- Both coins can be silver.

Ex.

- 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- We draw 1 item at random. What is the chance it will be a coin?

Ex.

- 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- We draw 1 item at random. What is the chance it will be a coin?
- 50%

Ex.

- 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- We draw 1 item at random. What is the chance it will be golden?

Ex.

- 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- We draw 1 item at random. What is the chance it will be golden?
- 50%

Ex.

- There are 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- We draw 1 item at random. What is the chance it will be a golden coin?

Ex.

- There are 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- We draw 1 item at random. What is the chance it will be a golden coin?
- That depends on whether material and shape are independent.

Ex.

- There are 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- How would the contents of the box look if material and shape were independent?

Ex.

- There are 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- How would the contents of the box look if material and shape were independent?
- 1 golden coin, 1 silver coin, 1 golden cube, 1 silver cube.

Ex.

- There are 4 items in a box. 2 are coins and 2 are cubes. 2 are silver and 2 are gold.
- How would the contents of the box look if material and shape were independent?
- 1 golden coin, 1 silver coin, 1 golden cube, 1 silver cube.
- Can you write this as a table?

Ex.

- There are 8 items in a box. 4 are coins and 4 are cubes. 4 are silver and 4 are gold. Material and shape are independent.
- We draw 1 item at random. What is the chance it will be a golden coin?

Ex.

- There are 8 items in a box. 4 are coins and 4 are cubes. 4 are silver and 4 are gold. Material and shape are independent.
- We draw 1 item at random. What is the chance it will be a golden coin?

  $P(\text{golden coin}) = P(\text{golden}) \times P(\text{coin})$

  $P(\text{golden coin}) = \frac{4}{8} \times \frac{4}{8} = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$

Multiplication Rule

- If 2 events are independent their joint probability is equal to the product of their probabilities.
- I.e. if

$$P(\mathrm{A}|\mathrm{B}) = P(\mathrm{A}|\text{not B})$$

then

$$P(\mathrm{A}\ \&\ \mathrm{B}) = P(\mathrm{A}) \times P(\mathrm{B})$$

Ex.

- 7 items in a box. 3 are coins and 4 are cubes. 5 are silver and 2 are gold. Material and shape are independent.
- We draw 1 item at random. What is the chance it will be a golden coin?

Ex.

- 7 items in a box. 3 are coins and 4 are cubes. 5 are silver and 2 are gold. Material and shape are independent.
- We draw 1 item at random. What is the chance it will be a golden coin?

  $P(\text{golden coin}) = P(\text{golden}) \times P(\text{coin})$

  $P(\text{golden coin}) = \frac{2}{7} \times \frac{3}{7} = \frac{6}{49}$

Ex.

- 4 items in a box. 2 are coins and 2 are cubes.
- We draw 1 item at random with replacement. Then we draw another item at random with replacement.
- What is the chance that at least one of them will be a coin?

Ex.

- 4 items in a box. 2 are coins and 2 are cubes.
- We draw 1 item at random with replacement. Then we draw another item at random with replacement.
- What is the chance that at least one of them will be a coin?
- The opposite of never getting a coin.

Ex.

- 4 items in a box. 2 are coins and 2 are cubes.
- We draw 1 item at random with replacement. Then we draw another item at random with replacement.
- What is the chance that at least one of them will be a coin?
- The opposite of never getting a coin.
- $1 - (\frac{2}{4} \times \frac{2}{4}) = 1 - \frac{1}{4} = \frac{3}{4}$

Probability of ‘At least 1 in N repetitions’

- Probability of getting a certain outcome at least once in several repetitions.
- It is the “opposite” of never getting the outcome.
- 1 minus the probability of never getting the outcome.

Ex.

- 6 items in a box. 3 are coins and 3 are cubes.
- We draw 1 item at random with replacement. Then we draw another item at random with replacement.
- What is the chance that at least of them will be a coin?

Ex.

- 6 items in a box. 3 are coins and 3 are cubes.
- We draw 1 item at random with replacement. Then we draw another item at random with replacement.
- What is the chance that at least of them will be a coin?
- $1 - \frac{3}{6} \times \frac{3}{6} = 1 - \frac{9}{36} = 1 - \frac{1}{4} = \frac{3}{4}$

Ex.

- 4 items in a box. 1 is a coin, 1 is a cube and 2 are sticks.
- We draw 1 item at random.
- What is the chance that it is a coin or a cube?

Ex.

- 4 items in a box. 1 is a coin, 1 is a cube and 2 are sticks.
- We draw 1 item at random.
- What is the chance that it is a coin or a cube?
- $P(Cube) + P(Coin) = \frac{1}{4} + \frac{1}{4} = \frac{2}{4} = \frac{1}{2}$

Addition Rule

- Two events are **mutually exclusive** if the occurence of one prevents the occurence of the other one.
- If two events are **mutually exclusive** we can calculate the probability that at least one of them happens by adding up their probabilities.

Peer Review Example

- What is peer review in scientific journals?

Peer Review Example

- What is peer review in scientific journals?
- A journal receives 100 submissions. 80 of them are bad, the rest is good.
- The chance of a bad one getting published is 10%.
- The chance of a good one getting published is 90%.
- How many good and bad articles will get published?

Summarizing a Single Variable

How to summarize the following information?

x
1
4
0
3
3
2

Summarizing a Single Variable

- Average (mean)
- Mode
- Median
- Midrange

Average (Mean)

Average (Mean)

- Sum of all values divided by the number of values.

Mode

Mode

- The most common value.

Median

Median

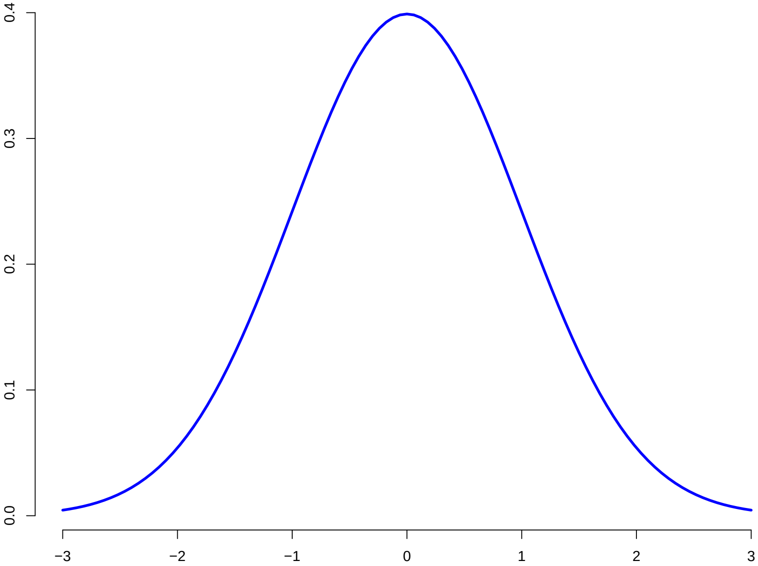
- Half of the observations have less, half more.

Midrange

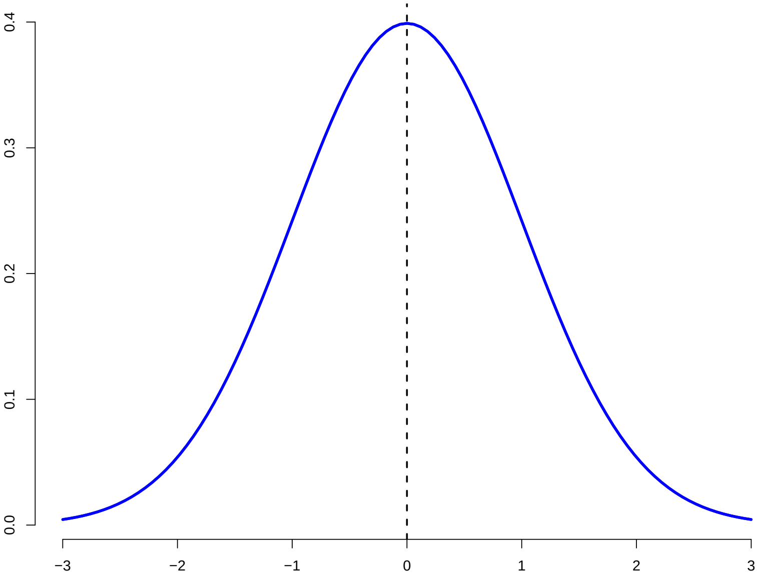
Midrange

- Middle between maximal value and minimal value.

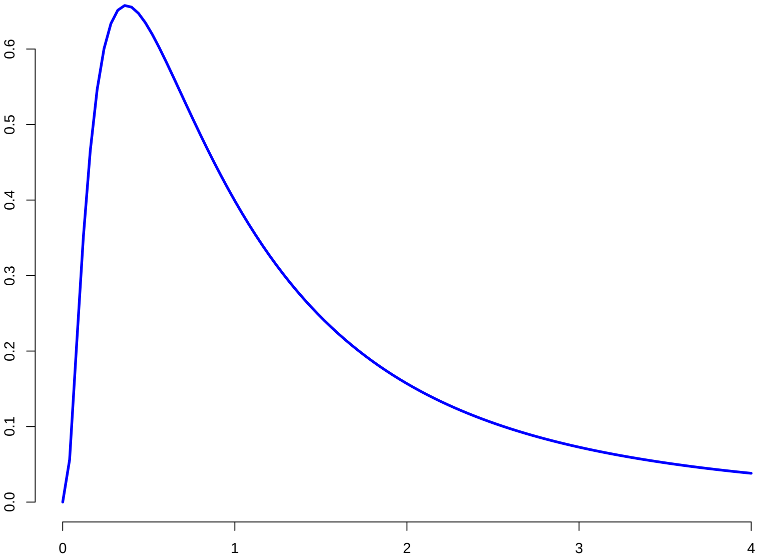
A Continuous Variable



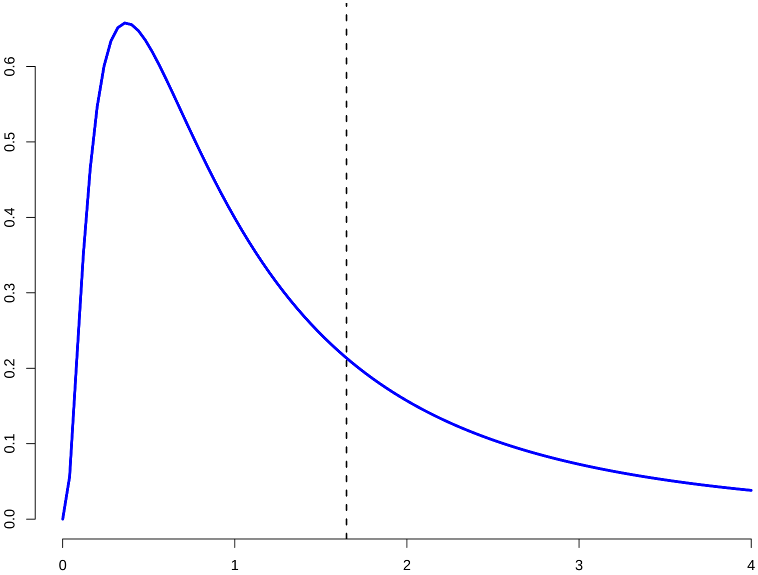
Mean, Median, Mode, Midrange



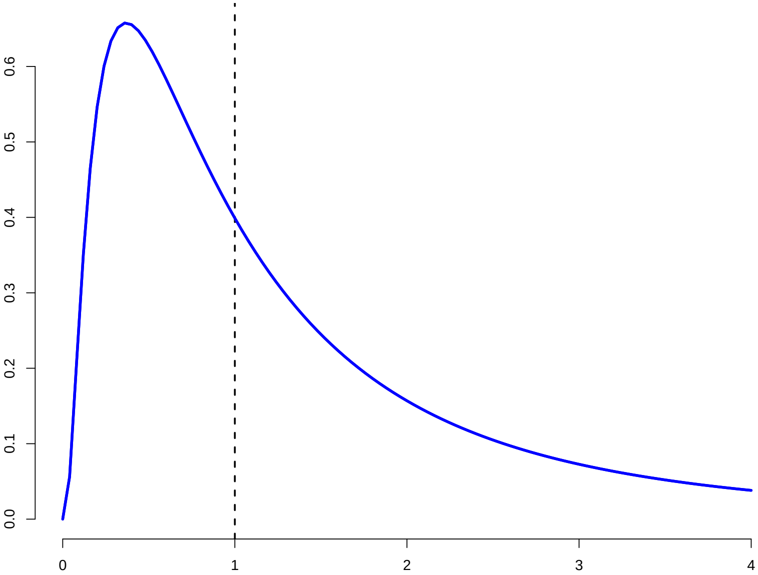
A Continuous Variable



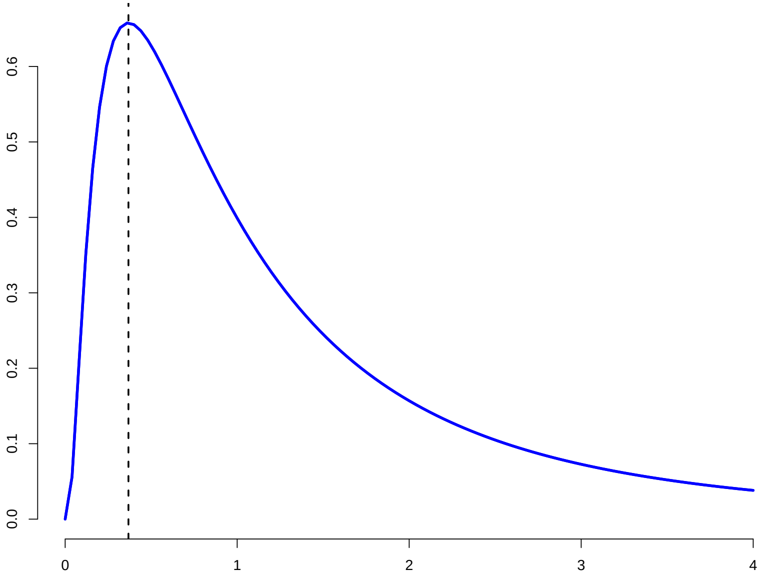
Mean



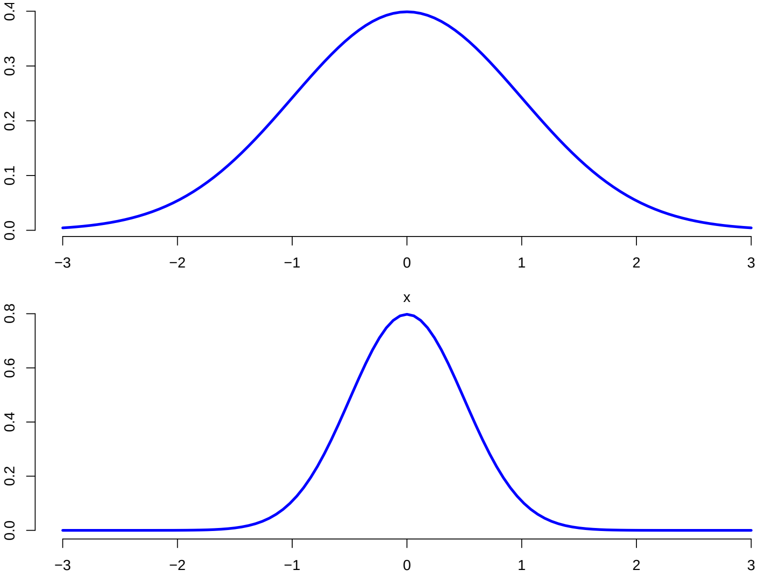
Median



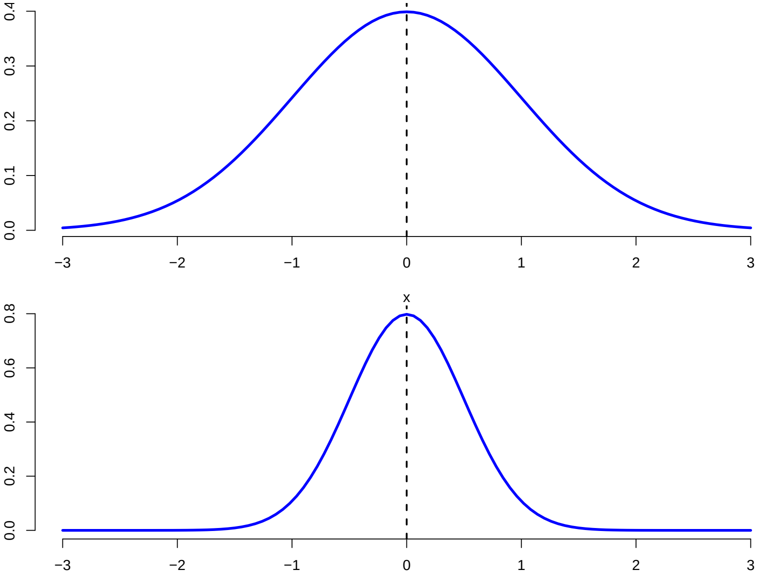
Mode



Two Continuous Variables



Same Mean



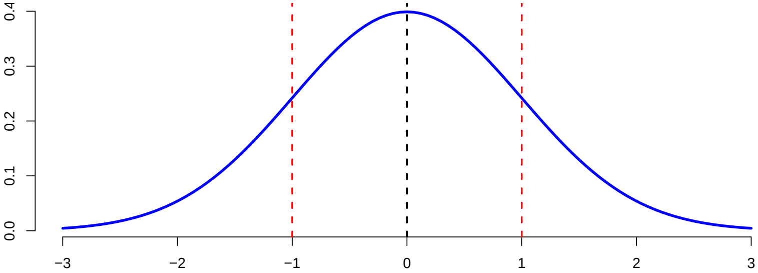
Standard Deviation

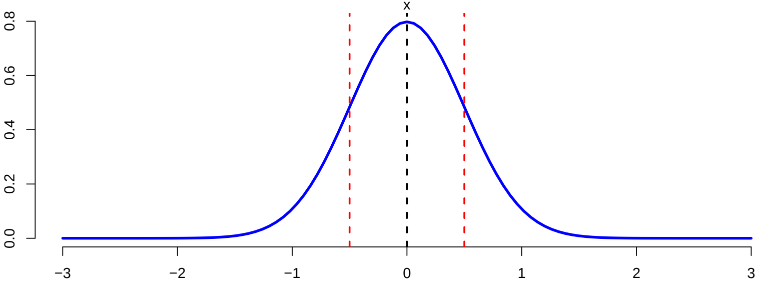
A measure of dispersion from the average.

Square root of the average squared distance from the average.

$$\sigma = \sqrt{\frac{\Sigma(\mu - x_i)^2}{N}}$$

Standard Deviation





Summarizing Two Variables

How to summarize the following information?

x	y
1	1
0	1
1	0
1	0
0	0
1	0

Summarizing Two Variables

- Summarize each of them separately.
- Capture information about their **association.**

Cross Table

		y	
		0	1
x	1	3	1
	0	1	1

Cross Table

		y	
		0	1
x	1	a	b
	0	c	d

Cross Product Ratio

A measure of association between two binary variables also know as odds ratio.

If cross product ratio $= 1$ then the variables are independent.

$$cpr = \frac{a \times d}{b \times c}$$

Cross Sum Ratio

An alternative to the cross product ratio.

$$csr = \frac{a + d}{b + c}$$

Relative Risk Ratio

A measure of association between two binary variables.

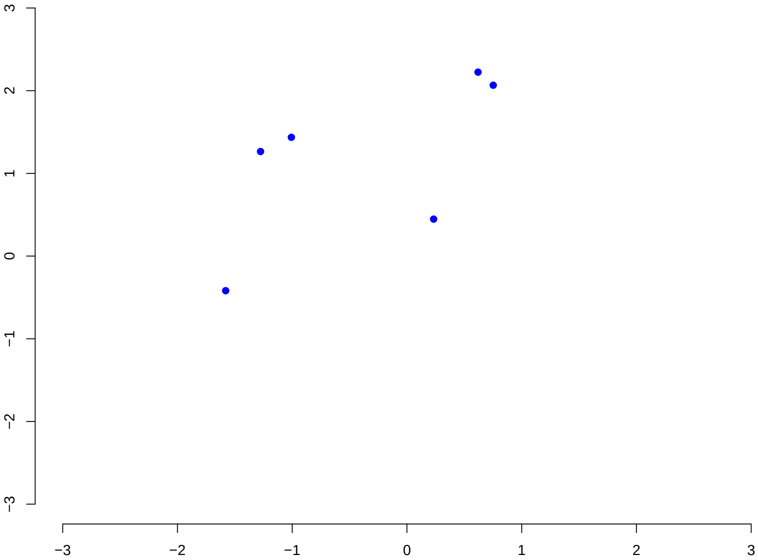
$$rr = \frac{\frac{a}{a+b}}{\frac{c}{c+d}}$$

Summarizing Two Variables

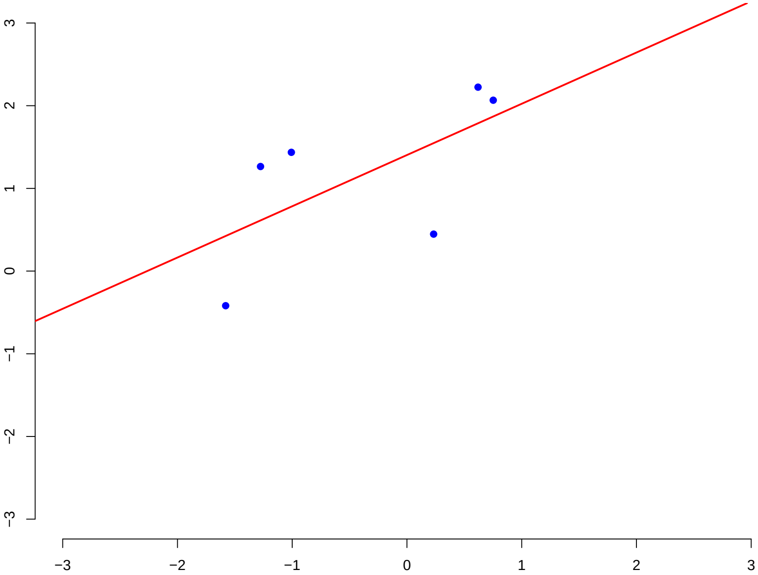
How to summarize the following information?

y	x
0.8	2.1
0.6	2.2
-1.6	-0.4
0.2	0.5
-1.3	1.3
-1.0	1.4

Scatter Plot



Correlation



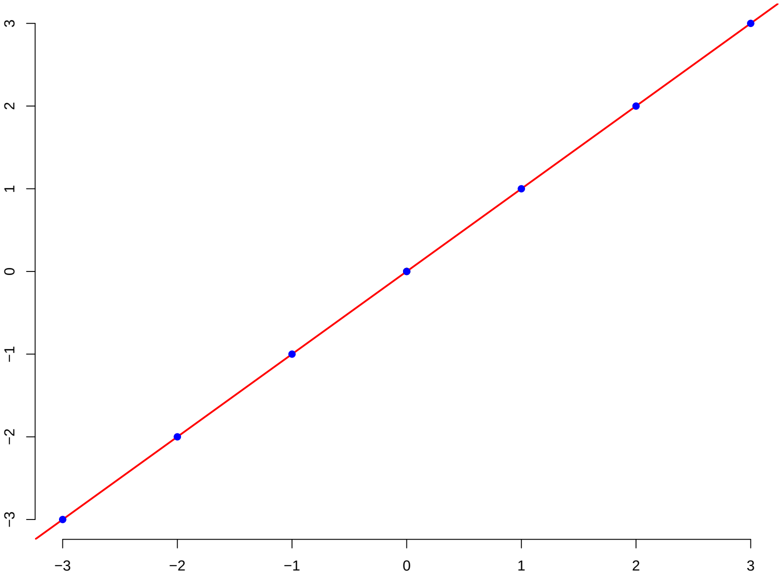
Correlation

A measure of association between two continuous variables.

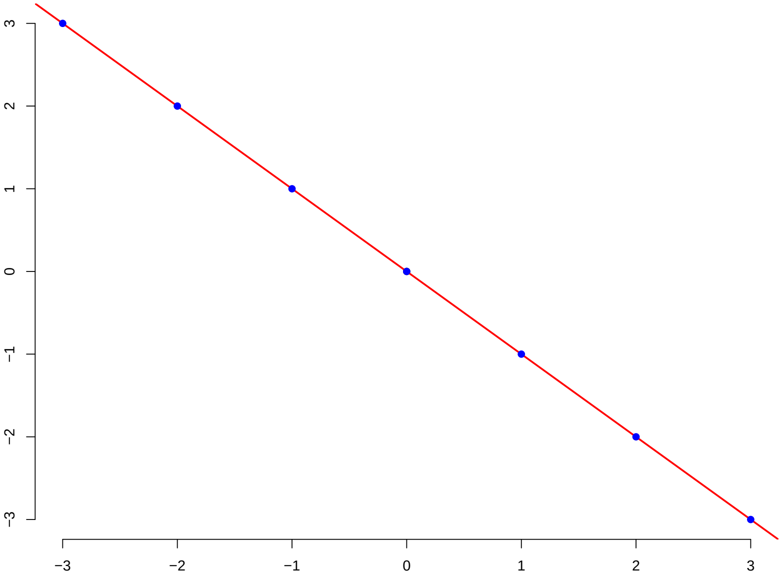
$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y}$$

Ranges from -1 to 1 on a closed interval.

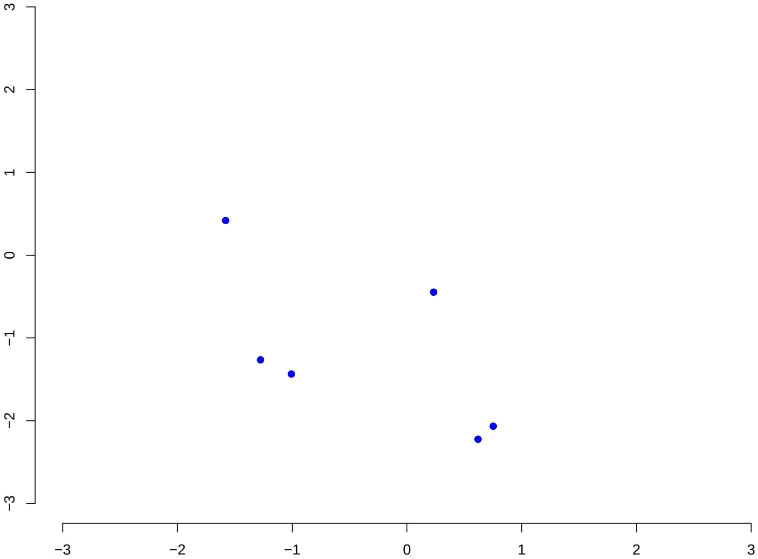
$\rho = 1$



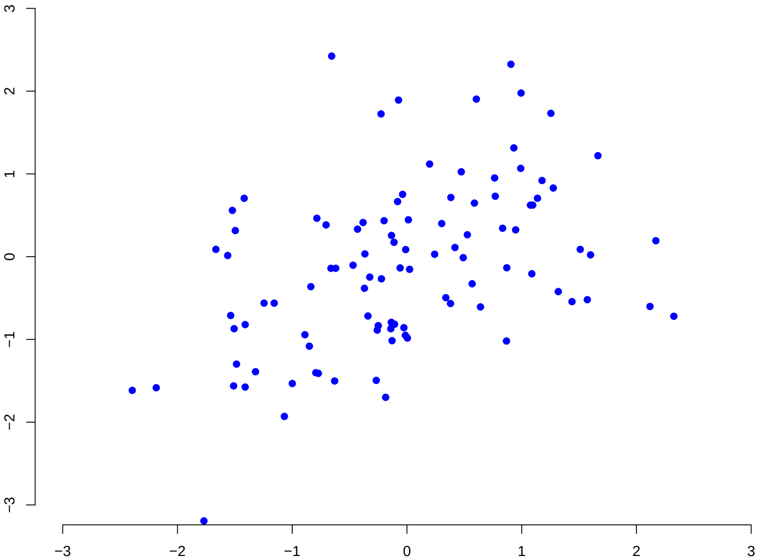
$$\rho = -1$$



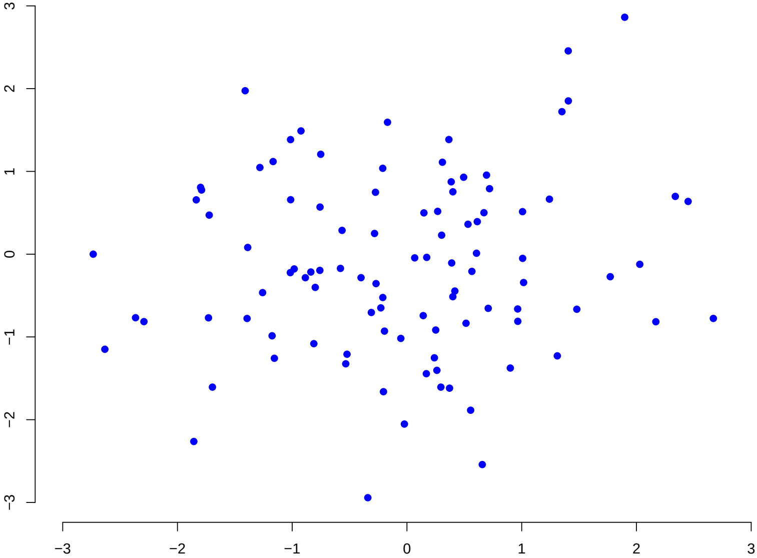
$$\rho = -0.63$$



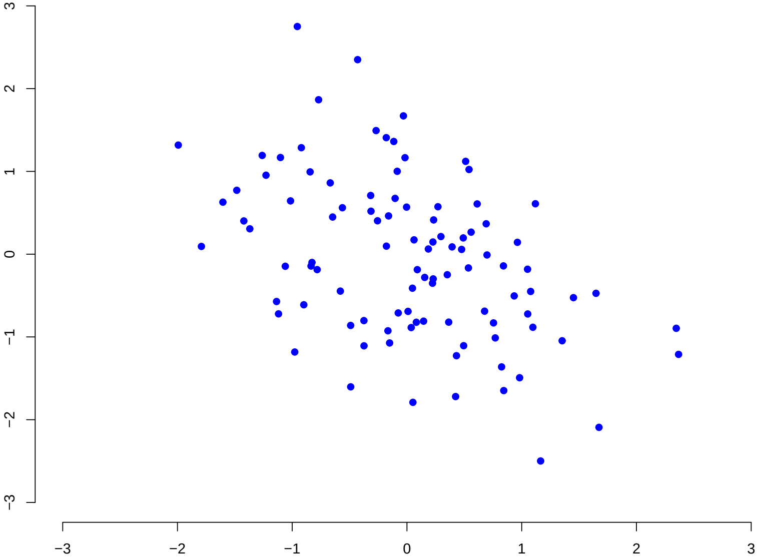
$$\rho = 0.44$$



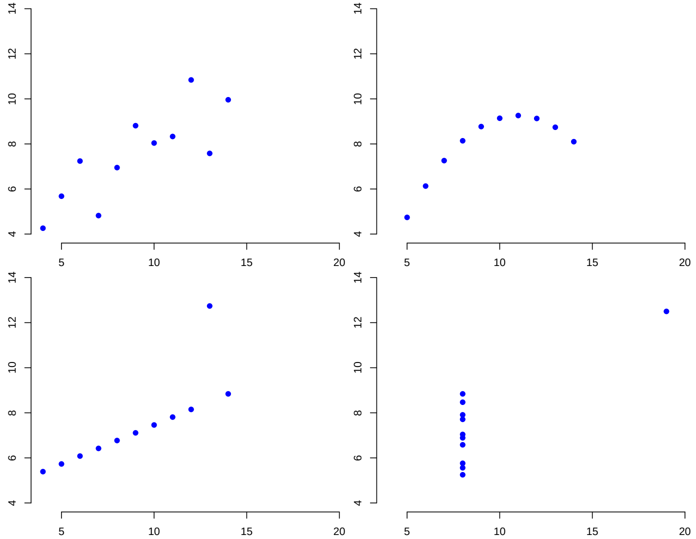
$\rho = 0.09$



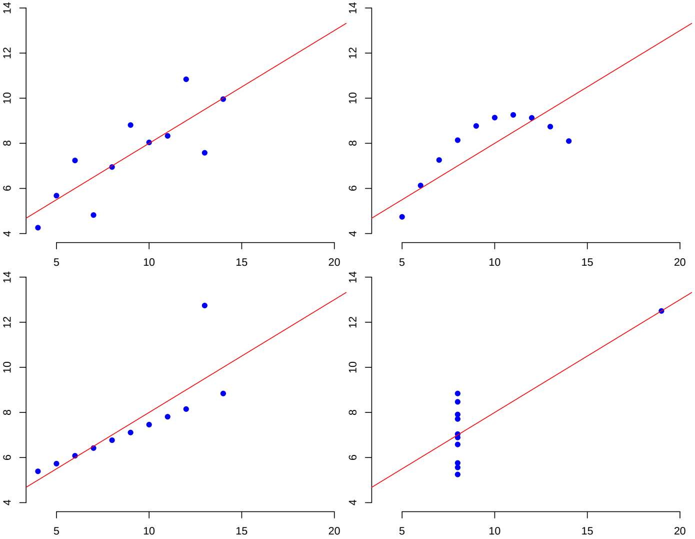
$\rho = -0.46$



Anscombe's Quartet



$$\rho = 0.82$$



14
12
10
8
6
4
5
10
15
20
14
12
10
8
6
4
5
10
15
20
14
12
10
8
6
4
5
10
15
20
14
12
10
8
6
4
5
10
15
20

Simpson's Paradox

The whole sample:

	heal	didn't
drug	20	20
no drug	16	24

Simpson's Paradox

Females:

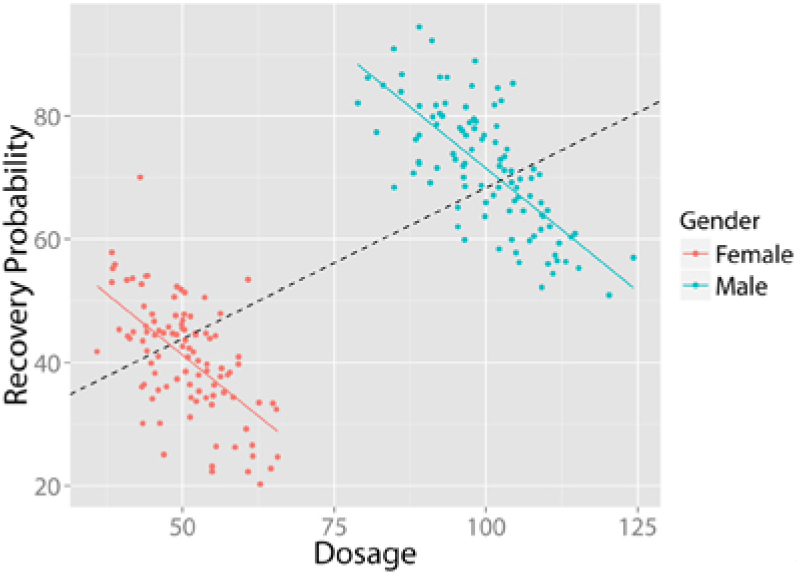
	heal	didn't
drug	2	8
no drug	9	21

Simpson's Paradox

Males:

	heal	didn't
drug	18	12
no drug	7	3

Simpson's Paradox



Simpson's Paradox

- In the whole population association in one direction.
- In subsets of the population association in the opposite direction.
- Not really a paradox when you think about it.
- A serious problem is that people rush ahead with causal interpretations.

Association & Causality

- Non-statisticians say *“correlation does not imply causation.”*
- Statisticians say *“association does not imply causation.”*
- Calling all association “correlation” is like calling all motor vehicles “cars.”

Homework #2

- Any questions?