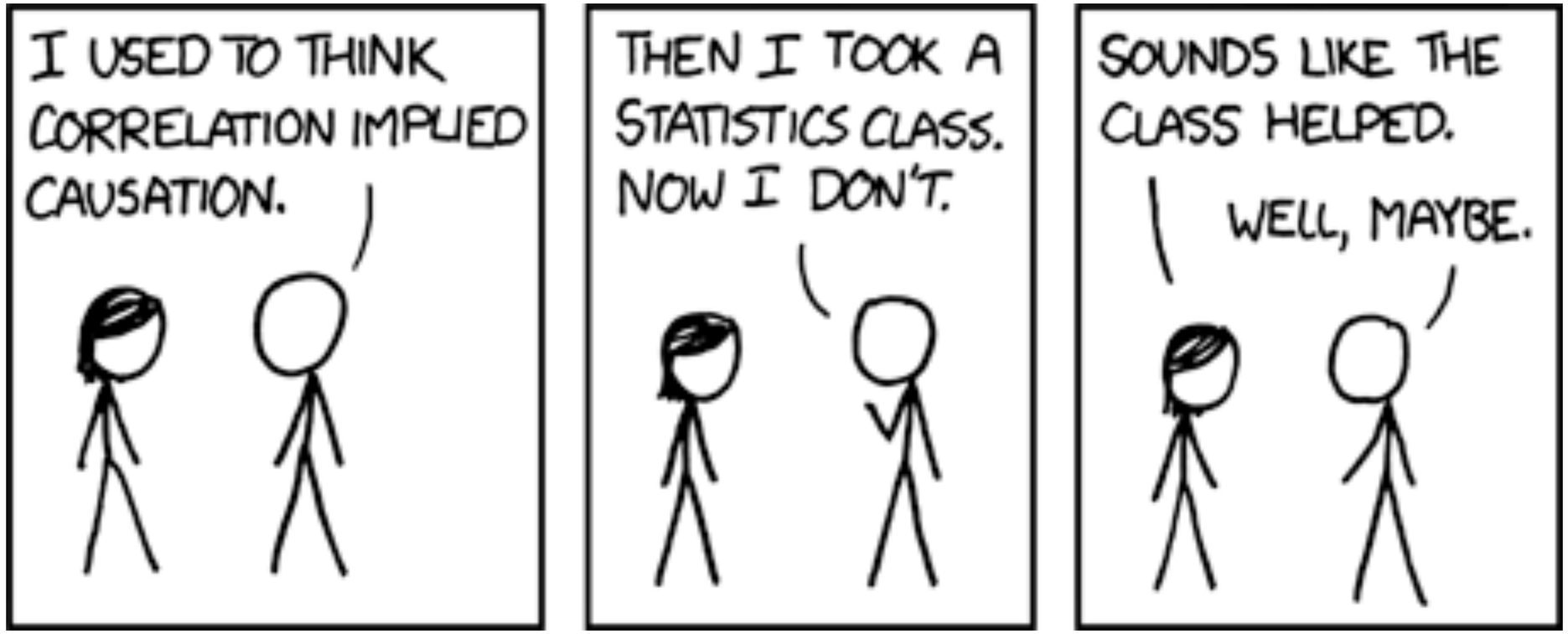


Lecture 12

11.11.21

You are running a case-control GWAS for Type 2 Diabetes. Of the 500,000 variants you test, one variant (rs4514, which has 2 alleles, A and G) near the *sweetums* gene has good separation between cases and controls. The data are shown below*. Use a chi-square test to determine if this variant is associated with Type 2 Diabetes.

Bonus: consider how many tests you are performing…

	Case	Control
A	1360	1160
G	640	840

Degrees of Freedom	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
1	—	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086	16.750

**Data shows number of alleles — there were 1000 cases and 1000 controls tested*

You are running a case-control GWAS for Type 2 Diabetes. Of the 500,000 variants you test, one variant (rs4514, which has 2 alleles, A and G) near the *sweetums* gene has good separation between cases and controls. The data are shown below*. Use a chi-square test to determine if this variant is associated with Type 2 Diabetes.

Bonus: consider how many tests you are performing…

	Case	Control	Total
A	1360	1160	**2520**
G	640	840	**1480**
Total	**2000**	**2000**	**4000**

**Data shows number of alleles — there were 1000 cases and 1000 controls tested*

You are running a case-control GWAS for Type 2 Diabetes. Of the 500,000 variants you test, one variant (rs4514, which has 2 alleles, A and G) near the *sweetums* gene has good separation between cases and controls. The data are shown below*. Use a chi-square test to determine if this variant is associated with Type 2 Diabetes.

Bonus: consider how many tests you are performing…

	Case	Control	Total
A	1360	1160	**2520**
G	640	840	**1480**
Total	**2000**	**2000**	**4000**

$$e = \frac{(row_T)(column_T)}{Grand_T} = \frac{(2520)(2000)}{4000} = 1260$$

**Data shows number of alleles — there were 1000 cases and 1000 controls tested*

You are running a case-control GWAS for Type 2 Diabetes. Of the 500,000 variants you test, one variant (rs4514, which has 2 alleles, A and G) near the *sweetums* gene has good separation between cases and controls. The data are shown below*. Use a chi-square test to determine if this variant is associated with Type 2 Diabetes.

Bonus: consider how many tests you are performing…

	Case	Control	Total
A	1360 (1260)	1160 (1260)	**2520**
G	640 (740)	840 (740)	**1480**
Total	**2000**	**2000**	**4000**

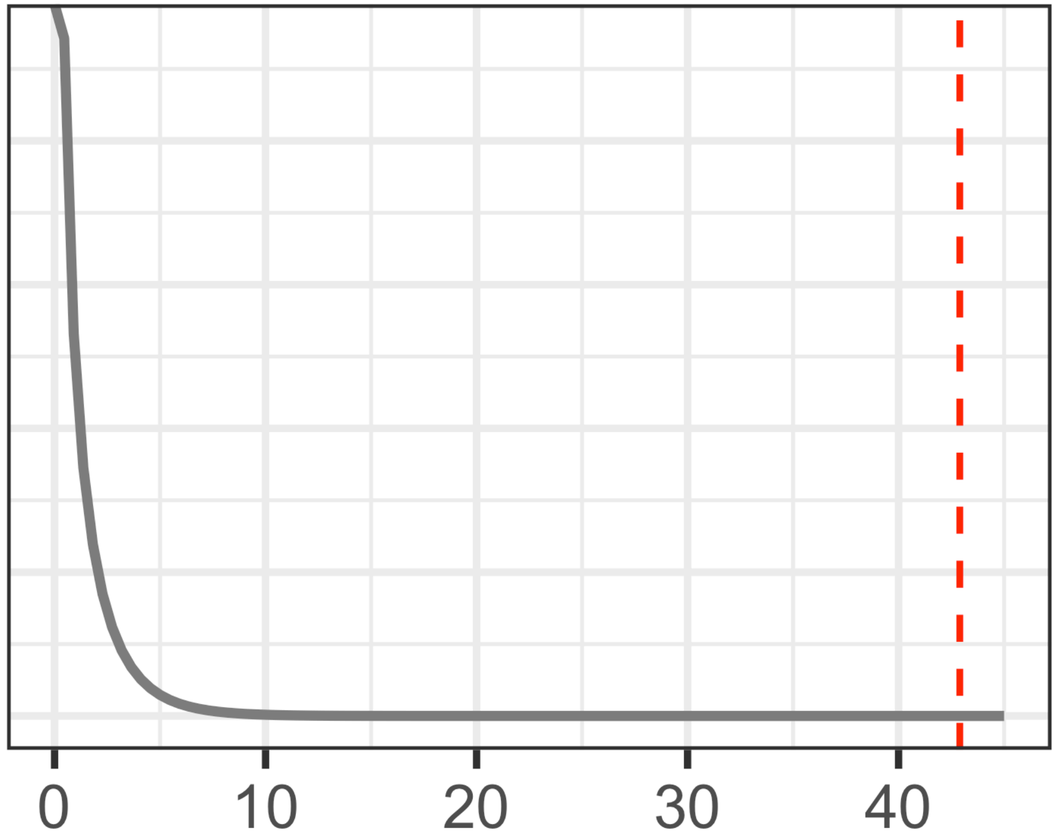
$$df = 1$$

$$\boxed{\chi_s^2 = \sum \frac{(o_i - e_i)^2}{e_i}} = \frac{(1360 - 1260)^2}{1260} + \frac{(1160 - 1260)^2}{1260} + \frac{(640 - 740)^2}{740} + \frac{(840 - 740)^2}{740} = 42.9$$

**Data shows number of alleles — there were 1000 cases and 1000 controls tested*

You are running a case-control GWAS for Type 2 Diabetes. Of the 500,000 variants you test, one variant (rs4514, which has 2 alleles, A and G) near the *sweetums* gene has good separation between cases and controls. The data are shown below*. Use a chi-square test to determine if this variant is associated with Type 2 Diabetes.

Bonus: consider how many tests you are performing…



$$\chi^2 = 42.9; df = 1$$

```
> pchisq(42.9, 1, lower.tail = F)
```

[1] 5.761066e-11

```
> 5.761066e-11 * 500000
```

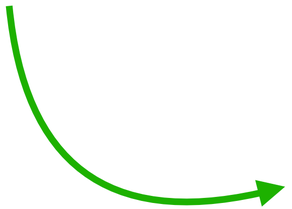
[1] 2.880533e-05 REJECT the null

α = 0.05 / 500000 = [1] 1e-07

Two contexts for contingency tables

(Migraine example)

- **Two independent samples** with **a dichotomous** observed variable
- **One sample** with **two dichotomous** observed variables



	Dark hair	Light hair	Total
Dark eyes	726	131	**857**
Light eyes	3,129	2,814	**5,943**
Total	**3,855**	**2,945**	**6,800**

- Sample of 6,800 people
- Some had dark hair, some had light hair

Math is same in both contexts, but conclusions and interpretation of hypothesis can be different

Dichotomous = two possible values

Independence and association

Responses

	Real surgery	Sham surgery	Total
Success	41 (36.59)	15 (19.41)	**56**
No success	8 (12.41)	11 (6.59)	**19**
Total	**49**	**26**	**75**

Treatments

Variable 2

	Dark hair	Light hair	Total
Dark eyes	726	131	**857**
Light eyes	3,129	2,814	**5,943**
Total	**3,855**	**2,945**	**6,800**

Variable 1

Pr{Success | Real} ?= Pr{Success | Sham}

Pr{Real | Success} ?= Pr{Sham | Success}

Pr{D.eyes | D.hair} ?= Pr{D.eyes | L.hair}

Pr{D.hair| D.eyes} ?= Pr{L.hair | D.eyes}

Chi-square “test of independence”

When the data is viewed as a single sample with two observed variables, the relationship expressed by H_0 is called **statistical independence**

Variable 2

	Dark hair	Light hair	Total
Dark eyes	726	131	**857**
Light eyes	3,129	2,814	**5,943**
Total	**3,855**	**2,945**	**6,800**

Variable 1

H_0: Hair color and eye color are independent.

These hypotheses are the same

Pr{Dark hair | Dark eyes} ?= Pr{Dark hair | light eyes}

Pr{Dark eyes | Dark hair} ?= Pr{Dark eyes | light hair}

Language of “association”

“… you should say what you mean,” the March hare went on.

“I do,” Alice hastily replied; “at least - at least I mean what I say - that’s the same thing you know.”

“Not the same thing a bit!” Said the Hatter. “Why, you might just as well say that ‘I see what I eat’ is the same thing as ‘I eat what I see’!”

Alice in Wonderland (Lewis Carroll)

Language of “association”

	Dark hair	Light hair	Total
Dark eyes	726	131	**857**
Light eyes	3,129	2,814	**5,943**
Total	**3,855**	**2,945**	**6,800**

$$\chi^2 = 314; p \approx 0$$

P(Dark eyes | Dark hair) > Pr(Dark eyes | Light hair)

There is sufficient evidence to conclude that dark-haired men have a greater tendency to be dark-eyed than do light-haired men.

Language of “association”

	Dark hair	Light hair	Total
Dark eyes	726	131	**857**
Light eyes	3,129	2,814	**5,943**
Total	**3,855**	**2,945**	**6,800**

$$\chi^2 = 314; p \approx 0$$

P(Dark eyes | Dark hair) > Pr(Dark eyes | Light hair)

There is sufficient evidence to conclude that dark-haired men have a greater tendency to be dark-eyed than do light-haired men.

than to be light-eyed.

Can be misleading…

P(Dark eyes | Dark hair) > Pr(Light eyes | Dark hair)

There is sufficient evidence to conclude that dark hair is associated with dark eyes

There is sufficient evidence to conclude that most dark-haired men are dark-eyed.

Conditions for chi-square test

- **Design conditions**: Data must be either (1) two or more independent random samples observed with respect to a categorical variable OR (2) one random sample observed with respect to two categorical variables
- **Sample size conditions**: “large enough” sample size. Rule of thumb — each expected frequency (e) must be at least 5
- **Form of H_0**: the row variable and the columns variable are independent
- **Scope of inference**: if data arise from experiments we can draw a causal inference; if the data arise from observational study we can only infer that the observed association is not due to chance but cannot rule out other explanations

The *r* x *k* contingency table

Randomized, double-blind experiment was conducted in which patients with Alzheimer's disease were given either a treatment (EGb) or placebo for one year. The change in each patient's ADAS-Cog score was measured (if score went down, the patient improved)

Treatments	< -4	-3 to -2	-1 to +1	+2 to +3	> +4	Total
EGb	22	18	12	7	16	**75**
Placebo	10	11	19	11	24	**75**
Total	**32**	**29**	**31**	**18**	**40**	**150**

Responses

Treatments	< -4	-3 to -2	-1 to +1	+2 to +3	> +4	Total
EGb	22	18	12	7	16	**75**
Placebo	10	11	19	11	24	**75**
Total	**32**	**29**	**31**	**18**	**40**	**150**

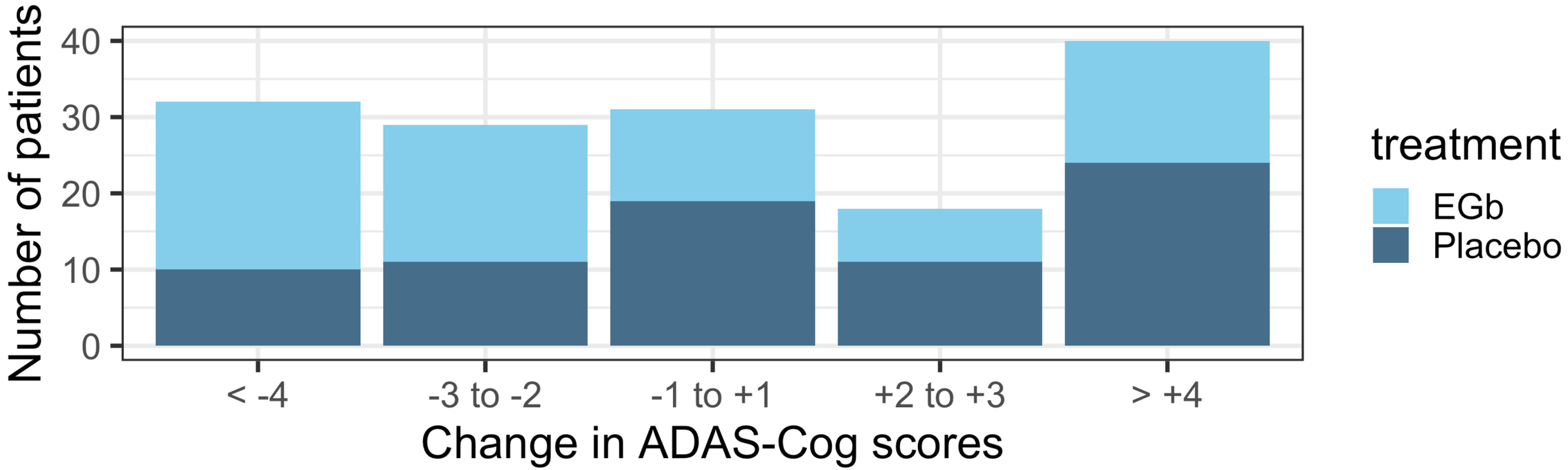
Responses

$Pr\{<-4 \mid EGb\} = Pr\{<-4 \mid Placebo\}$ $Pr\{-3 \text{ to } -2 \mid EGb\} = Pr\{-3 \text{ to } -2 \mid Placebo\}$ Etc.

H_0: Change of ADAS-Cog score is <u>independent</u> of treatment

H_A: Change of ADAS-Cog score is <u>related</u> to the treatment

	< -4	-3 to -2	-1 to +1	+2 to +3	> +4	Total
EGb	22	18	12	7	16	**75**
Placebo	10	11	19	11	24	**75**
Total	**32**	**29**	**31**	**18**	**40**	**150**



alz

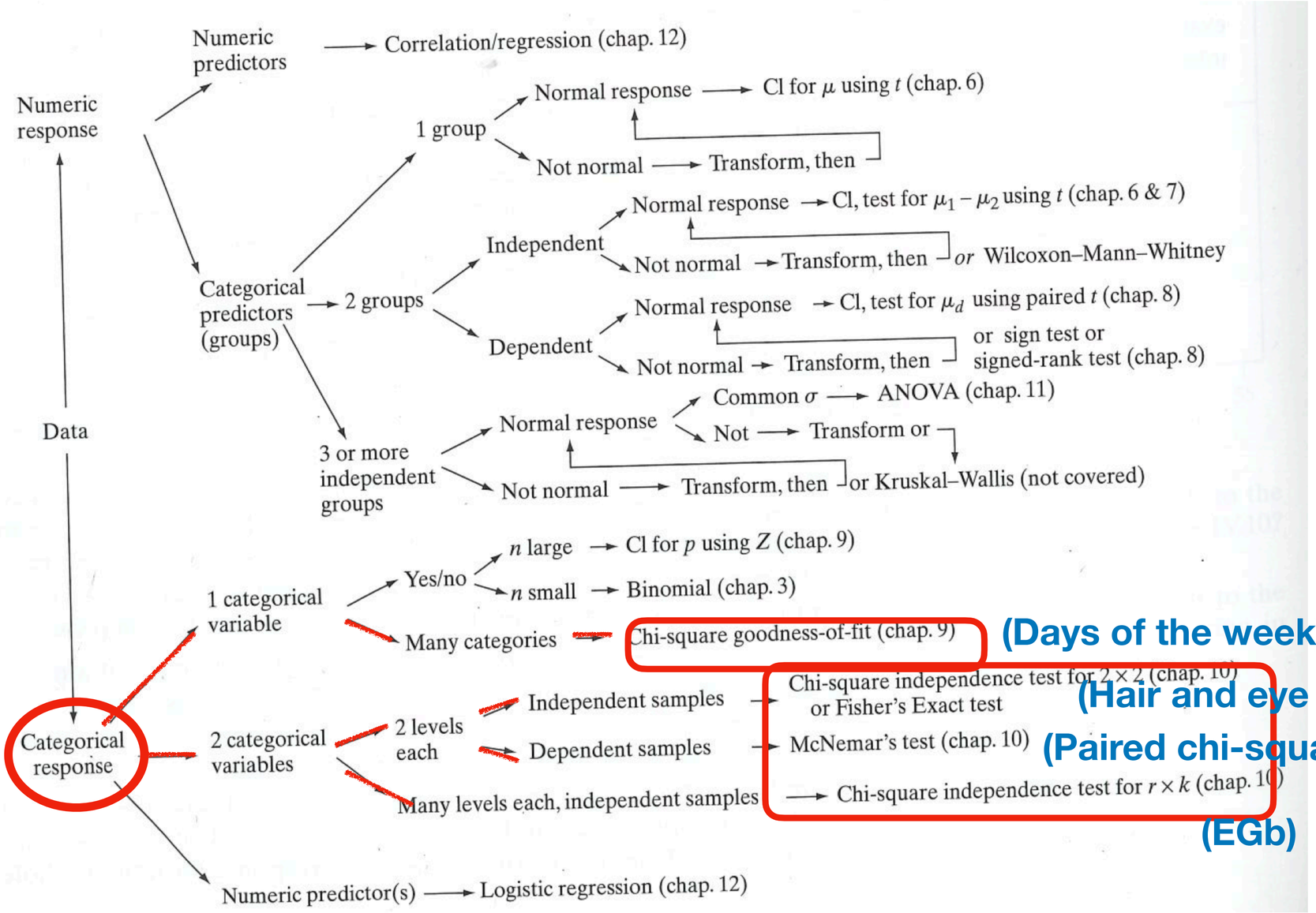
	< -4	-3 to -2	-1 to +1	+2 to +3	> +4	Total
EGb	22 (16)	18 (14.5)	12 (15.5)	7 (9)	16 (20)	**75**
Placebo	10 (16)	11 (14.5)	19 (15.5)	11 (9)	24 (20)	**75**
Total	**32**	**29**	**31**	**18**	**40**	**150**

```
chisq.test(data, correct = F)
```

```
        Pearson's Chi-squared test

data:  alz
X-squared = 10.259, df = 4, p-value = 0.03628
```

At alpha = 0.05, we reject the null hypothesis and conclude that there is a relationship between treatment and change in ADAS-Cog scores

When to use Chi-square test



(Days of the week)

(Hair and eye color)

(Paired chi-square)

(EGb)

Relative risk = p_1 / p_2

(Ratio of probabilities)

Lung Cancer?	Smoker	Former smoker
Yes	89	37
No	6063	5711
Total	**6152**	**5748**

p_1 = Pr{lung cancer | smoker}

p_1 = 89 / 6152

p_2 = Pr{lung cancer | former smoker}

p_2 = 37 / 5748

$$\frac{p_1}{p_2} = \frac{0.01447}{0.00644} = 2.247$$

"The risk (conditional probability) of developing lung cancer is about 2.2 times as great for smokers as for former smokers"

Odds ratio: another way to compare two probabilities

$$Odds = \frac{Pr(E)}{1 - Pr(E)}$$

$$\theta = \frac{odds_1}{odds_2}$$

Odds ratio

Lung Cancer?	Smoker	Former smoker
Yes	89	37
No	6063	5711
Total	**6152**	**5748**

$$Odds_1 = \frac{89/6152}{1 - (89/6152)} = 0.01468$$

(Smokers)

$$Odds_2 = \frac{37/5748}{1 - (37/5748)} = 0.00648$$

(Former smokers)

"The odds of developing lung cancer are about 2.3 times as great for smokers as non-smokers."

$$\theta = \frac{0.01468}{0.00648} = 2.265$$

Comparing odds ratio and relative risk

Risk = $\frac{89}{6152}$ **Odds** = $\frac{89}{6063}$

Lung Cancer?	Smoker	Former smoker
Yes	89	37
No	6063	5711
Total	**6152**	**5748**

Similar with p_1 and p_2 are small

$$\text{Odds ratio} = (\text{relative risk})\left(\frac{1-p_2}{1-p_1}\right)$$

- Although relative risk is easier to interpret than odds ratio, odds ratio has certain advantages
- **Odds ratio can be estimated even though p_1 and p_2 cannot be estimated**

Comparing odds ratio and relative risk

Risk = $\frac{89}{6152}$ **Odds** = $\frac{89}{6063}$

Odds ratio is constant row-wise and column-wise but relative risk is NOT

Lung Cancer?	Smoker	Former smoker
Yes	89	37
No	6063	5711
Total	6152	5748

$$OR = \frac{89/6063}{37/5711} = \frac{89/37}{6063/5711} = 2.27$$

$$RR_c = \frac{89/(89+6063)}{37/(37+5711)} = 2.25$$

Odds ratio can be estimated even if p_1 and p_2 cannot be estimated

$$RR_r = \frac{89/(89+37)}{6063/6063+5711)} = 1.37$$

Comparing odds ratio and relative risk

Let's say this is a case-control study where we chose 500 men with lung cancer and 500 men without lung cancer and then surveyed smoking histories

- **Odds ratio can be estimated even though p_1 and p_2 cannot be estimated**

Lung Cancer?	Smoker	Former smoker	Total
Yes	273	277	**500**
No	173	327	**500**

$$\Pr\{\text{smoker} \mid \text{lung cancer}\} = \frac{273}{500}$$

$$\Pr\{\text{lung cancer} \mid \text{smoker}\} = ?$$

We don't actually know the probabilities of having lung cancer given smoker because the relative frequencies of lung cancer was fixed in this study.

Relative risk with 23andMe

Elevated Risk

NAME	CONFIDENCE	YOUR RISK	AVG. RISK	COMPARED TO AVERAGE
Atrial Fibrillation	★★★★	33.9%	27.2%	1.25x
Prostate Cancer ♂	★★★★	23.8%	17.8%	1.33x
Gallstones	★★★★	11.1%	7.0%	1.58x
Exfoliation Glaucoma	★★★★	2.2%	0.7%	2.90x
Ulcerative Colitis	★★★★	1.00%	0.77%	1.30x
Esophageal Squamous Cell Carcinoma (ESCC)	★★★★	0.43%	0.36%	1.21x
Stomach Cancer (Gastric Cardia Adenocarcinoma)	★★★★	0.28%	0.23%	?
Abdominal Aortic Aneurysm	★★★			
Alopecia Areata	★★★			

What is the relative risk (and odds ratio) of this person getting stomach cancer compared to the average person?

Relative risk with 23andMe

Elevated Risk

NAME	CONFIDENCE	YOUR RISK	AVG. RISK	COMPARED TO AVERAGE
Atrial Fibrillation	★★★★	33.9%	27.2%	1.25x
Prostate Cancer ♂	★★★★	23.8%	17.8%	1.33x
Gallstones	★★★★	11.1%	7.0%	1.58x
Exfoliation Glaucoma	★★★★	2.2%	0.7%	2.90x
Ulcerative Colitis	★★★★	1.00%	0.77%	1.30x
Esophageal Squamous Cell Carcinoma (ESCC)	★★★★	0.43%	0.36%	1.21x
Stomach Cancer (Gastric Cardia Adenocarcinoma)	★★★★	0.28%	0.23%	1.22x
Abdominal Aortic Aneurysm	★★★			
Alopecia Areata	★★★			

$$RR = p_1 / p_2$$

Relative risk of 1.217x (0.0028/0.0023) means you are 22% more likely (than average) to develop the disease

Relative risk with 23andMe

Elevated Risk

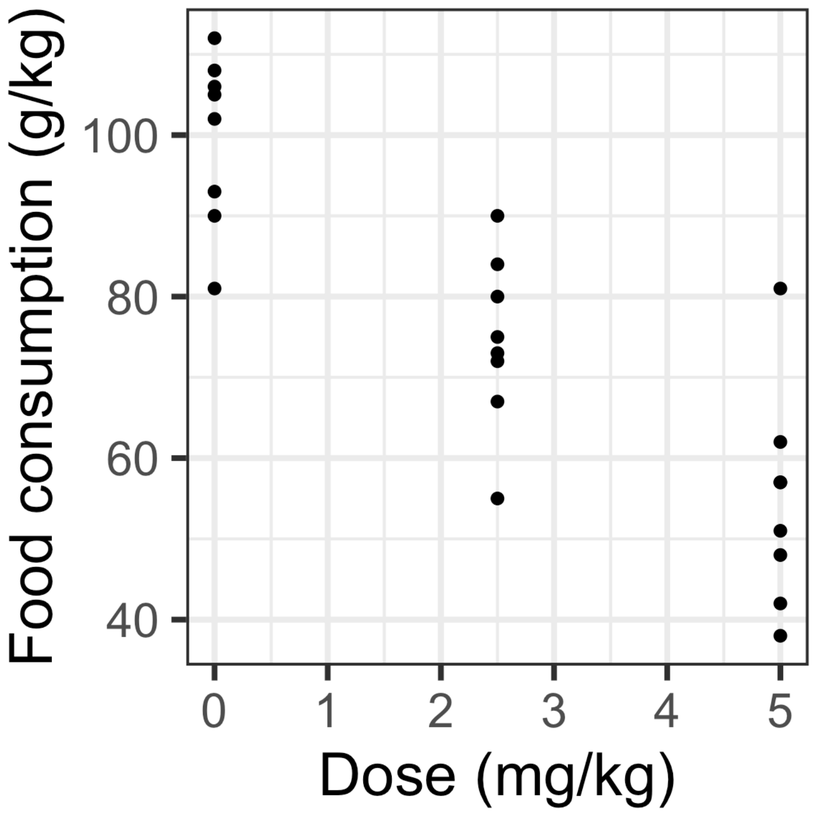
NAME	CONFIDENCE	YOUR RISK	AVG. RISK	COMPARED TO AVERAGE
Atrial Fibrillation	★★★★	33.9%	27.2%	1.25x
Prostate Cancer ♂	★★★★	23.8%	17.8%	1.33x
Gallstones	★★★★	11.1%	7.0%	1.58x
Exfoliation Glaucoma	★★★★	2.2%	0.7%	2.90x
Ulcerative Colitis	★★★★	1.00%	0.77%	1.30x
Esophageal Squamous Cell Carcinoma (ESCC)	★★★★	0.43%	0.36%	1.21x
Stomach Cancer (Gastric Cardia Adenocarcinoma)	★★★★	0.28%	0.23%	1.22x
Abdominal Aortic Aneurysm	★★★			
Alopecia Areata	★★★			

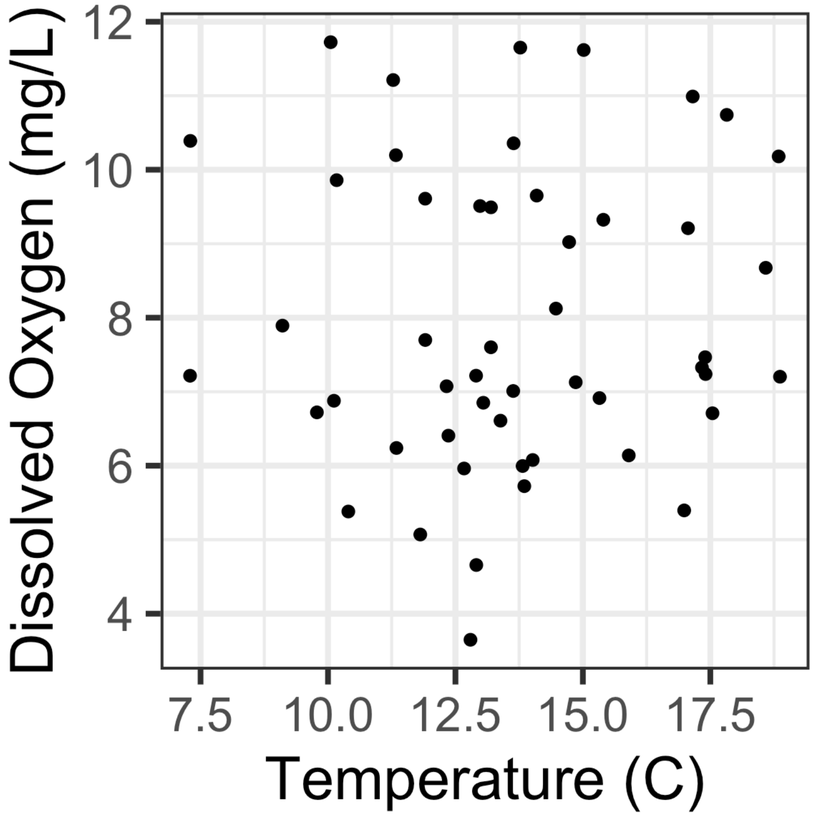
$$OR = \frac{p_1/(1-p_1)}{p_2/(1-p_2)}$$

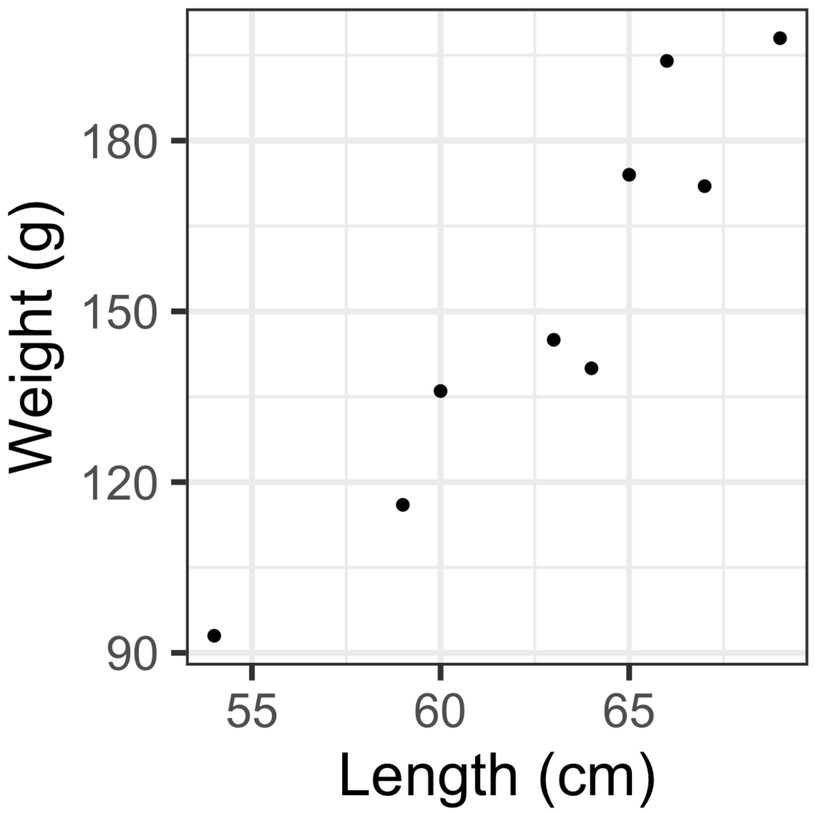
Odds ratio = (0.0028/0.9972)/(0.0023/0.9977) = 1.218

Describing relationships between variables

Pairs of observations can be used to describe the relationship or trend



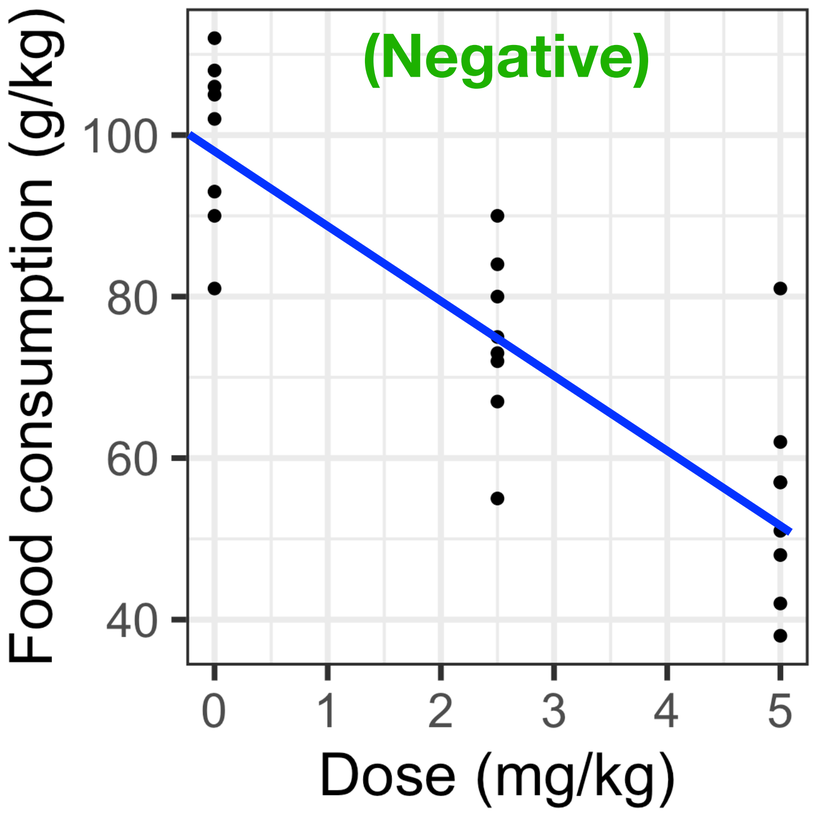


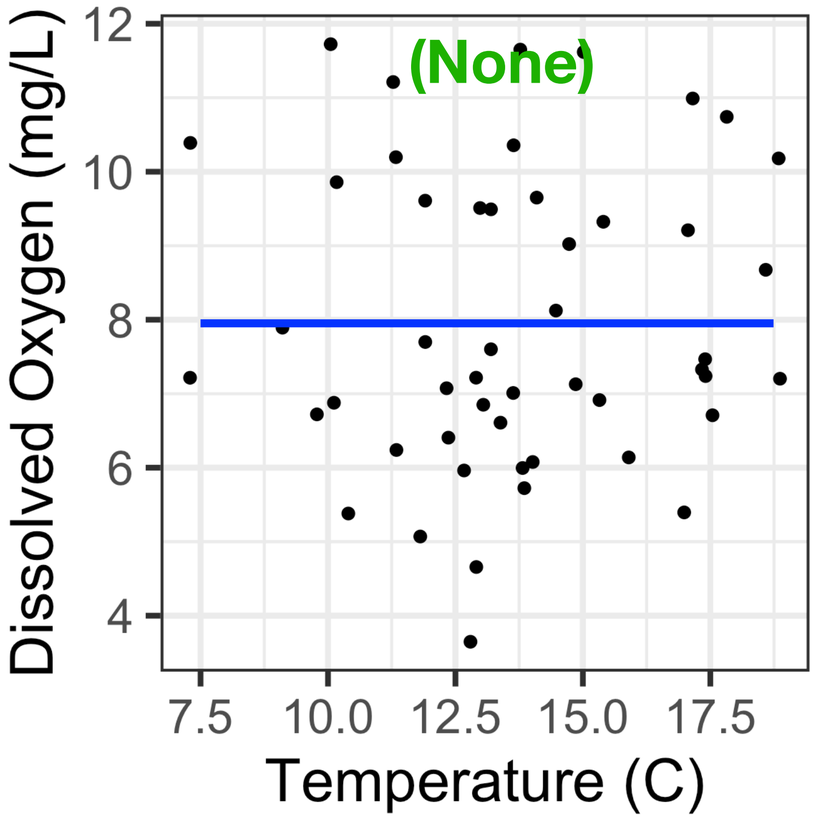


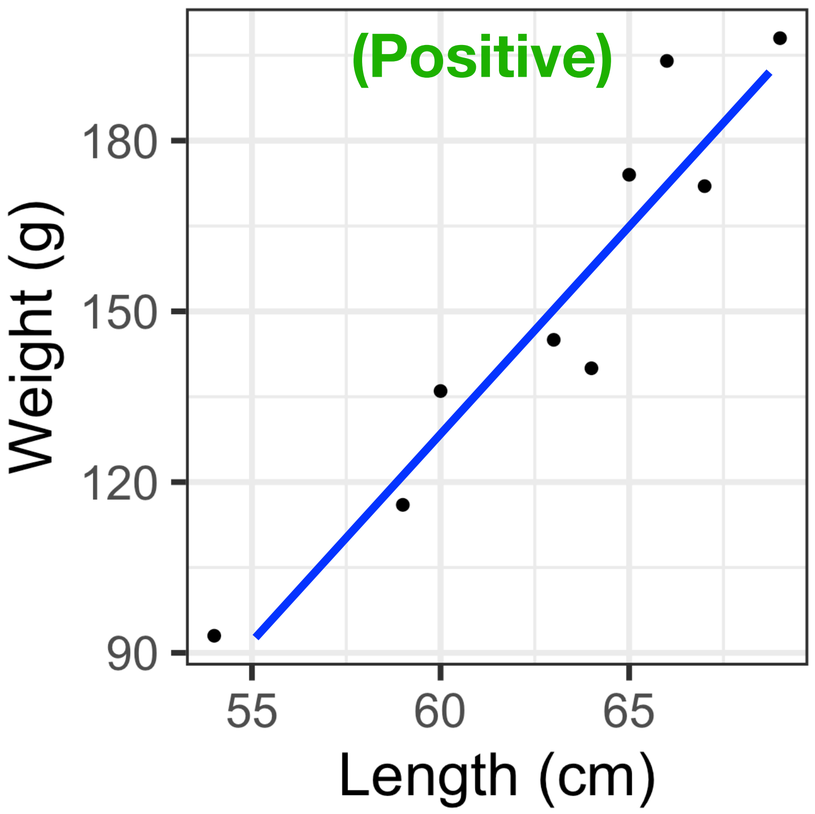
Describing relationships between variables

Direction of a linear association can be measured with the <u>covariance</u>

(Measure of how two variables vary together)





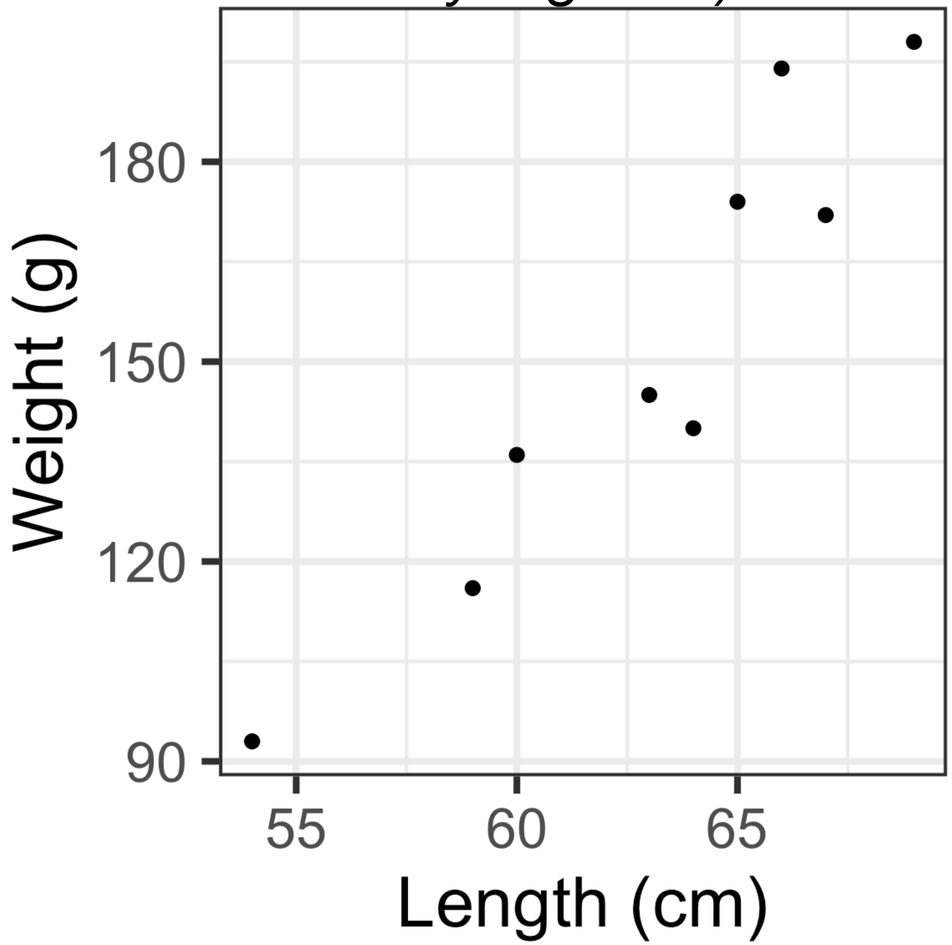


Defining the covariance of paired data

Direction of a linear association can be measured with the <u>covariance</u>

(Measure of how two variables vary together)

1. Find the mean of x and y



Defining the covariance of paired data

Direction of a linear association can be measured with the <u>covariance</u>

(Measure of how two variables vary together)

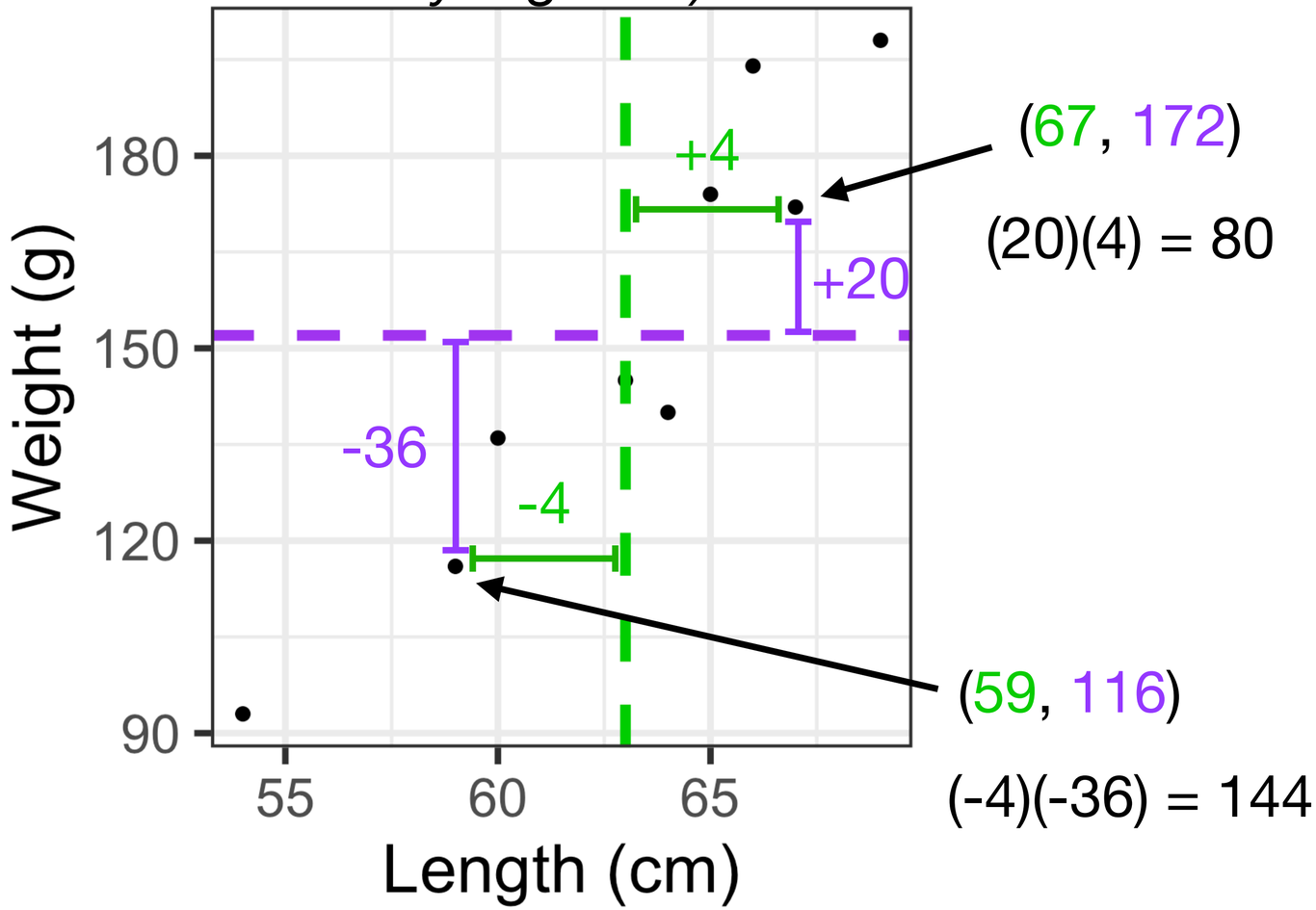
1. Find the mean of x and y

2. Calculate sum of product of deviances

$$\sum (x - \bar{x})(y - \bar{y}) = 1237$$

3. Divide by degrees of freedom (n-1)

1237/(9-1) = **154.62**



Defining the covariance of paired data

Direction of a linear association can be measured with the <u>covariance</u>

(Measure of how two variables vary together)

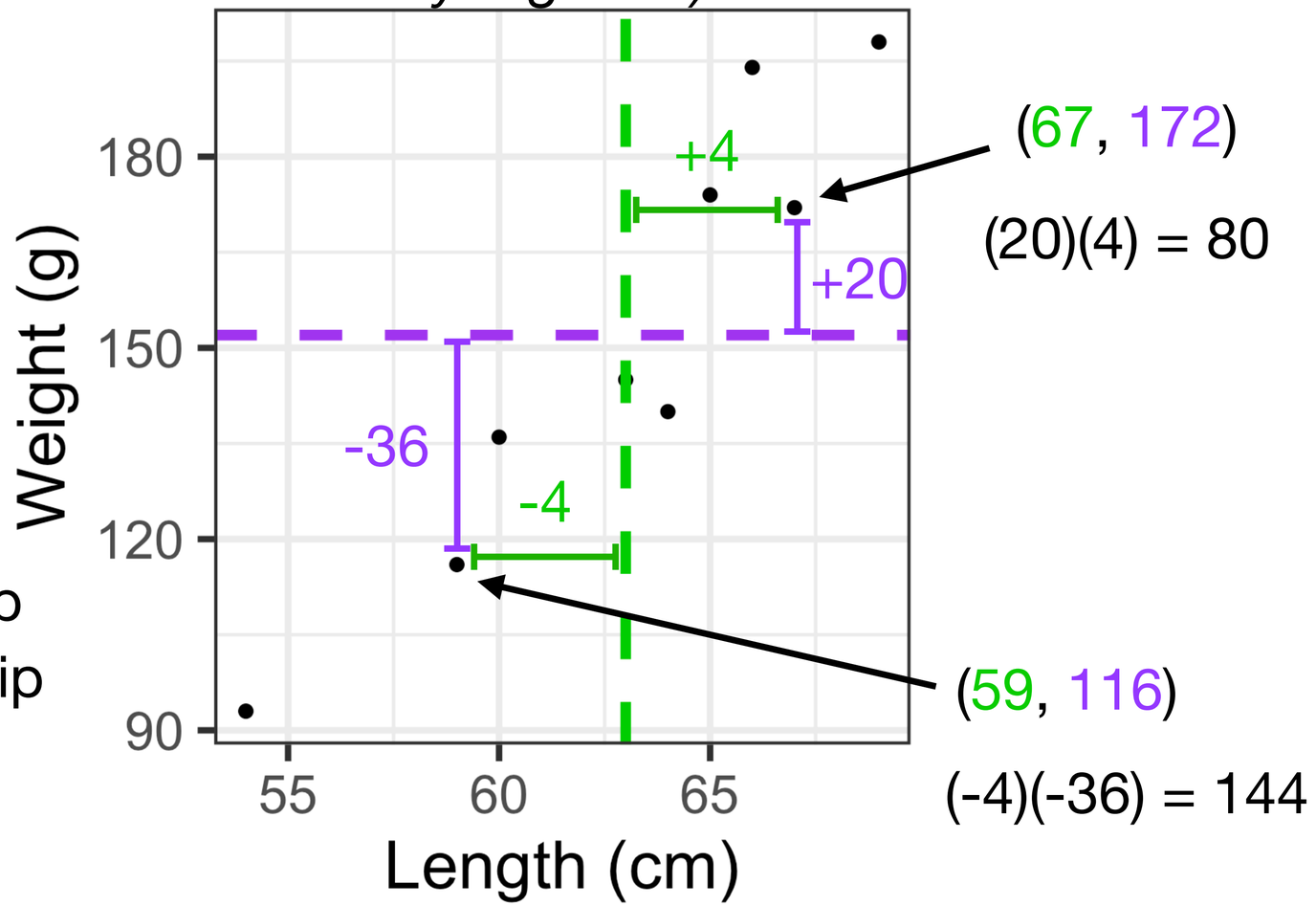
$$\frac{\sum (x - \bar{x})(y - \bar{y})}{n - 1}$$

= **154.62 cm*g**

Positive covariance = positive relationship

Negative covariance = negative relationship

Covariance is difficult to interpret and doesn't tell us the whole story



Defining the covariance of paired data

Direction of a linear association can be measured with the <u>covariance</u>

(Measure of how two variables vary together)

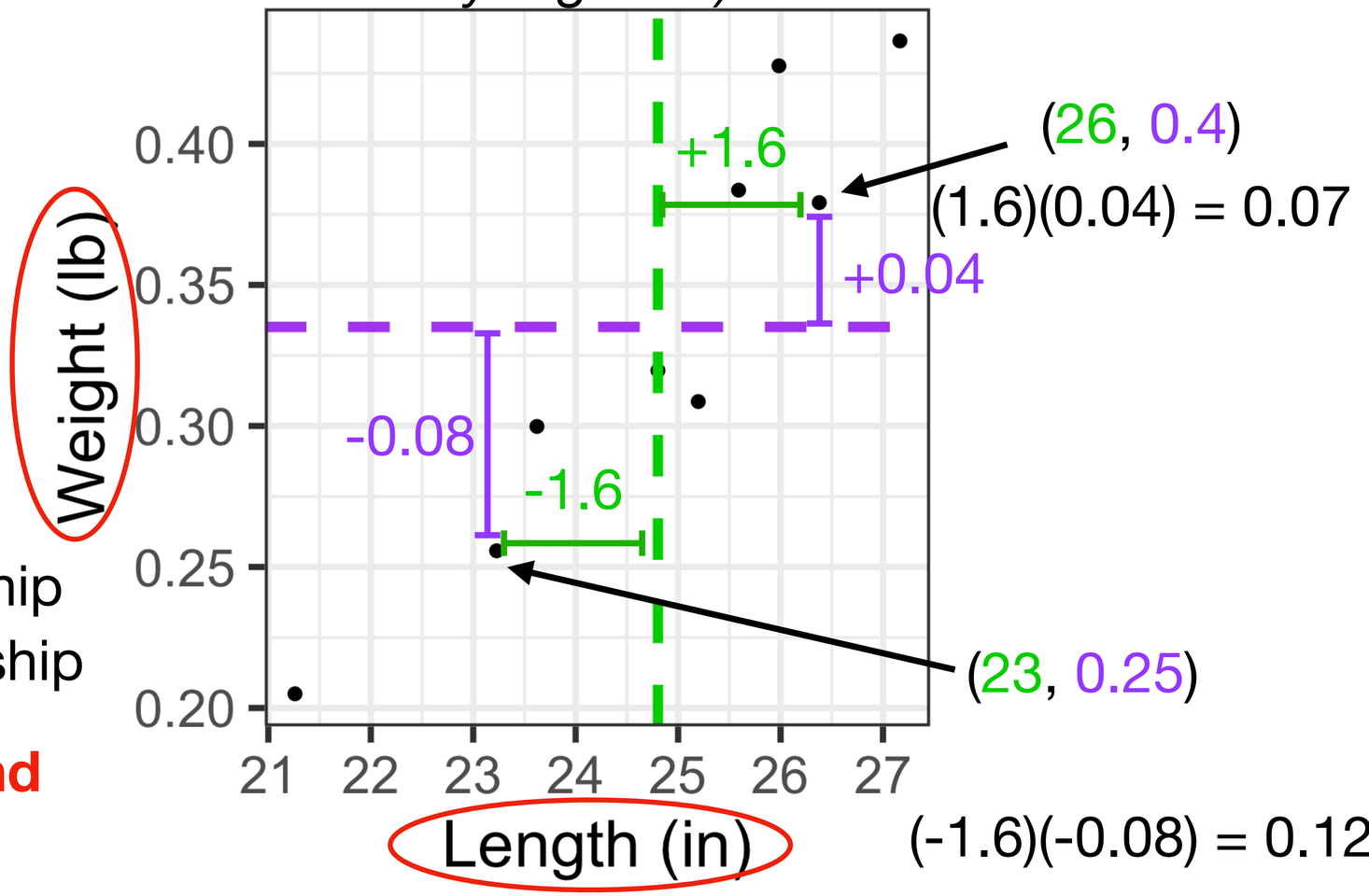
$$\frac{\sum (x - \bar{x})(y - \bar{y})}{n - 1}$$

= **1.07 in*lb**

Positive covariance = positive relationship

Negative covariance = negative relationship

Covariance is difficult to interpret and doesn't tell us the whole story



Defining the **correlation** of paired data

Direction and magnitude of a linear association can be measured with the correlation

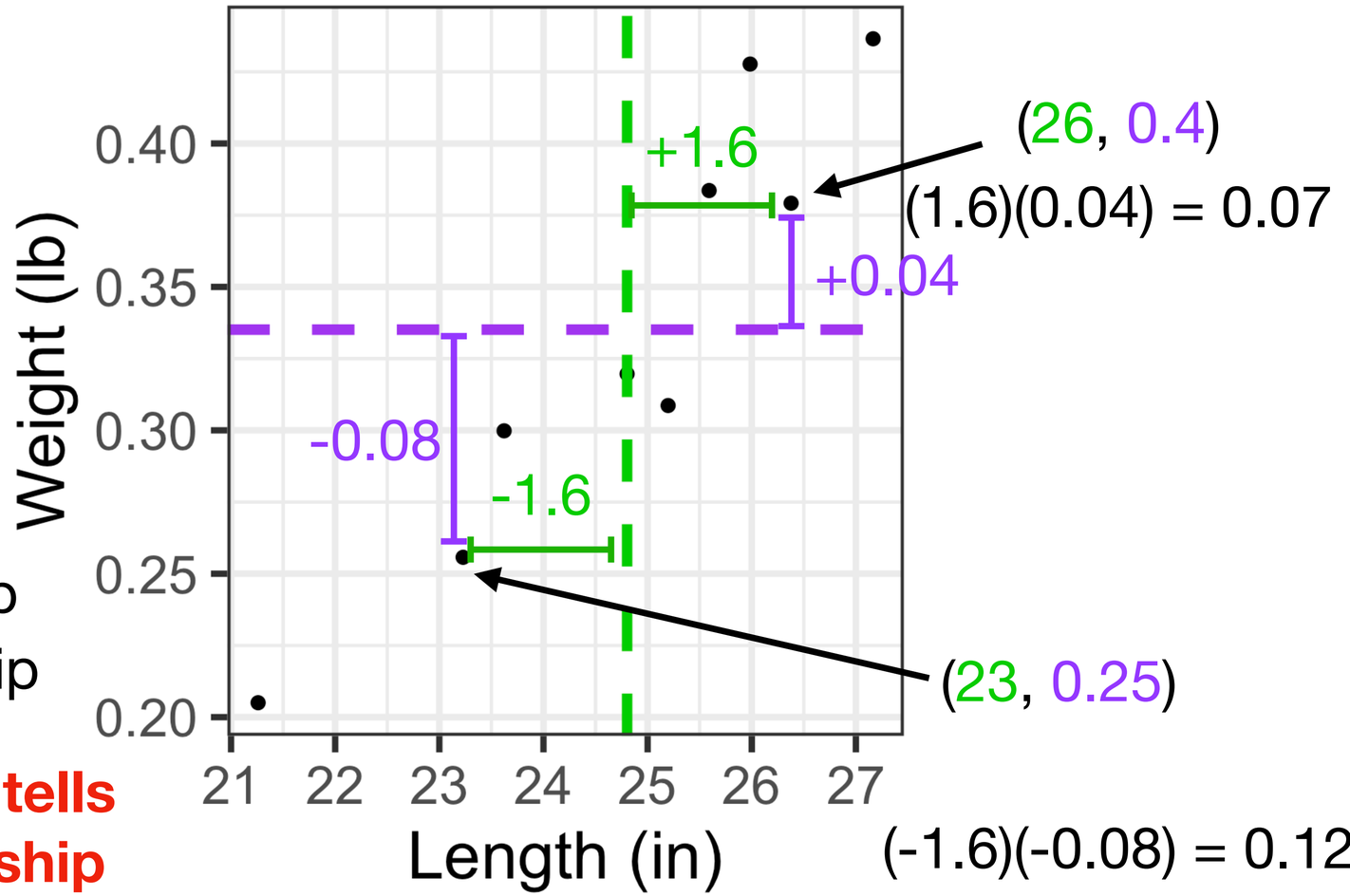
$$r = \frac{Cov}{s_x s_y}$$

1.07/(1.82*0.08) = 0.943

Positive correlation = positive relationship

Negative correlation = negative relationship

Correlation is scale invariant AND also tells us about the magnitude of the relationship



Defining the **correlation** of paired data

Direction and magnitude of a linear association can be measured with the correlation

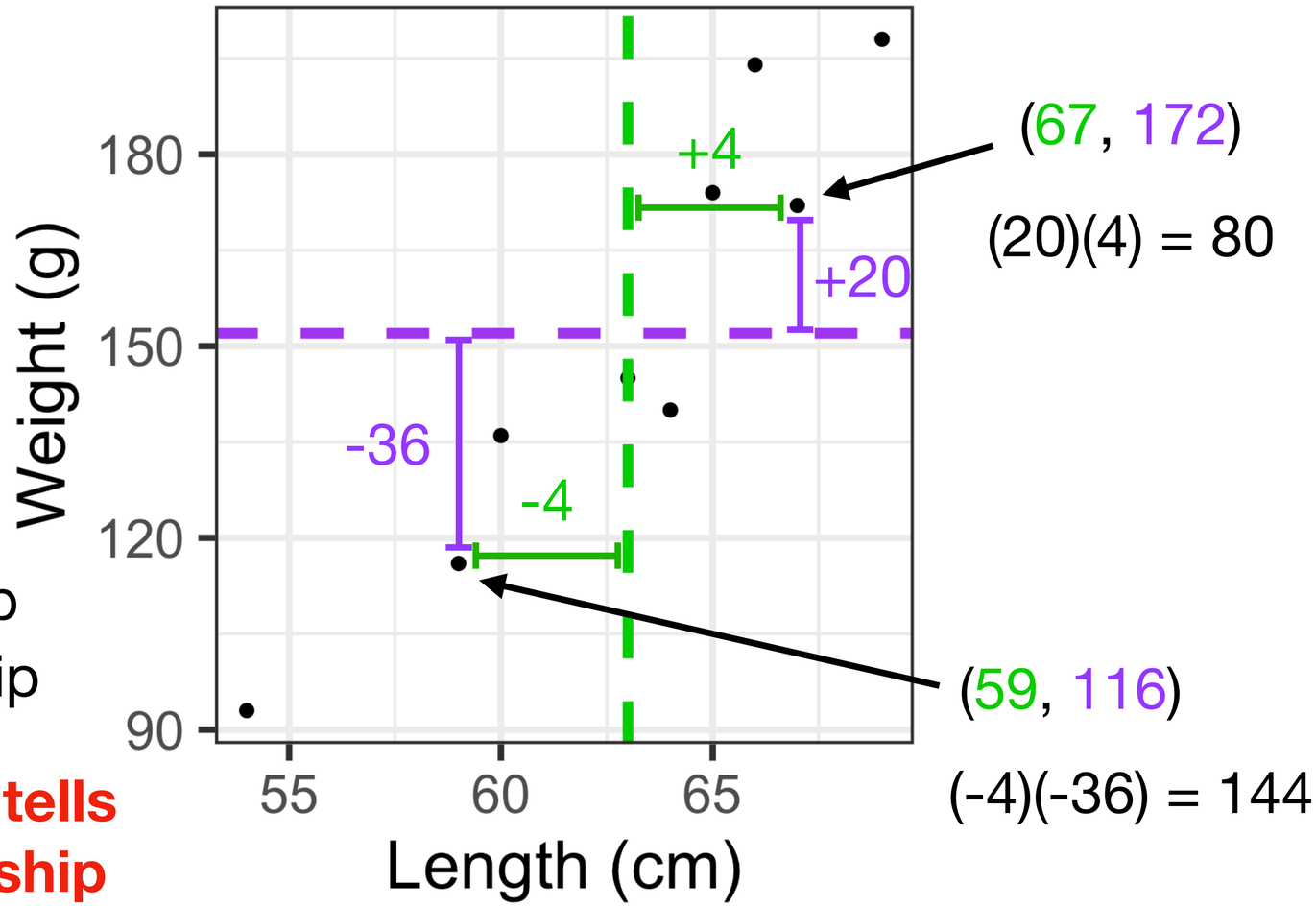
$$r = \frac{Cov}{s_x s_y}$$

154.62/(4.6*35.3) = 0.943

Positive correlation = positive relationship

Negative correlation = negative relationship

Correlation is scale invariant AND also tells us about the magnitude of the relationship



Defining the **correlation** of paired data

Direction and magnitude of a linear association can be measured with the correlation

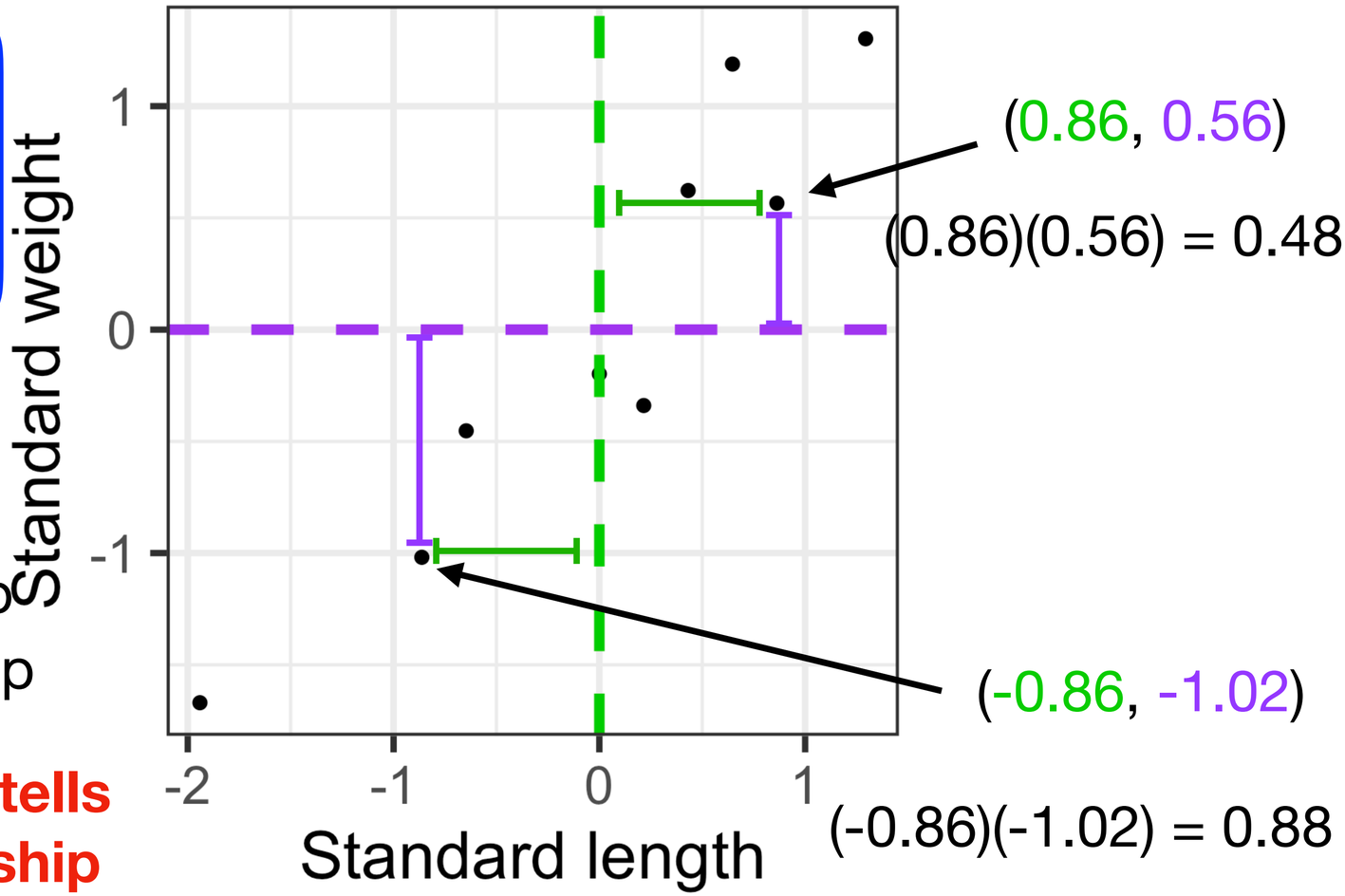
$$r = \frac{1}{1-n}\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{x_y}\right)$$

```
> cor(x, y)
```
= 0.943

Positive correlation = positive relationship

Negative correlation = negative relationship

Correlation is scale invariant AND also tells us about the magnitude of the relationship



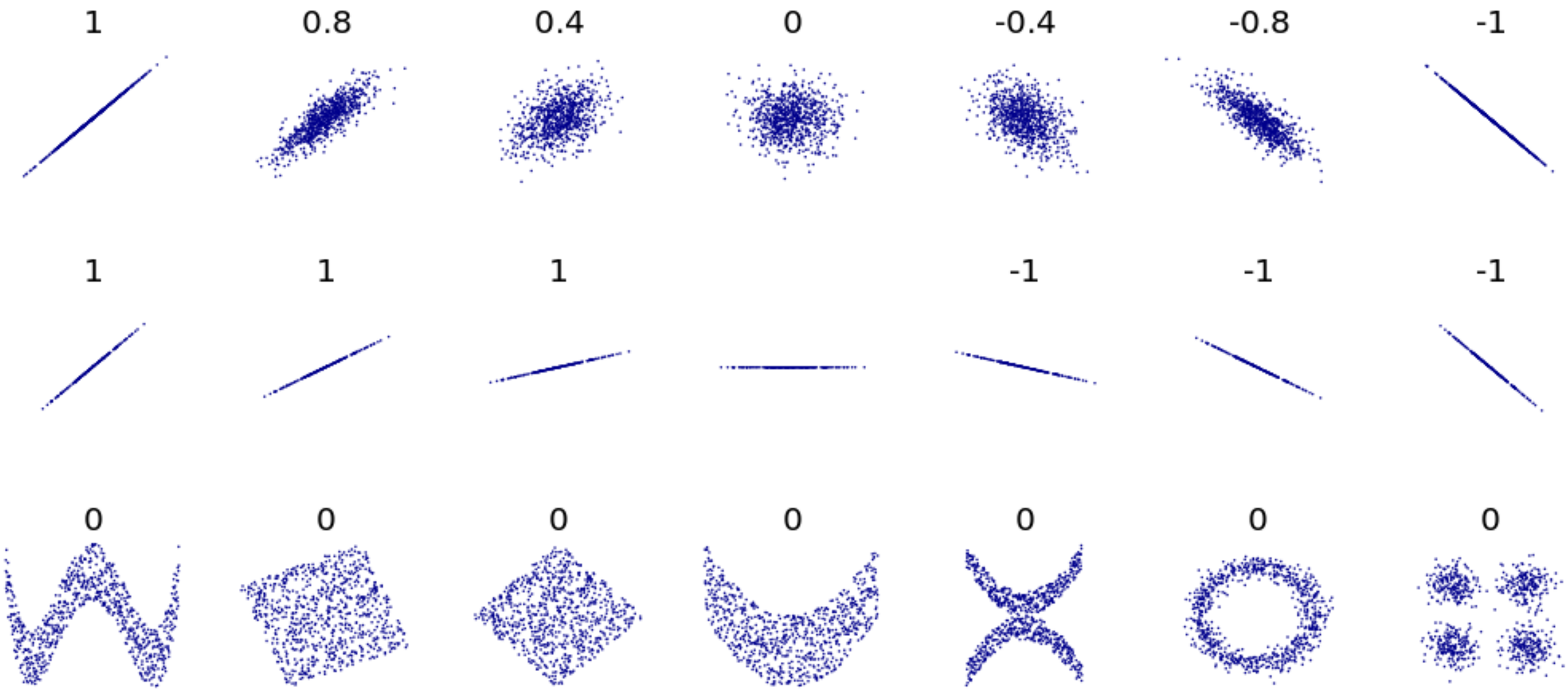
Interpreting the correlation coefficient

Direction and magnitude of a linear association can be measured with the correlation

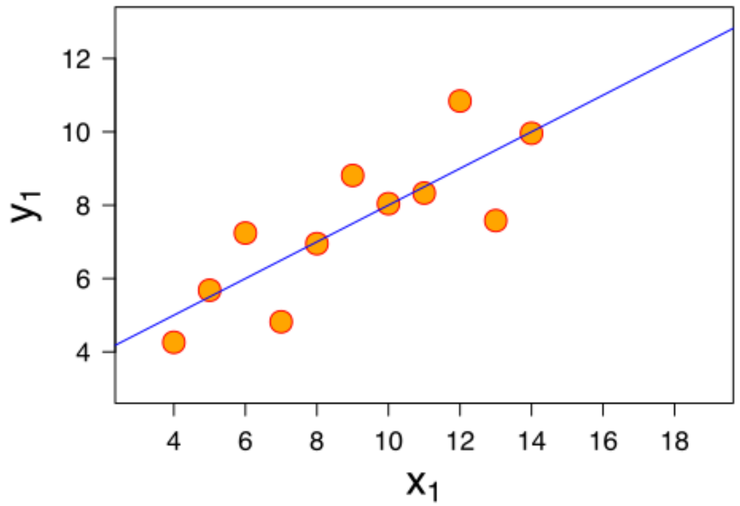
$$r = \frac{1}{1-n}\sum\left(\frac{x-\bar{x}}{s_x}\right)\left(\frac{y-\bar{y}}{x_y}\right)$$

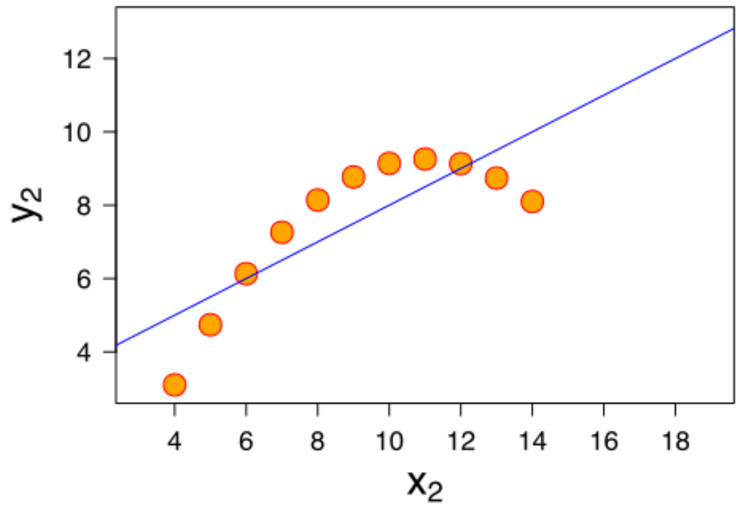
- Unit-free (scale invariant) between -1 and 1
- Sign of correlation = sign of relationship
- Closer correlation to 1 (or -1) = stronger the relationship
    - Correlation of 0 = no **linear** relationship
- X and Y can be interchanged

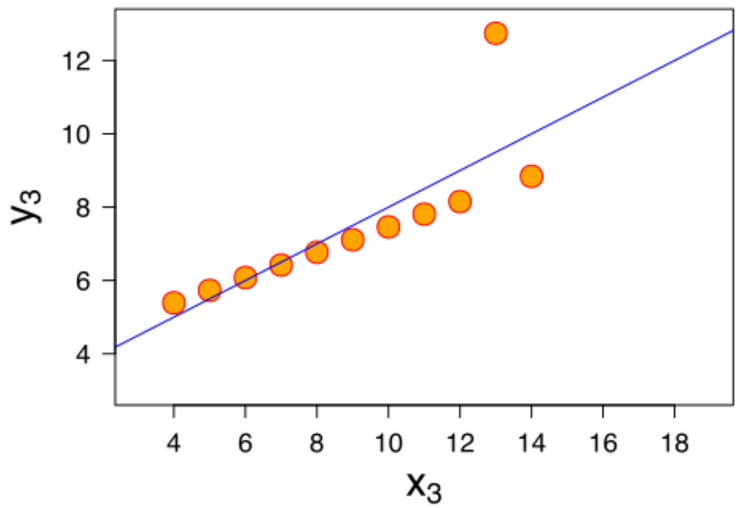
Interpreting the correlation coefficient

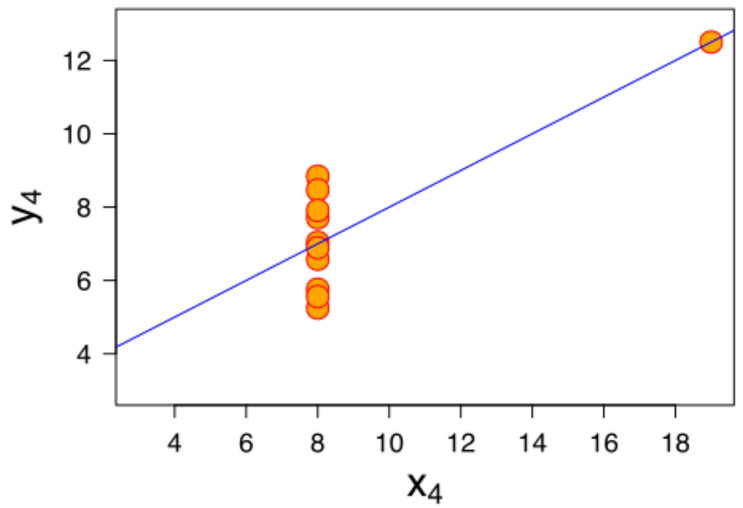


Interpreting the correlation coefficient









Assumptions for estimating correlation

Sample correlation *r* is an estimate of population correlation ρ (“rho”)

- Both X and Y values were selected at random (**bivariate random sampling model**)
    - Each *pair* is randomly sampled from a population of (x, y) pairs
    - This assumption is broken when the X values are specified by the experimenter (**random subsampling model**)

Testing the hypothesis $H_0 : \rho = 0$

1. Randomization test

Q: How likely is it that a correlation coefficient would be as far from zero as is our observed value of *r*, just by chance?

1. Calculate the sample correlation coefficient, *r*
2. Hold the X values constant and scramble the Y values (breaking the link)
3. Calculate the sample correlation coefficient for this random pairing
4. Record whether the | r | is at least as great as the real sample correlation
5. Repeat many times, count the fraction TRUE

Testing the hypothesis $H_0 : \rho = 0$

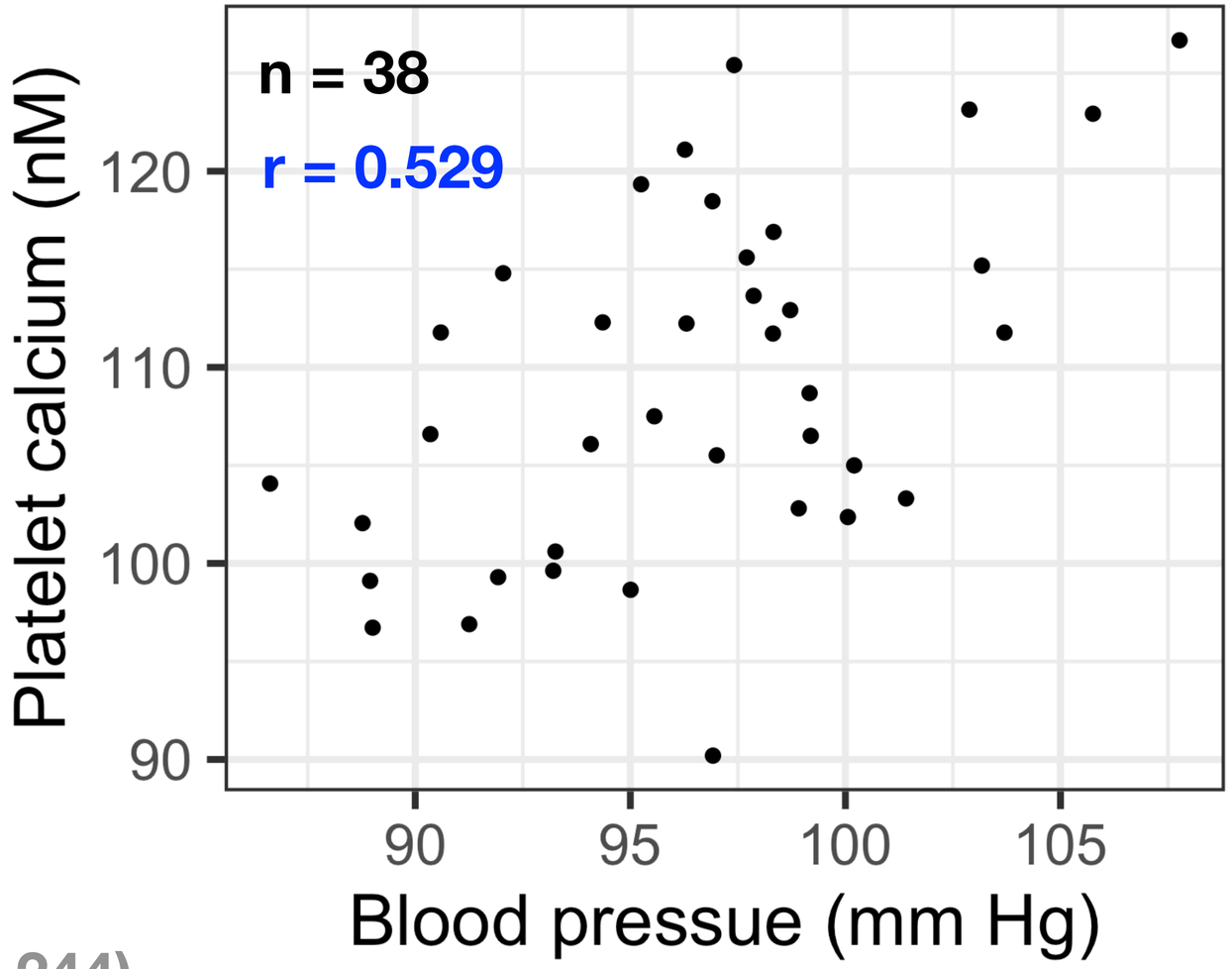
1. Randomization test

H_0: Platelet calcium **is not** linearly related to blood pressure

H_A: Platelet calcium **is** linearly related to blood pressure

X	Y	Y_1	Y_2
103	112	101	115
95	115	124	101
107	124	115	112
88	101	112	124

(0.529) (0.308) (0.0491) (-0.244)



Testing the hypothesis $H_0 : \rho = 0$

1. Randomization test

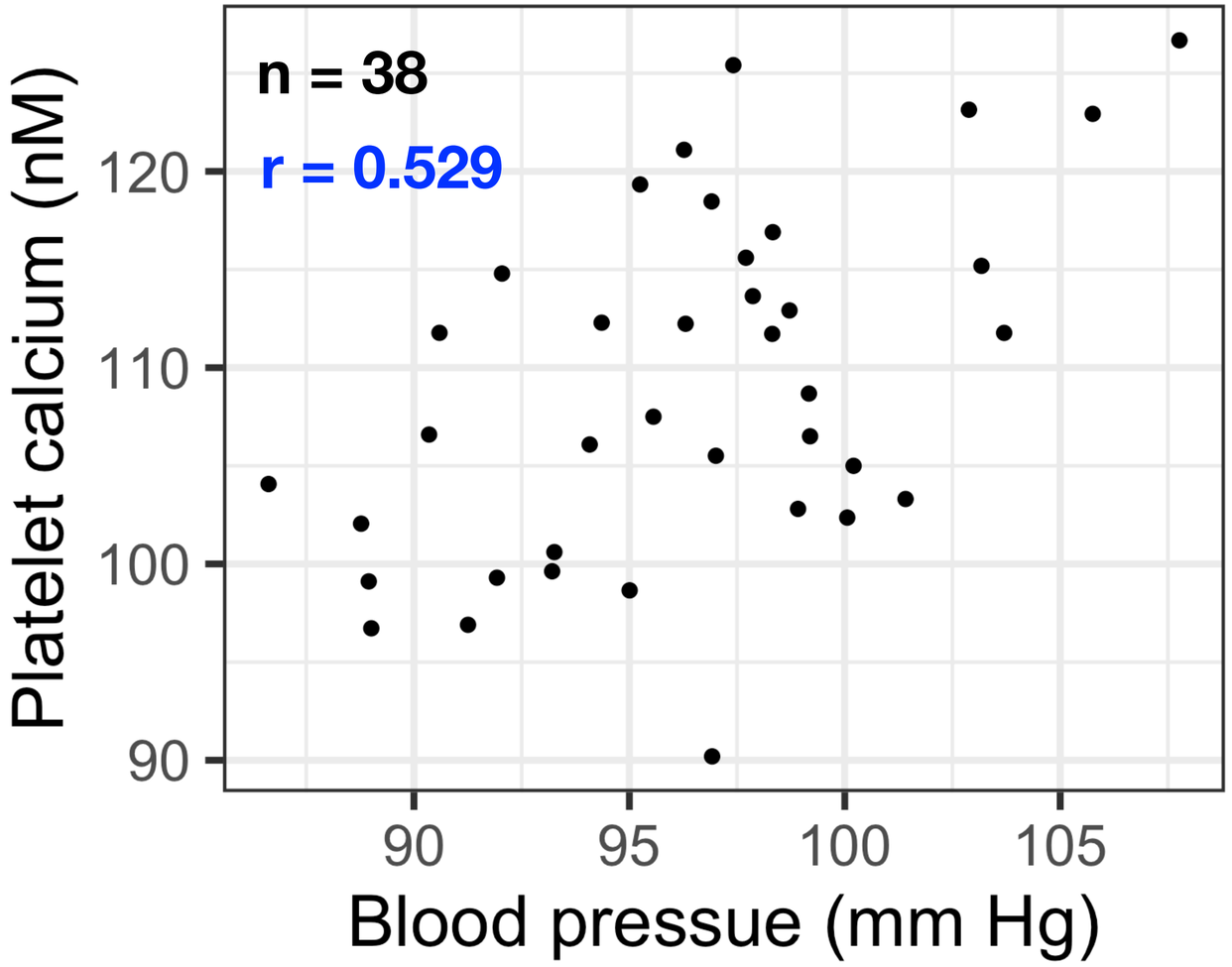
H_0: Platelet calcium **is not** linearly related to blood pressure

H_A: Platelet calcium **is** linearly related to blood pressure

```
> sum(abs(cors) >= 0.529)
```

```
4 / 10000 = 0.0004
```

Reject the null hypothesis ✓



Testing the hypothesis $H_0 : \rho = 0$

2. *t*-test

Q: How likely is it that a correlation coefficient would be as far from zero as is our observed value of *r*, just by chance?

1. Calculate the degrees of freedom **(n - 2)**
2. Calculate the test statistic
3. Calculate the *P*-value from the Student's *t*-distribution

Testing the hypothesis $H_0 : \rho = 0$

2. *t*-test

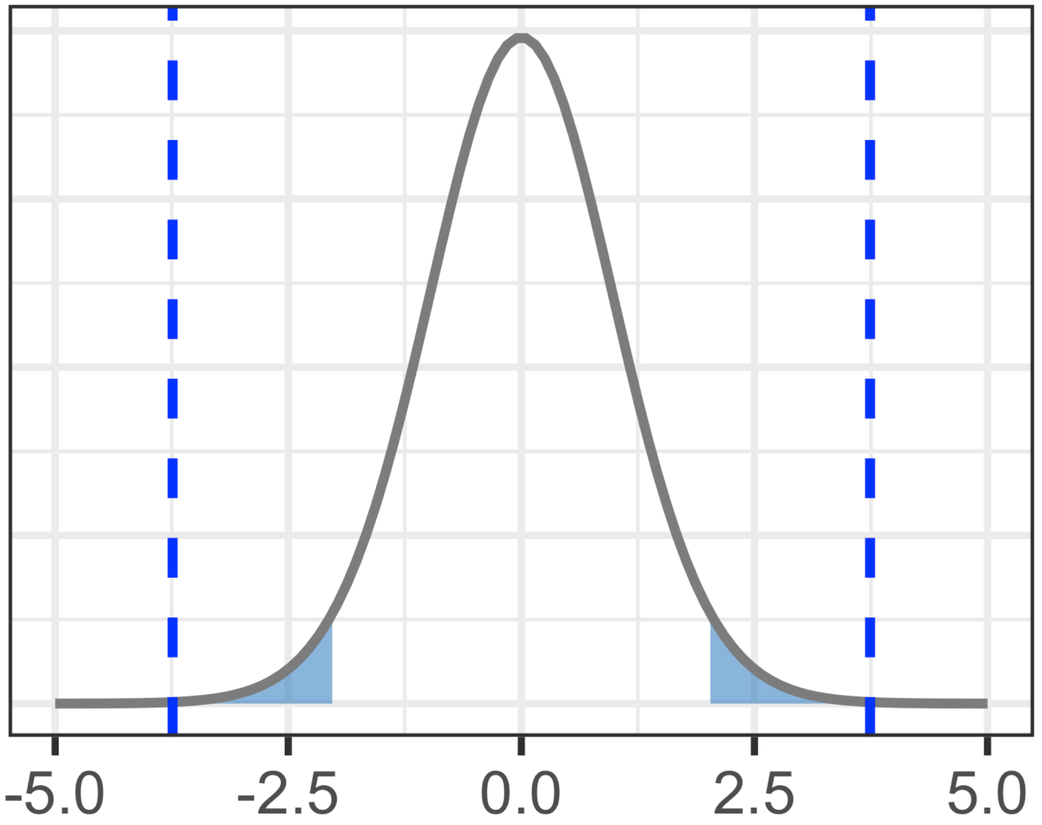
```
> cor.test(x, y)
```

$$df = n - 2 = 36$$

$$t_s = r\sqrt{\frac{n-2}{1-r^2}}$$

$$t_s = 0.529\sqrt{\frac{36}{1-0.529^2}}$$

$$t_s = 3.74$$




```
> pt(3.74, 36, lower.tail = F)*2
[1] 0.000639
```

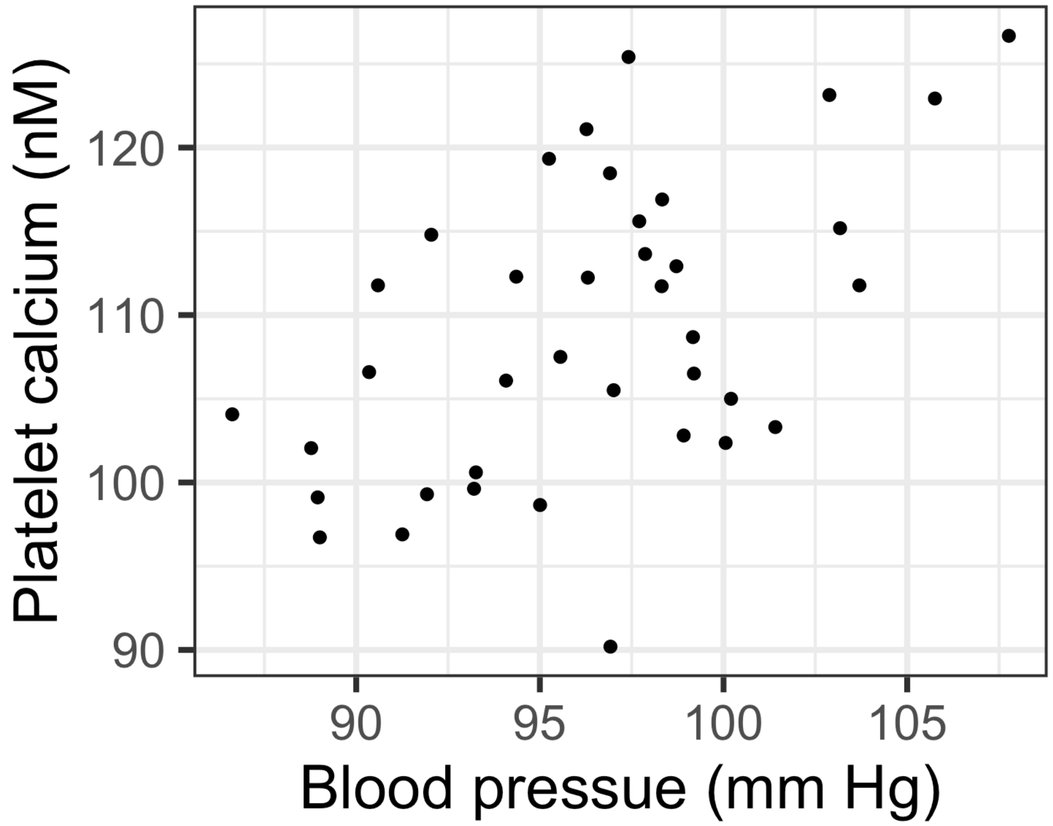
Reject the null hypothesis

Testing the hypothesis $H_0 : \rho = 0$

2. *t*-test

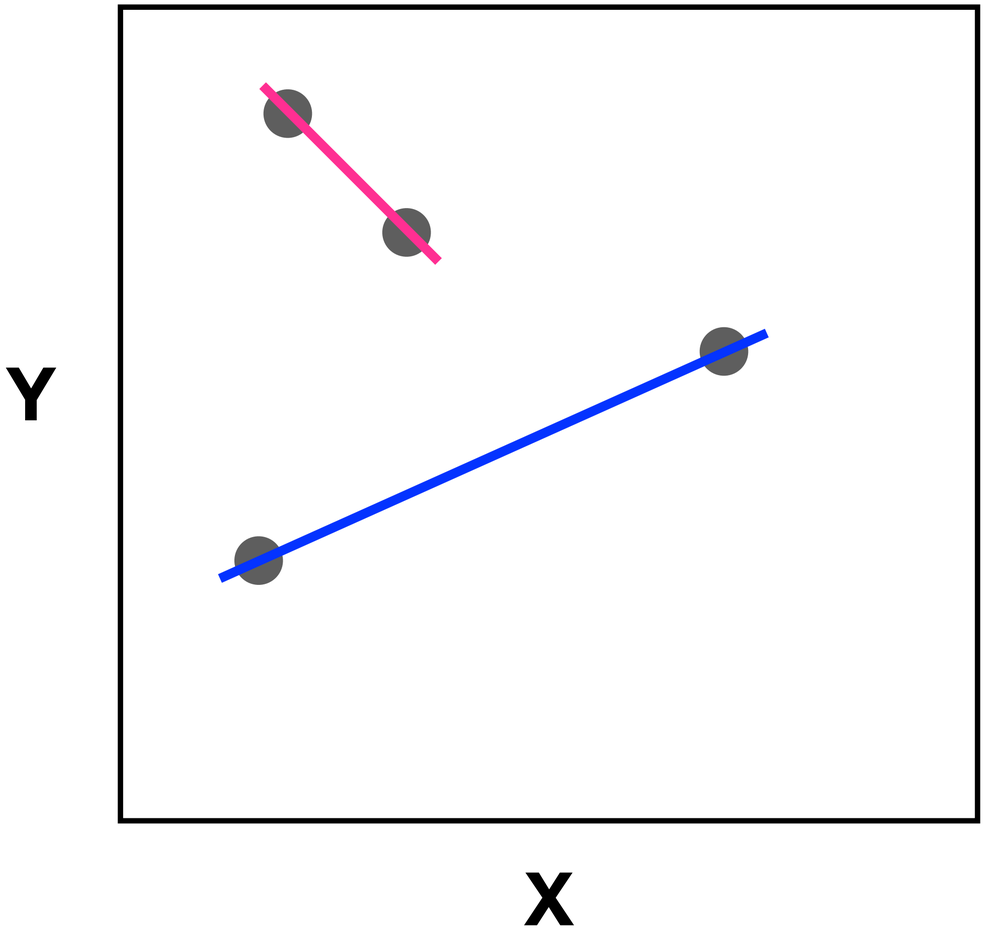
```
> cor.test(x, y)
```

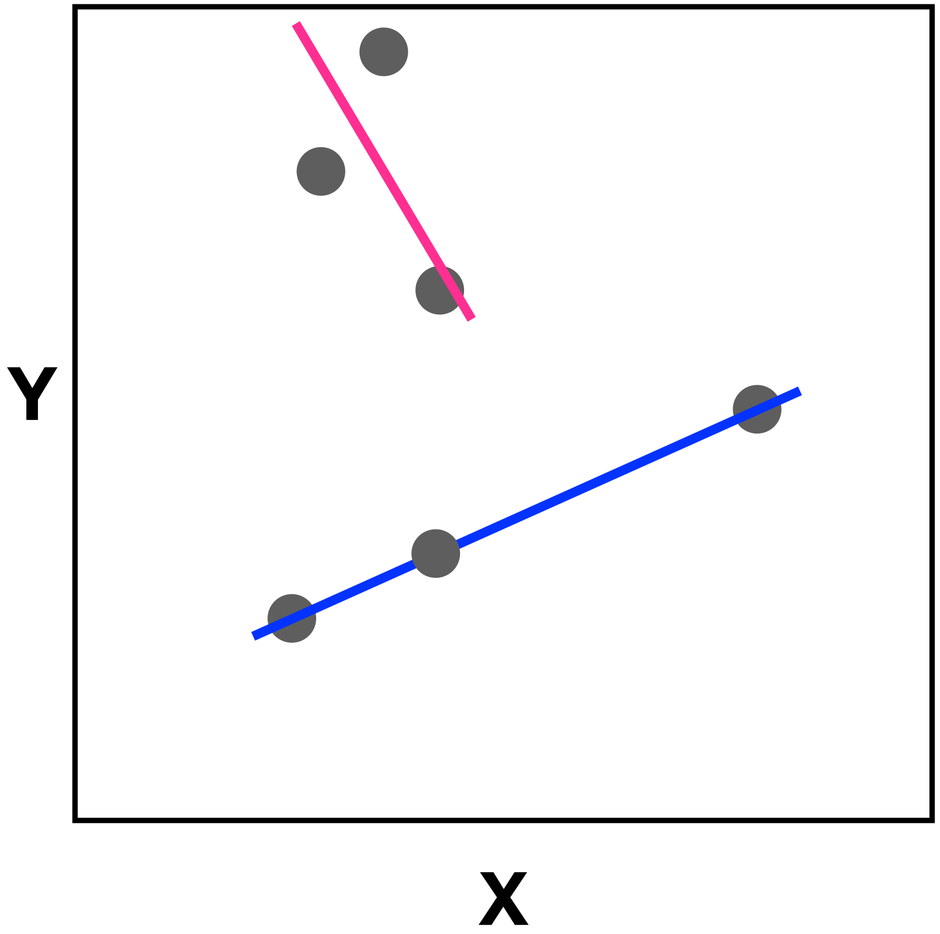
```
        Pearson's product-moment correlation

data:  df4$x and df4$y
t = 3.7406, df = 36, p-value = 0.000638
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.2519726 0.7259479
sample estimates:
      cor
0.529041
```



n = 38 **r = 0.529**

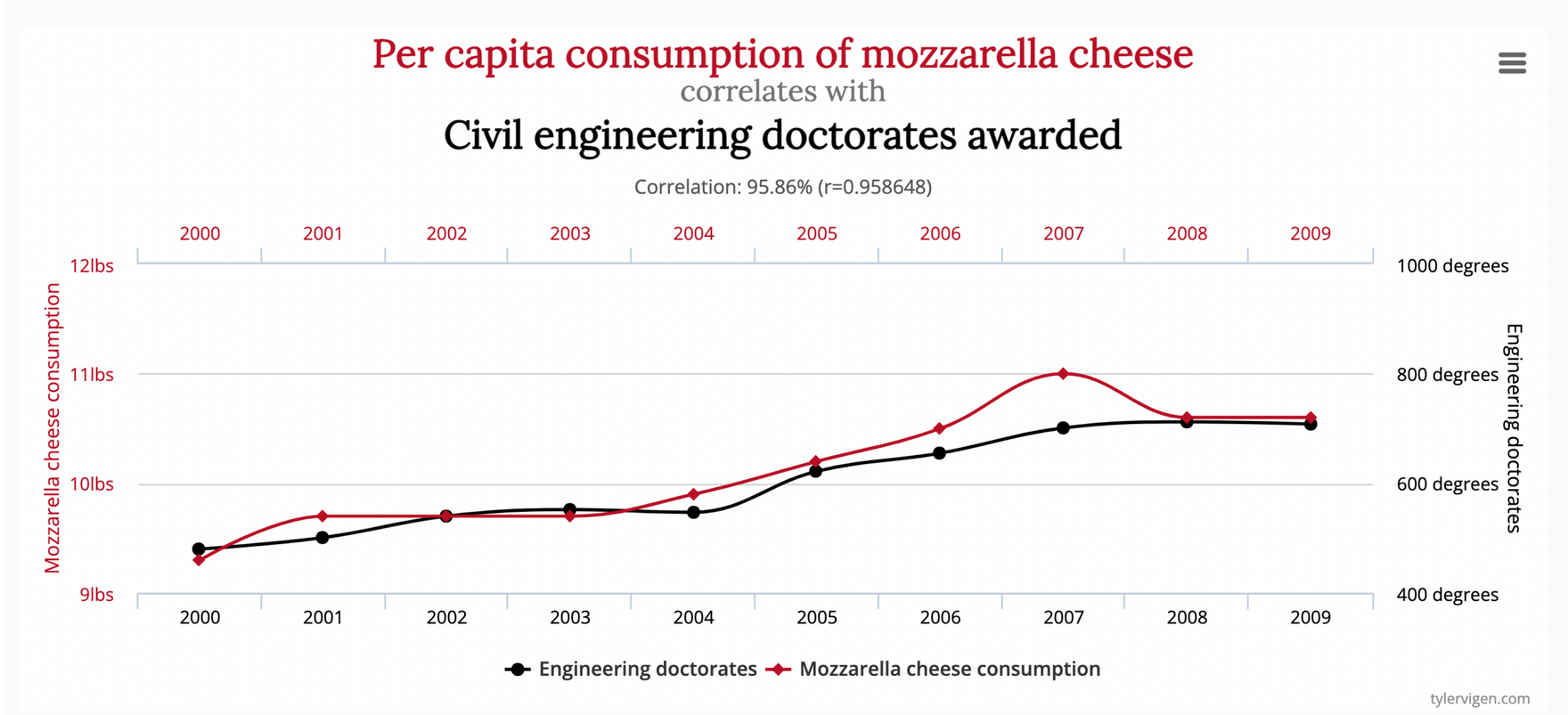
Degrees of freedom: **(n - 2)**





Correlation does not equal causation

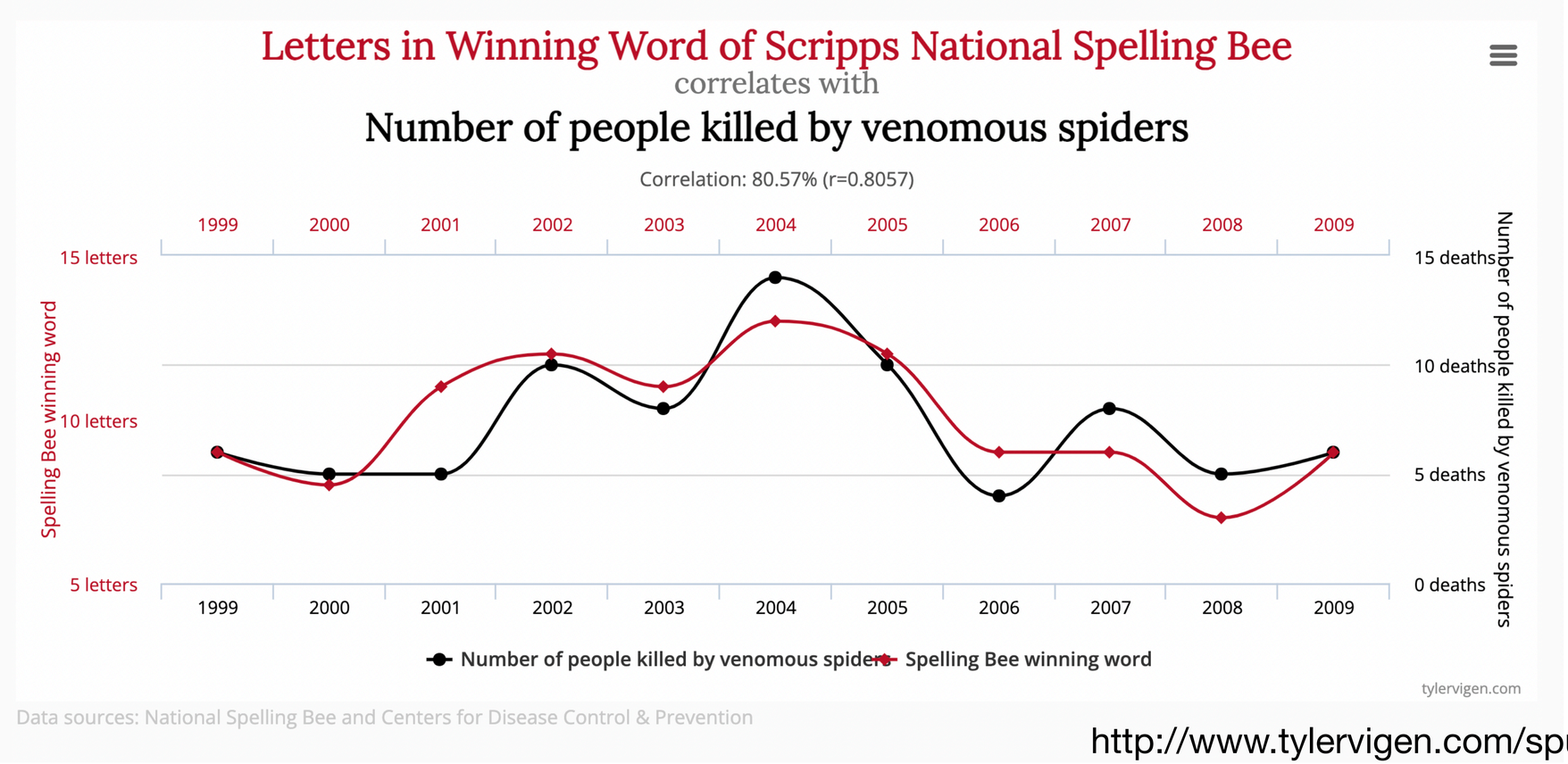
Observed association (i.e. correlation) between two variables DOES NOT indicate causal connection between them



http://www.tylervigen.com/spurious-correlations

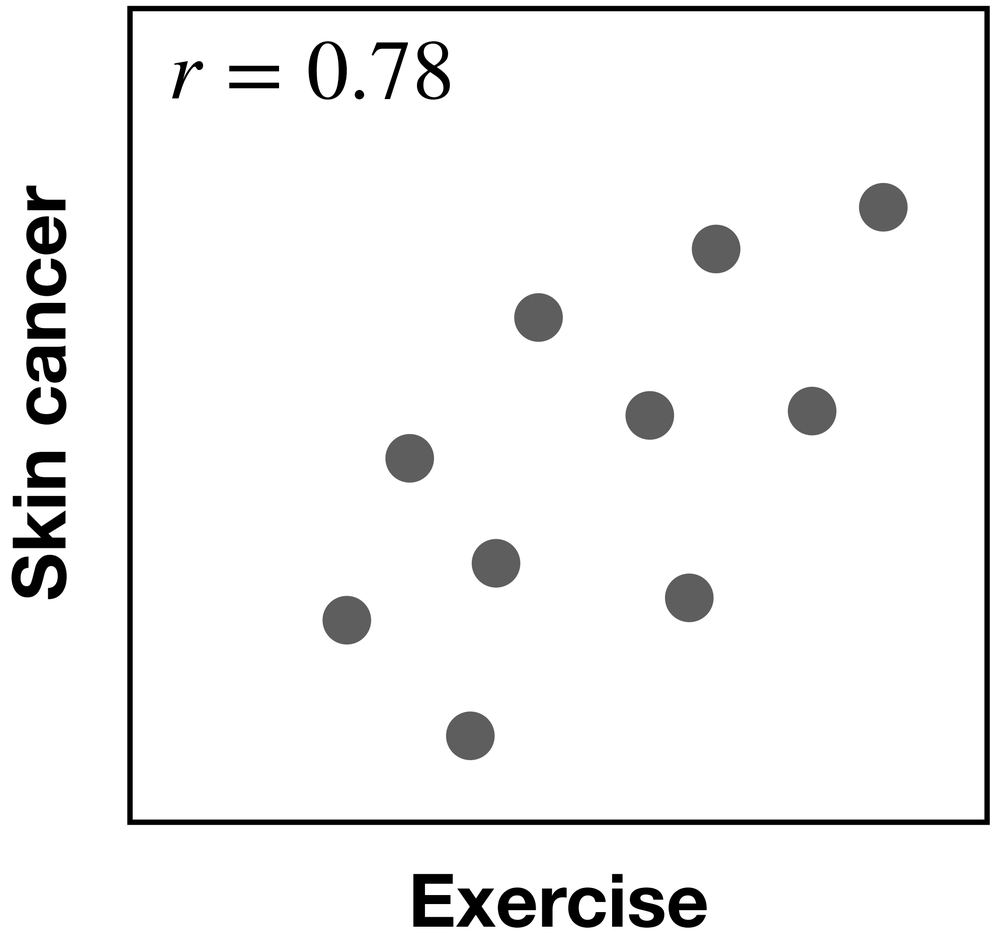
Correlation does not equal causation

Observed association (i.e. correlation) between two variables DOES NOT indicate causal connection between them



http://www.tylervigen.com/spurious-correlations

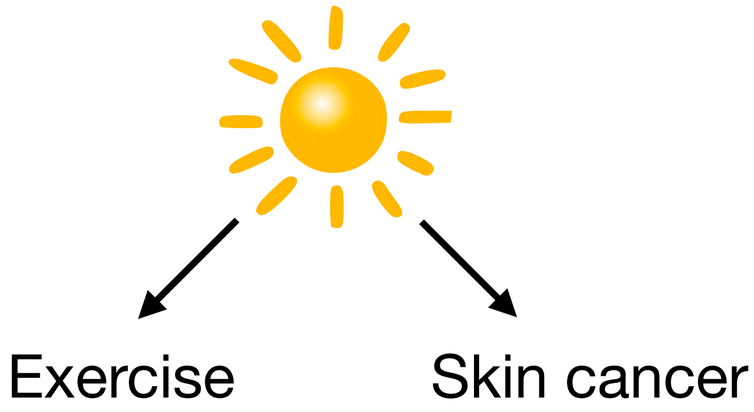
Correlation does not equal causation



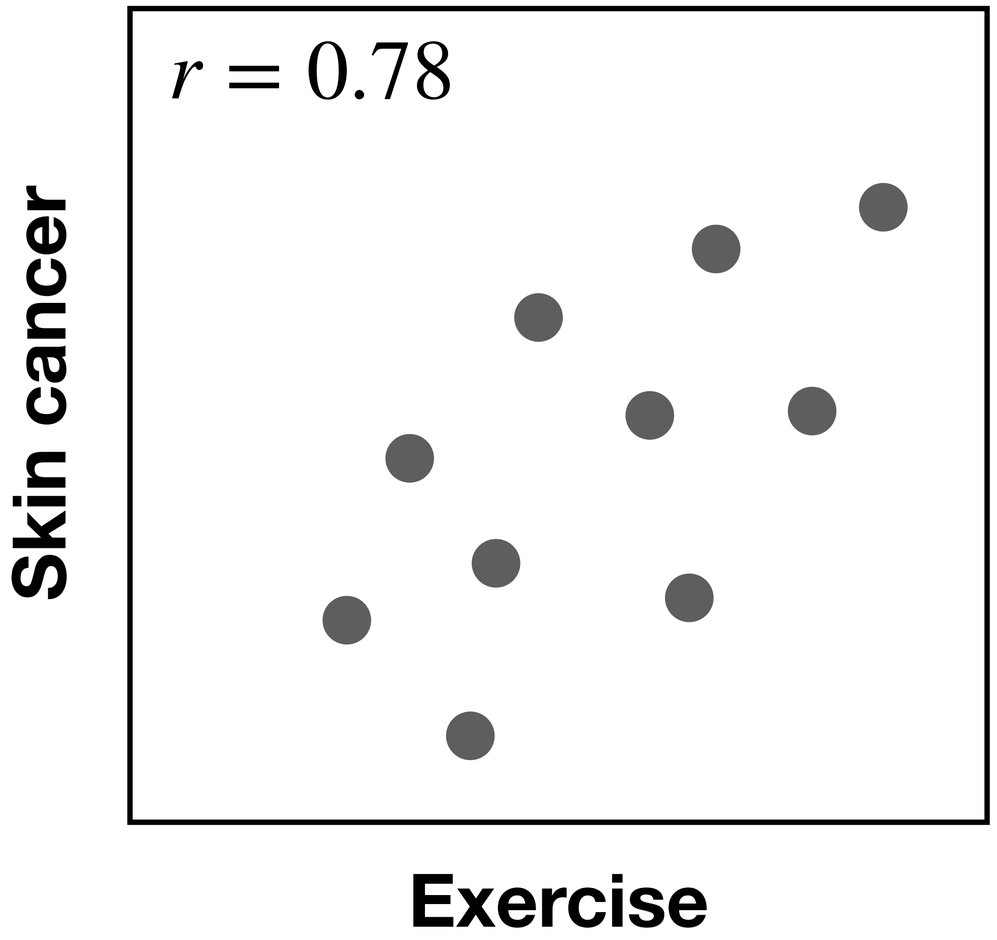
There is a strong positive correlation between exercise and skin cancer.

Does exercise cause skin cancer?

1. We might be missing a variable



Correlation does not equal causation



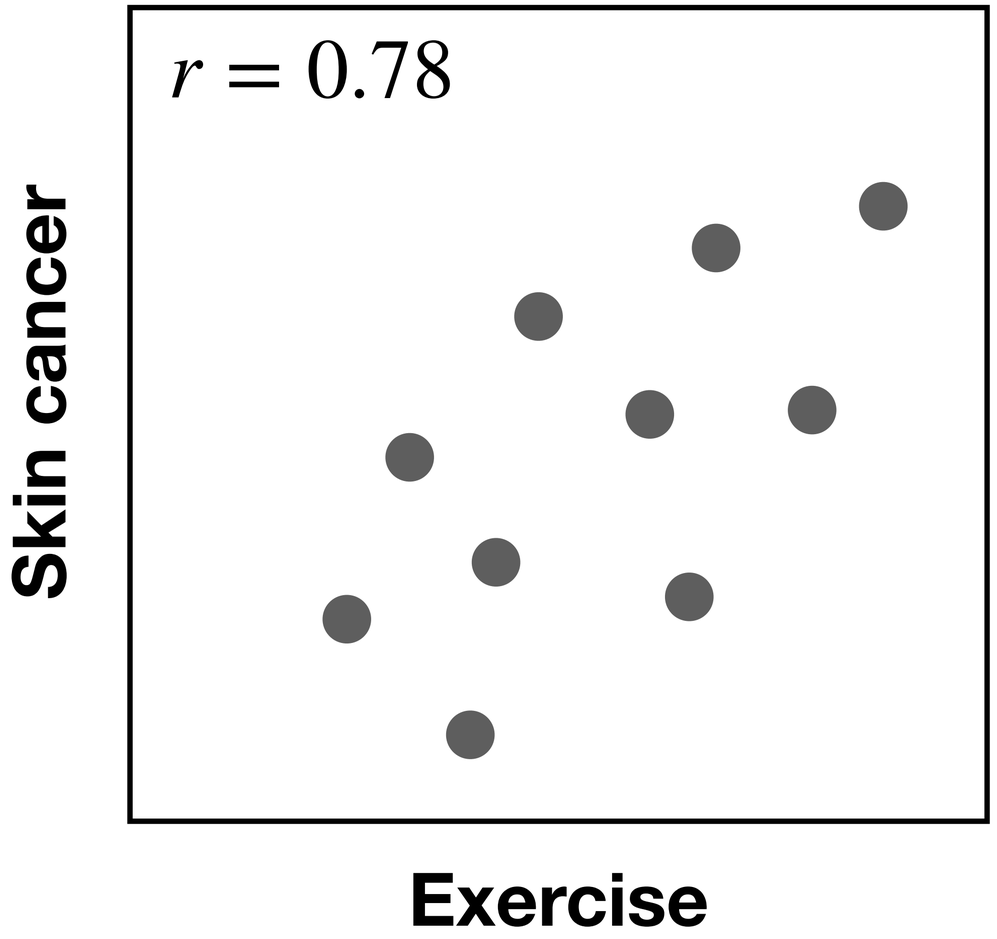
There is a strong positive correlation between exercise and skin cancer.

Does exercise cause skin cancer?

2. It might actually be reverse causality

Exercise ⟷ Skin cancer

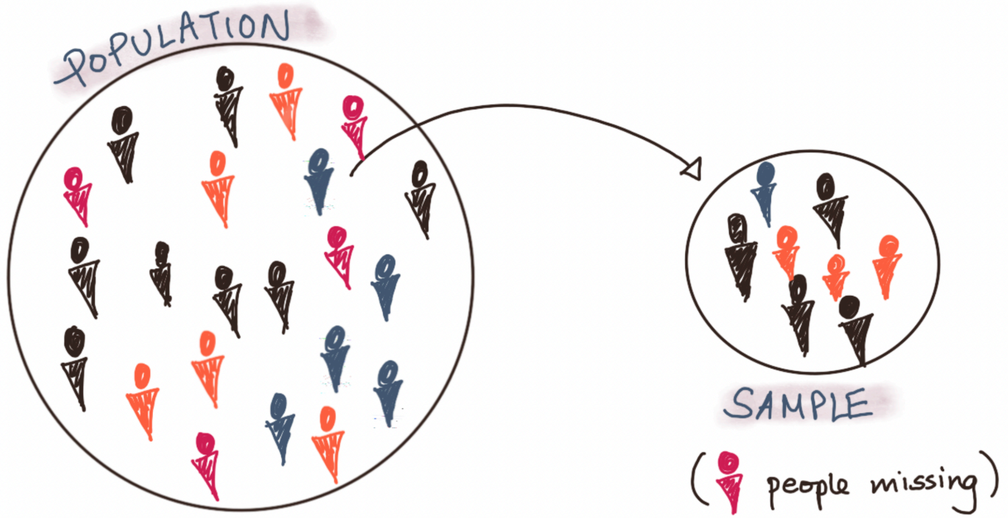
Correlation does not equal causation



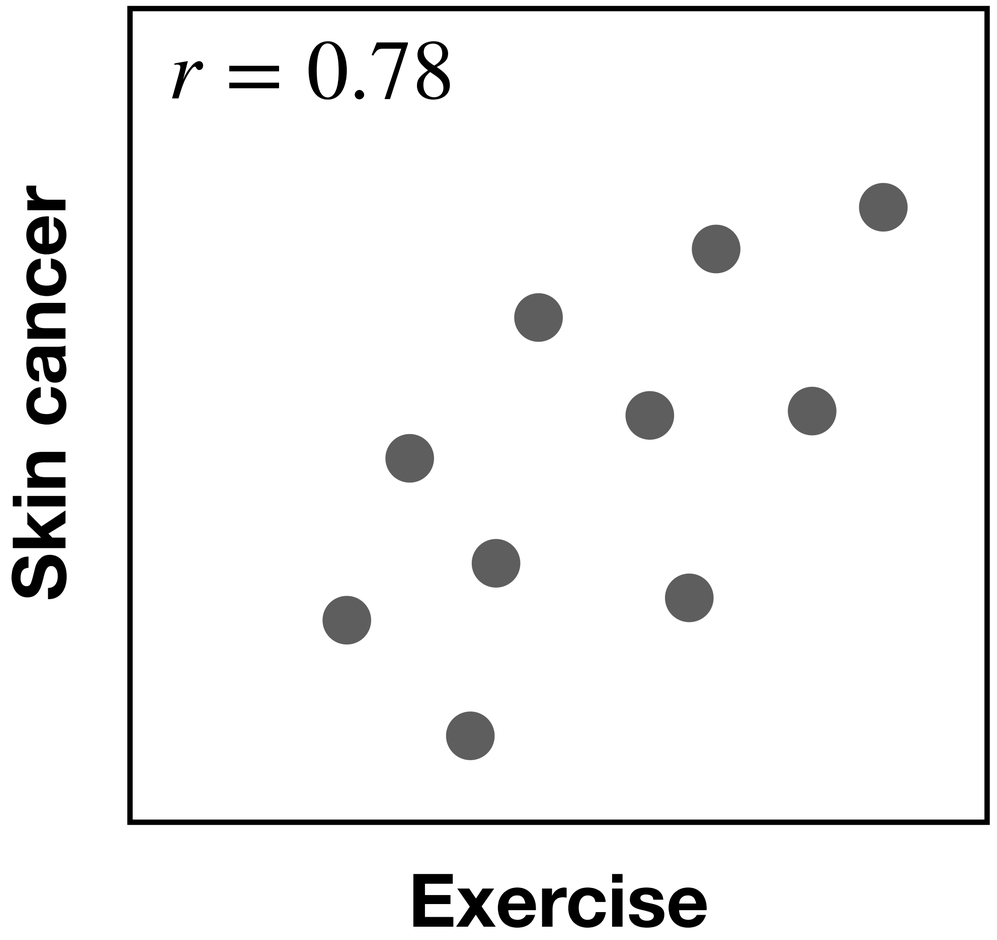
There is a strong positive correlation between exercise and skin cancer.

Does exercise cause skin cancer?

3. Could be influenced by our sample



Correlation does not equal causation



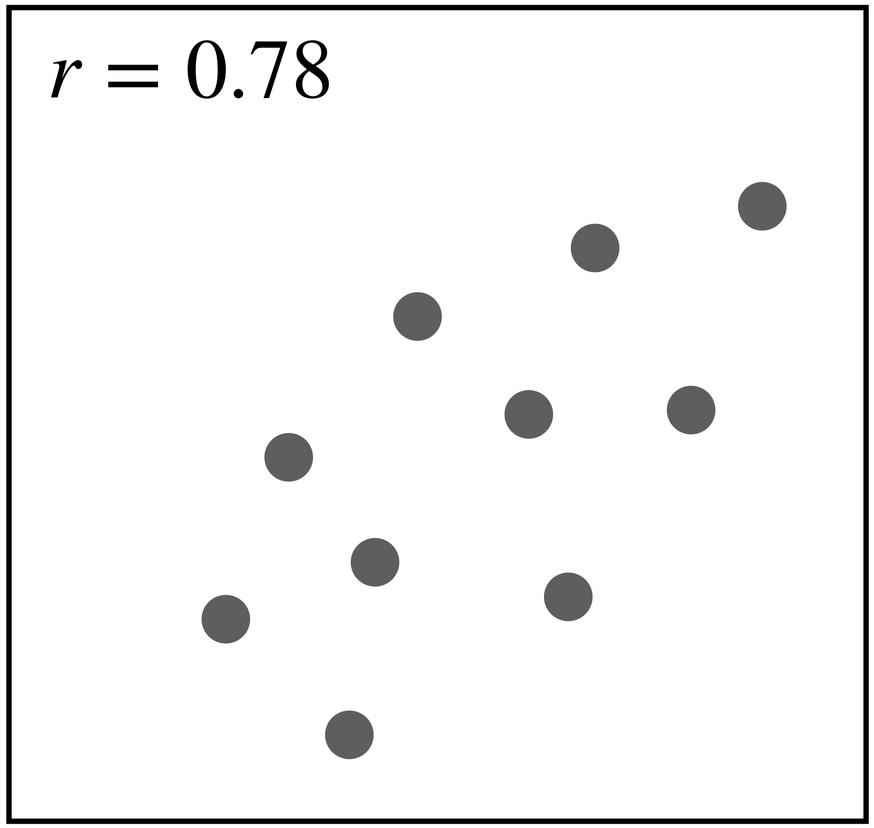
There is a strong positive correlation between exercise and skin cancer.

Does exercise cause skin cancer?

4. Difficult to measure variables

(If you were asked how often do you exercise… would your answer be reliable??)

Correlation does not equal causation



There is a strong positive correlation between exercise and skin cancer.

Does exercise cause skin cancer?

1. We might be missing a variable

2. It might actually be reverse causality

3. Could be influenced by our sample

4. Difficult to measure variables

Cannot prove causation with an observational study…

More on interpreting correlations

Just as with the t-test, a "significant" correlation might actually be a "weak" relationship…

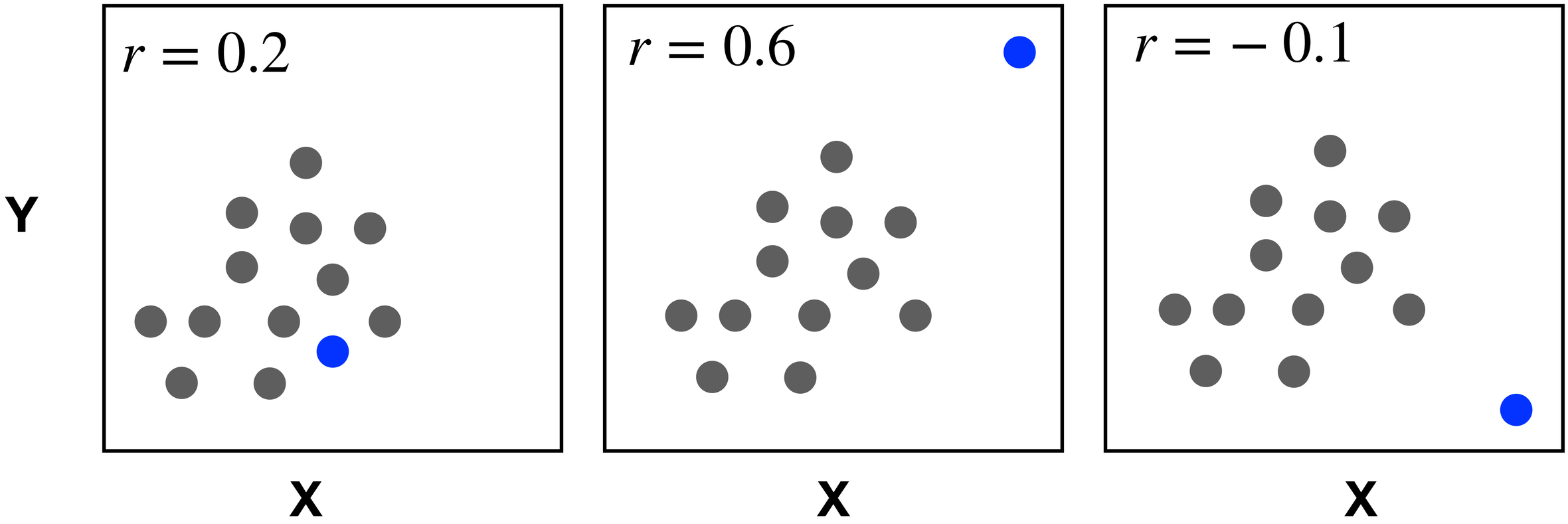
$$t_s = r\sqrt{\frac{n-2}{1-r^2}}$$

For a fixed correlation coefficient (r), test statistic increases with sample size (n).

(If your sample size is large enough, you could be "significantly" different from zero even if your correlation coefficient was 0.1)

More on interpreting correlations

The correlation coefficient is highly sensitive to extreme points



More on interpreting correlations

Q: Is correlation parametric or non-parametric?

A: The type of correlation we covered today (also known as Pearson's correlation) is parametric—it is calculating difference from the sample mean

Q: Is there a non-parametric alternative?

A: Yes! Most commonly, Spearman's correlation (another option: Kendall's correlation) uses rank order to correlate two samples

```
> cor(x, y, method = "spearman")
```

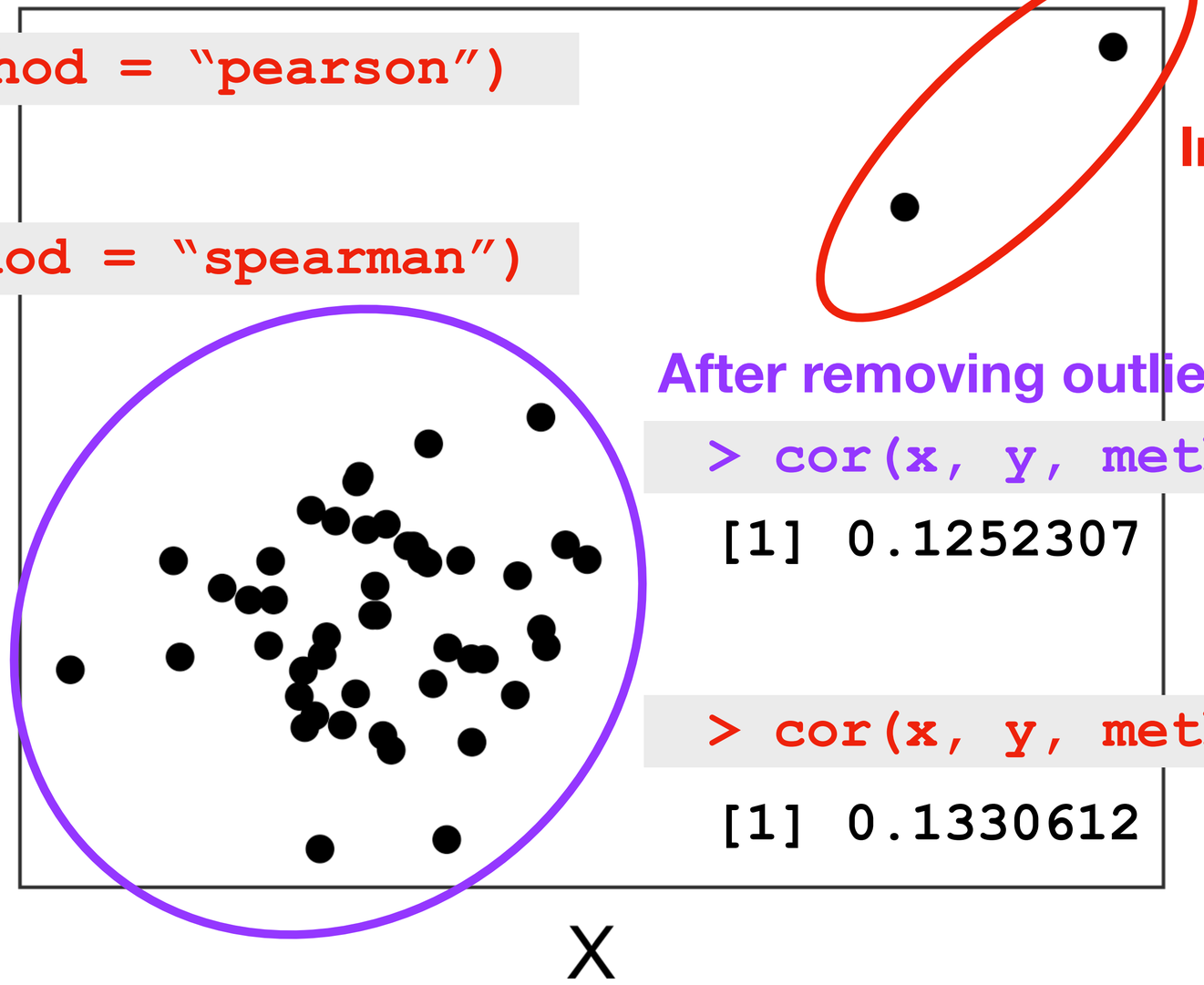
```
> cor.test(x, y, method = "spearman")
```

(Most common when one variable is ordinal (i.e. 1st, 2nd, 3rd OR strongly agree, agree, neutral, disagree, strongly disagree…)

Non-parametric correlation

```
> cor(x, y, method = "pearson")
[1] 0.6168412
```

```
> cor(x, y, method = "spearman")
[1] 0.2245858
```



Outliers/
Influential points

After removing outliers:

```
> cor(x, y, method = "pearson")
[1] 0.1252307
```

```
> cor(x, y, method = "kendall")
[1] 0.1330612
```