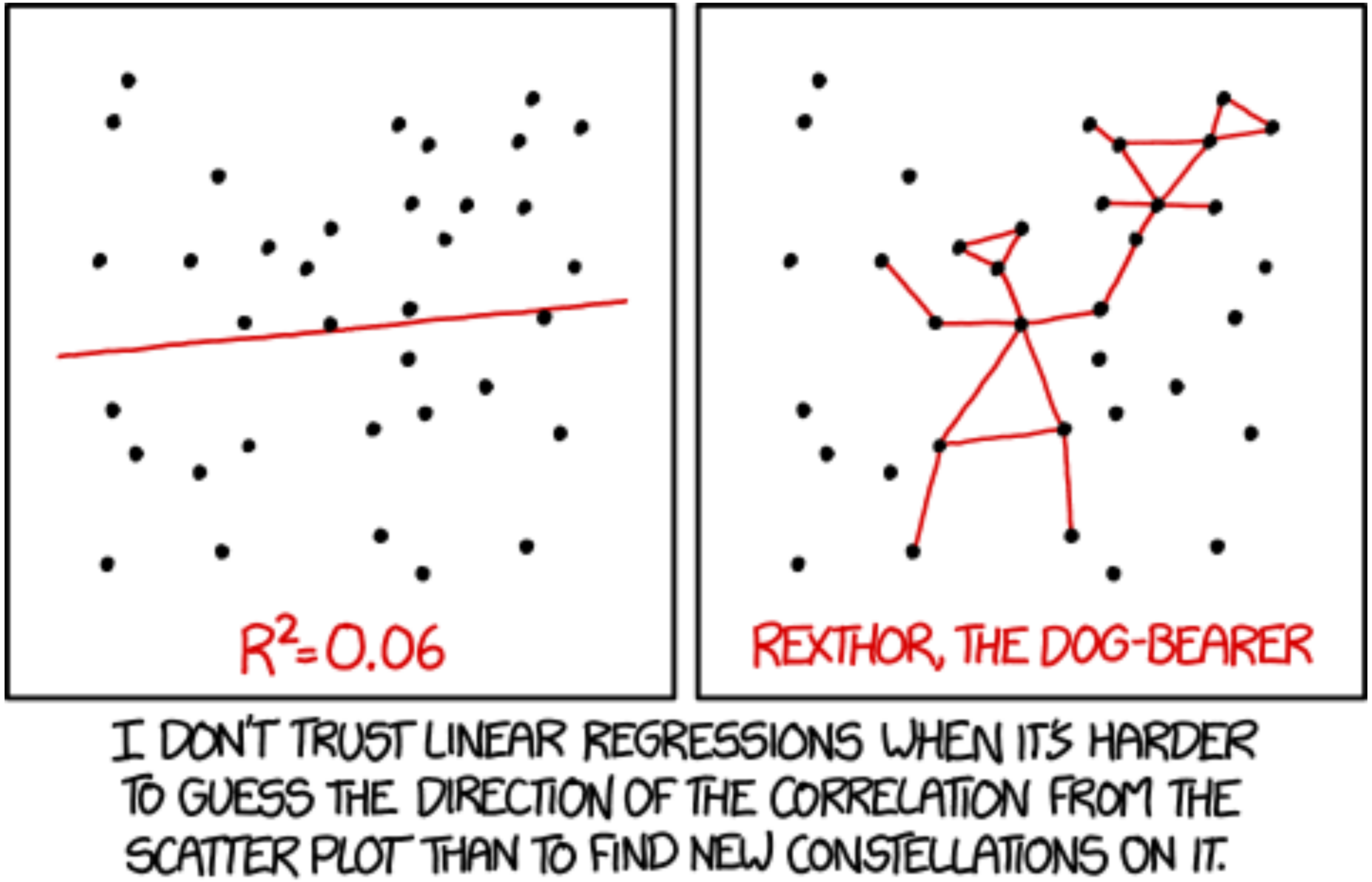
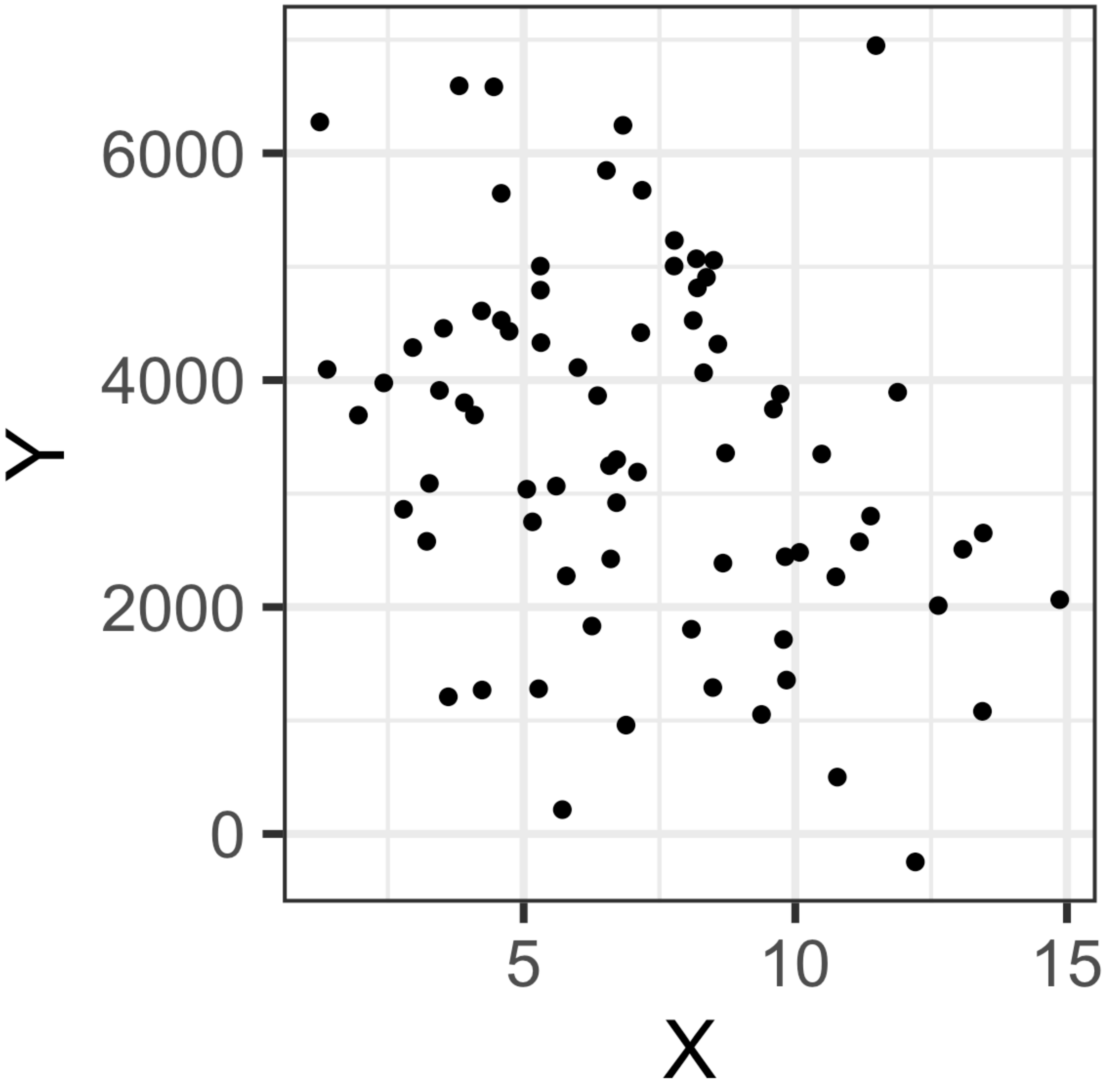


Lecture 13

(Taylor's Version)

11.16.21

Refresher Quiz

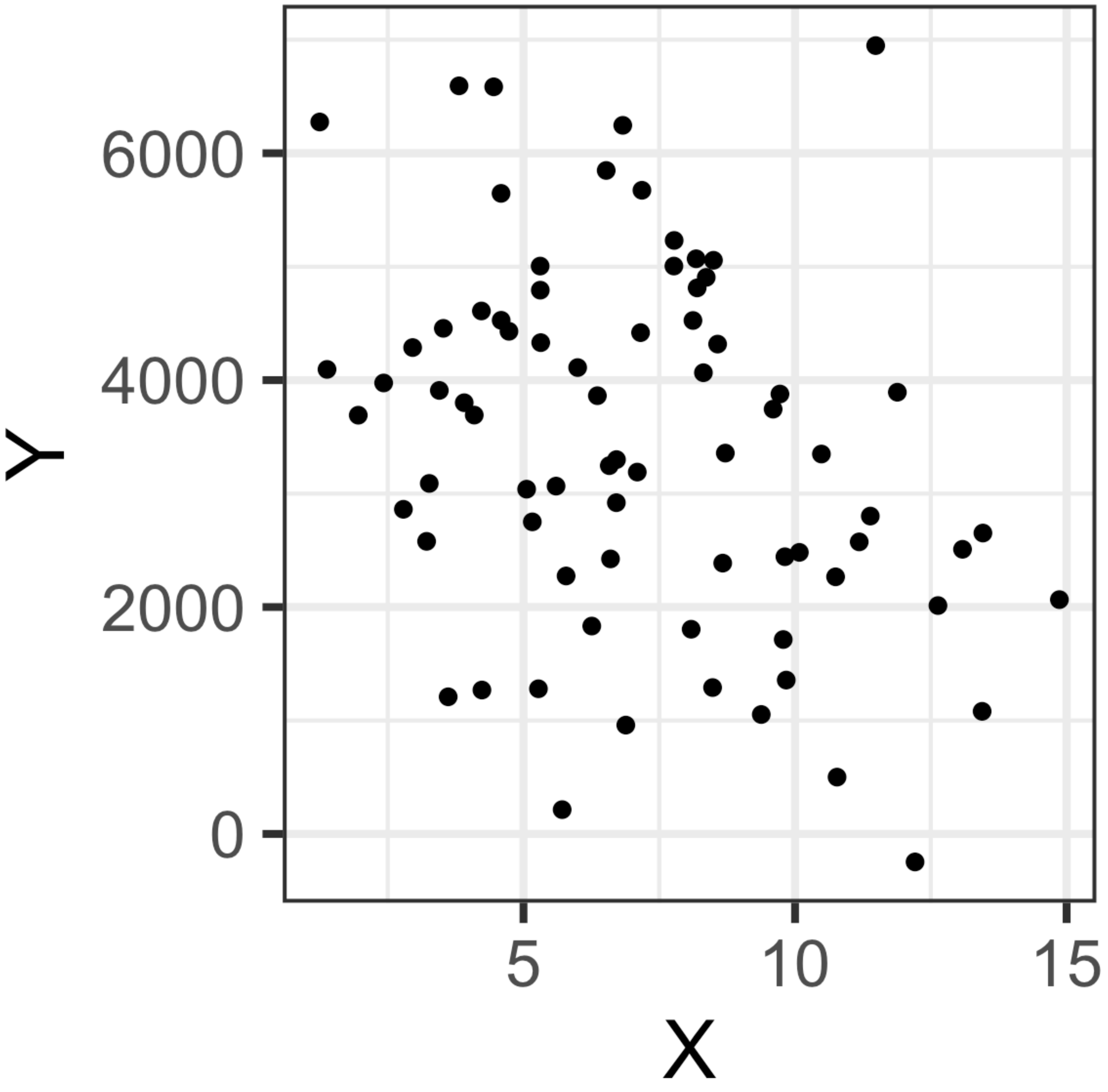


1. Examining the scatterplot on the left, does there seem to be a linear trend in the data? Is it increasing or decreasing? Is it weak or strong?

2. The correlation value is -0.321. You next test $H_0 : \rho = 0$ and get a p-value of 0.005. Explain how the evidence can be so strong even though the graph displays substantial scatter and the correlation value is far from -1.

3. Given this data, can we conclude that X affects Y?

Refresher Quiz



1. Examining the scatterplot on the left, does there seem to be a linear trend in the data? Is it increasing or decreasing? Is it weak or strong?

Looks like a weak, negative (decreasing) correlation

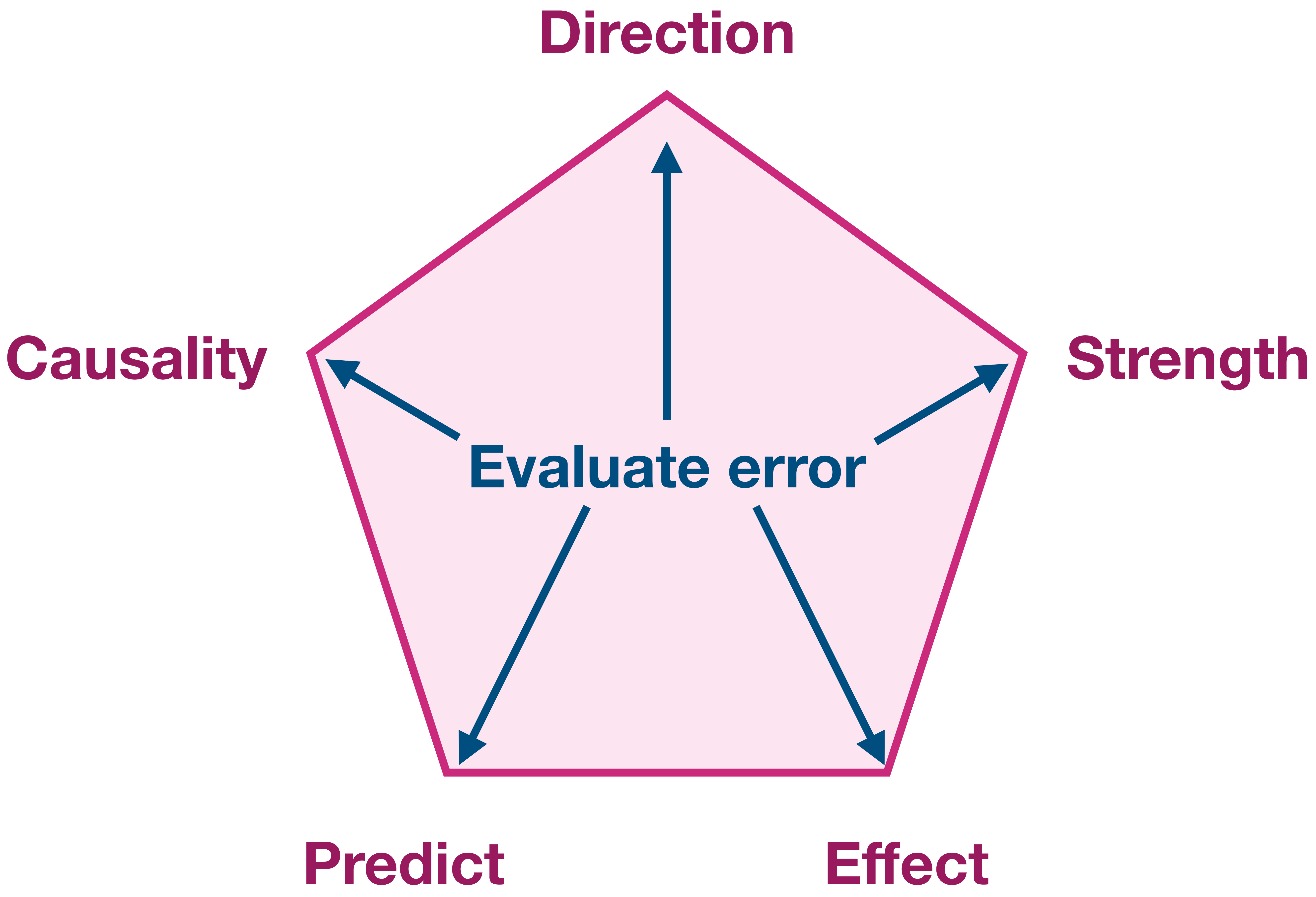
2. The correlation value is -0.321. You next test $H_0 : \rho = 0$ and get a p-value of 0.005. Explain how the evidence can be so strong even though the graph displays substantial scatter and the correlation value is far from -1.

Probably because we have a lot of samples!

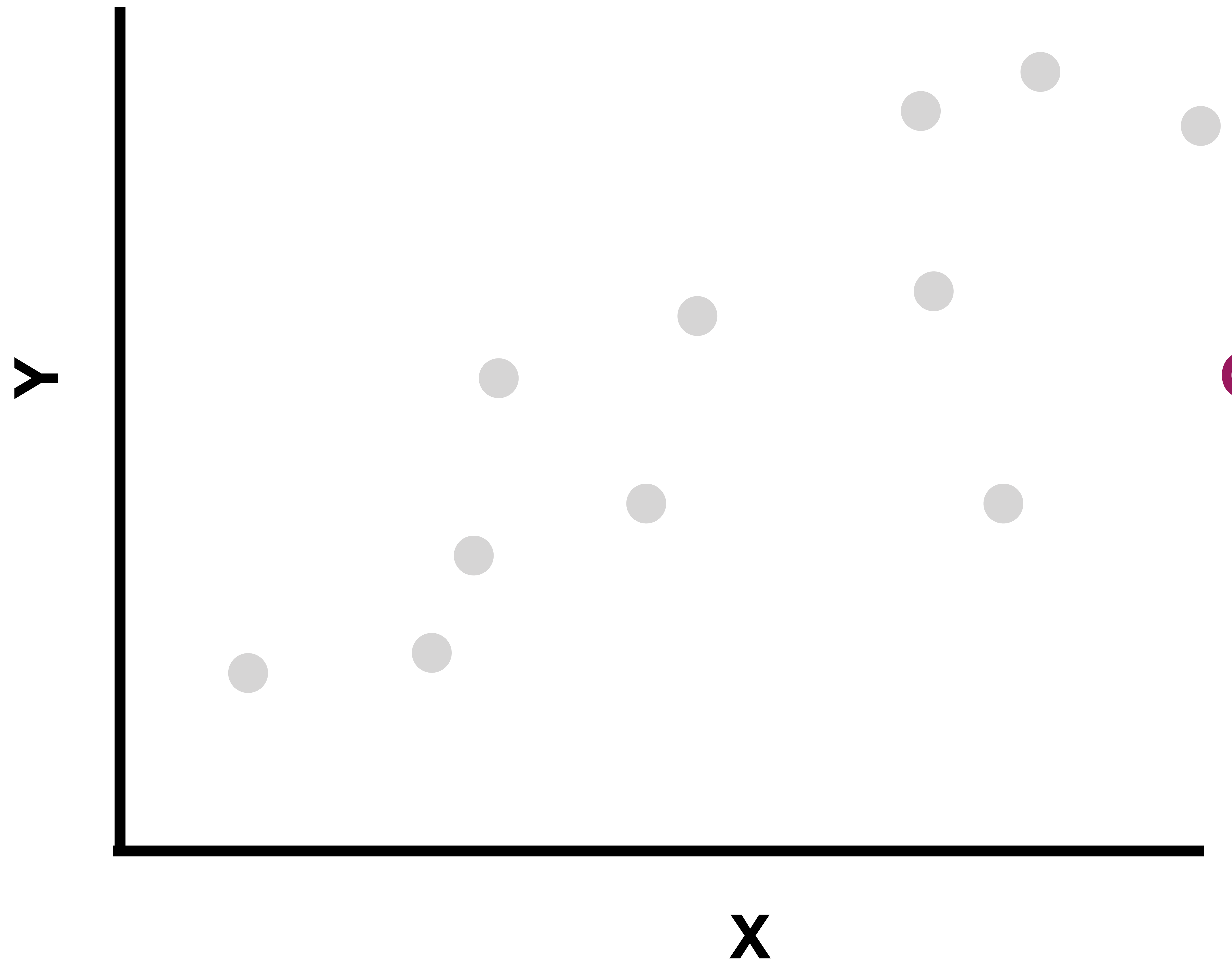
3. Given this data, can we conclude that X affects Y?

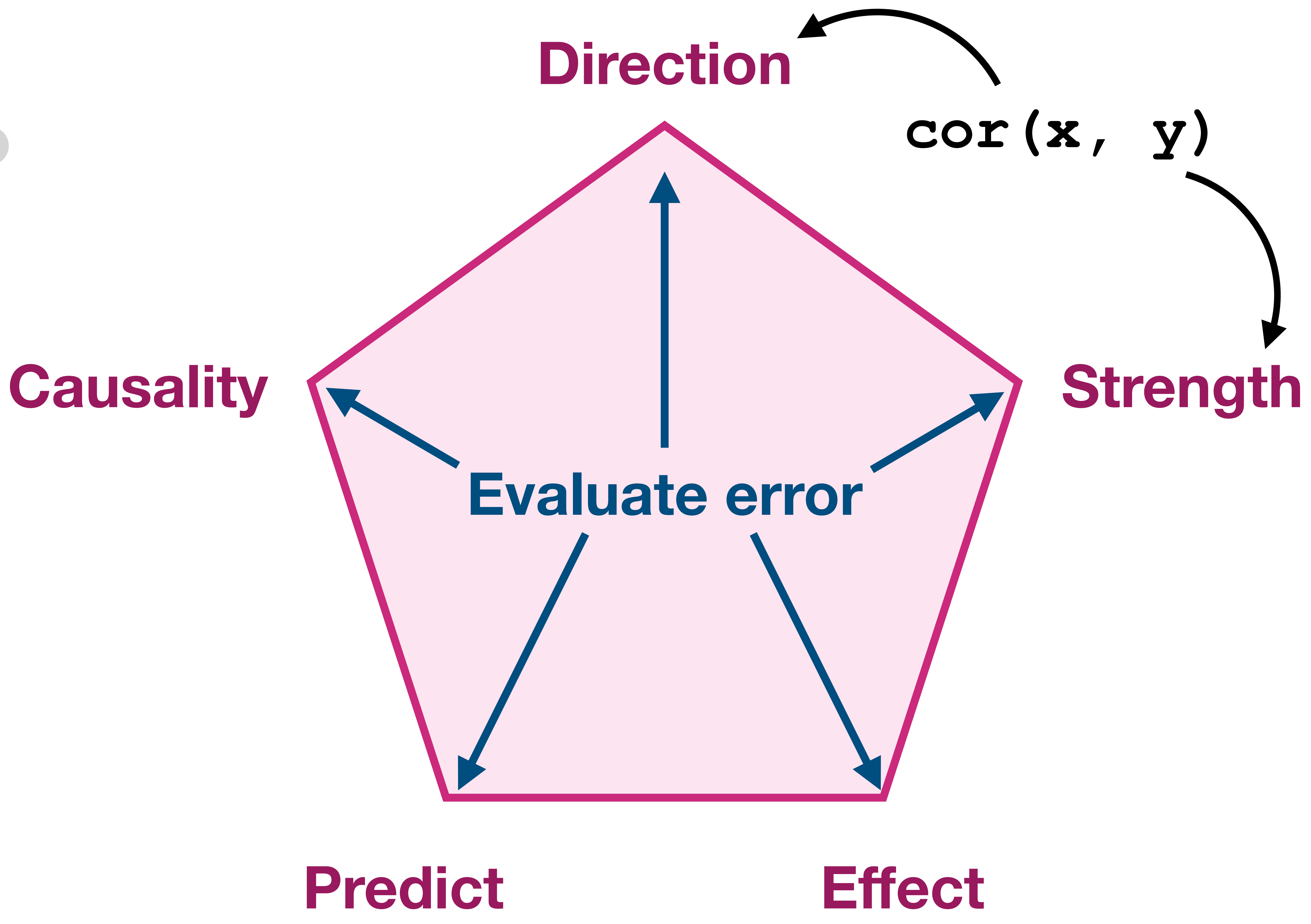
No! Correlation $\neq$ causation!

How to best define the relationship between two variables?



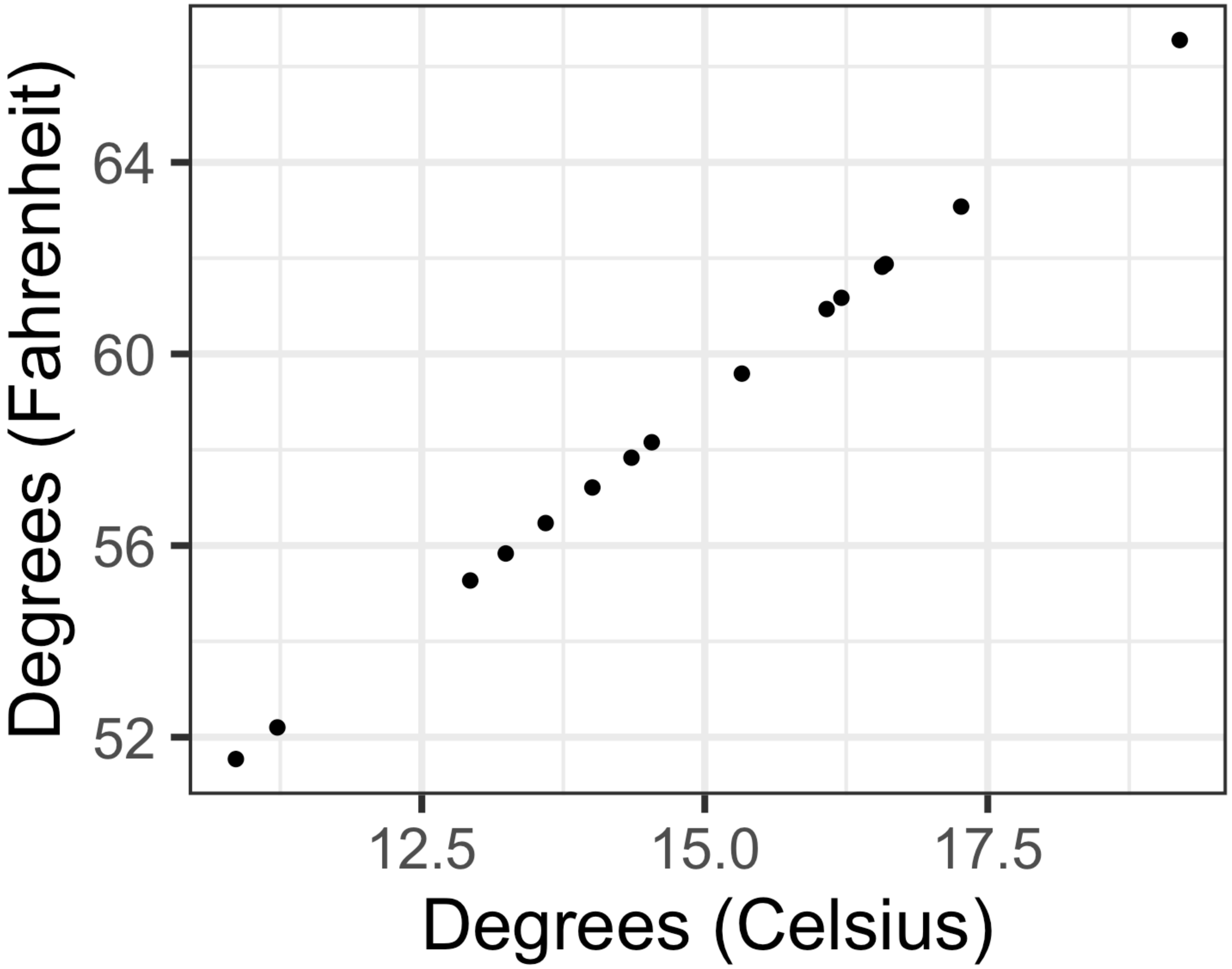
How to best define the relationship between two variables?



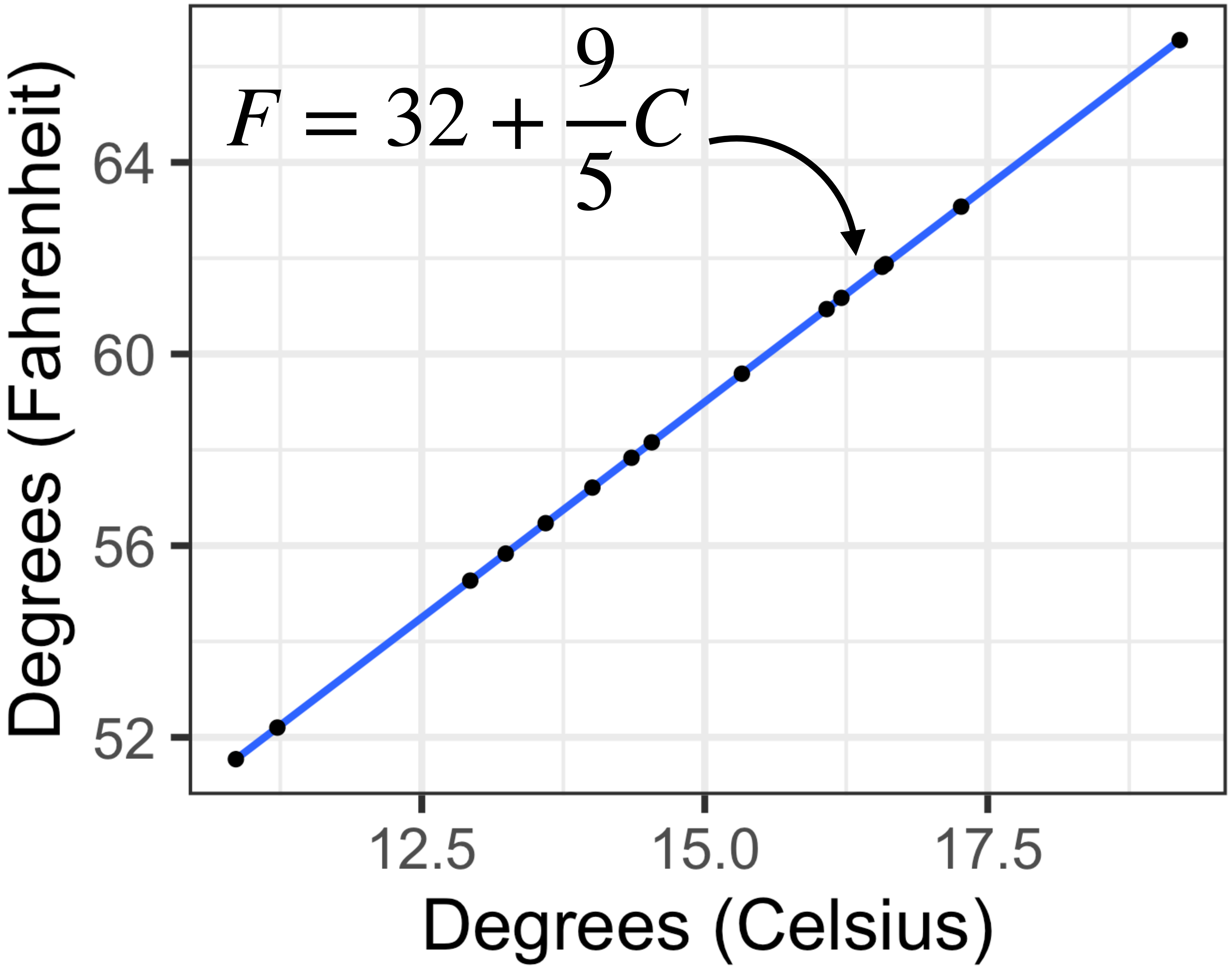


Regression models try to explain the relationship between two variables

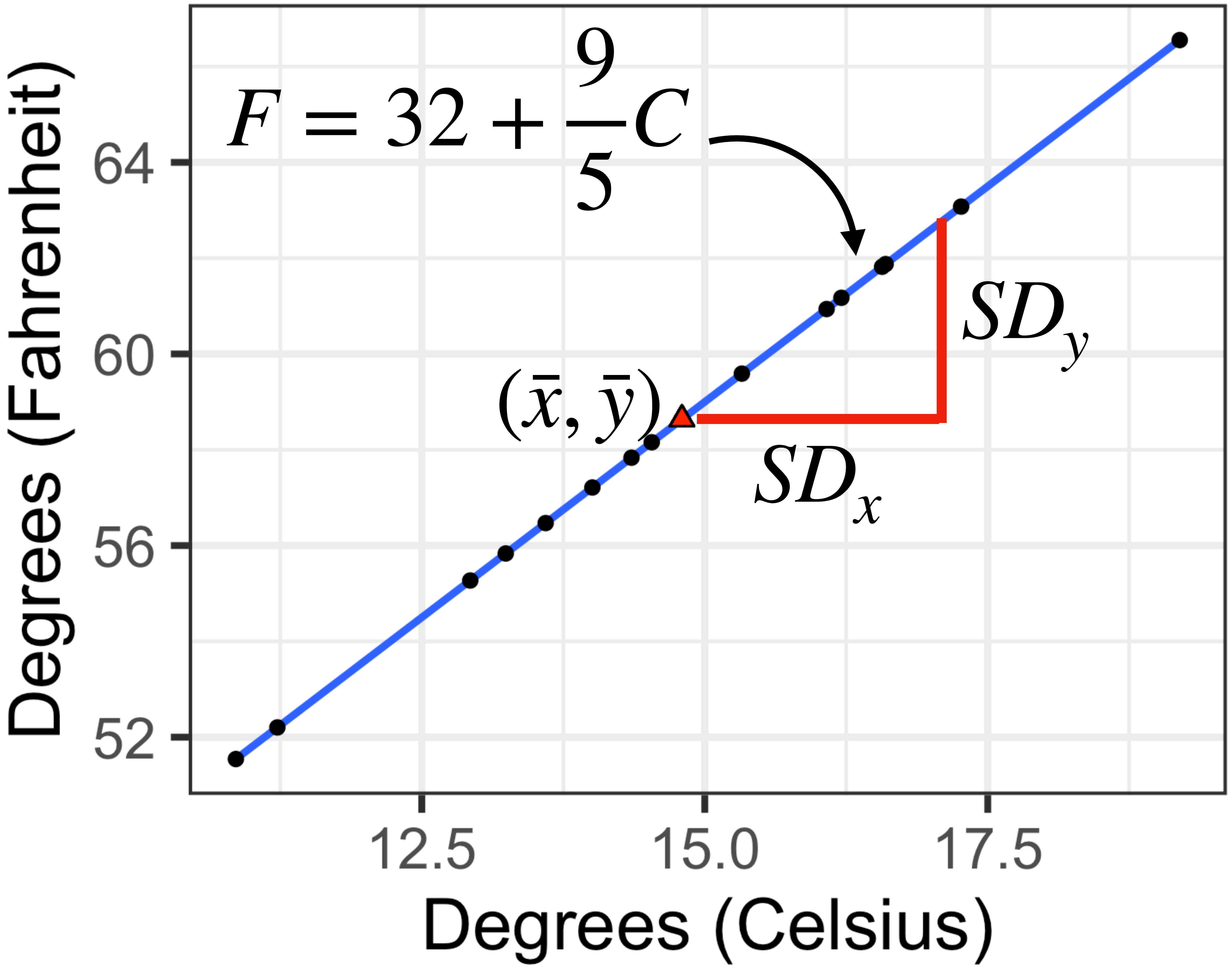
The fitted regression line



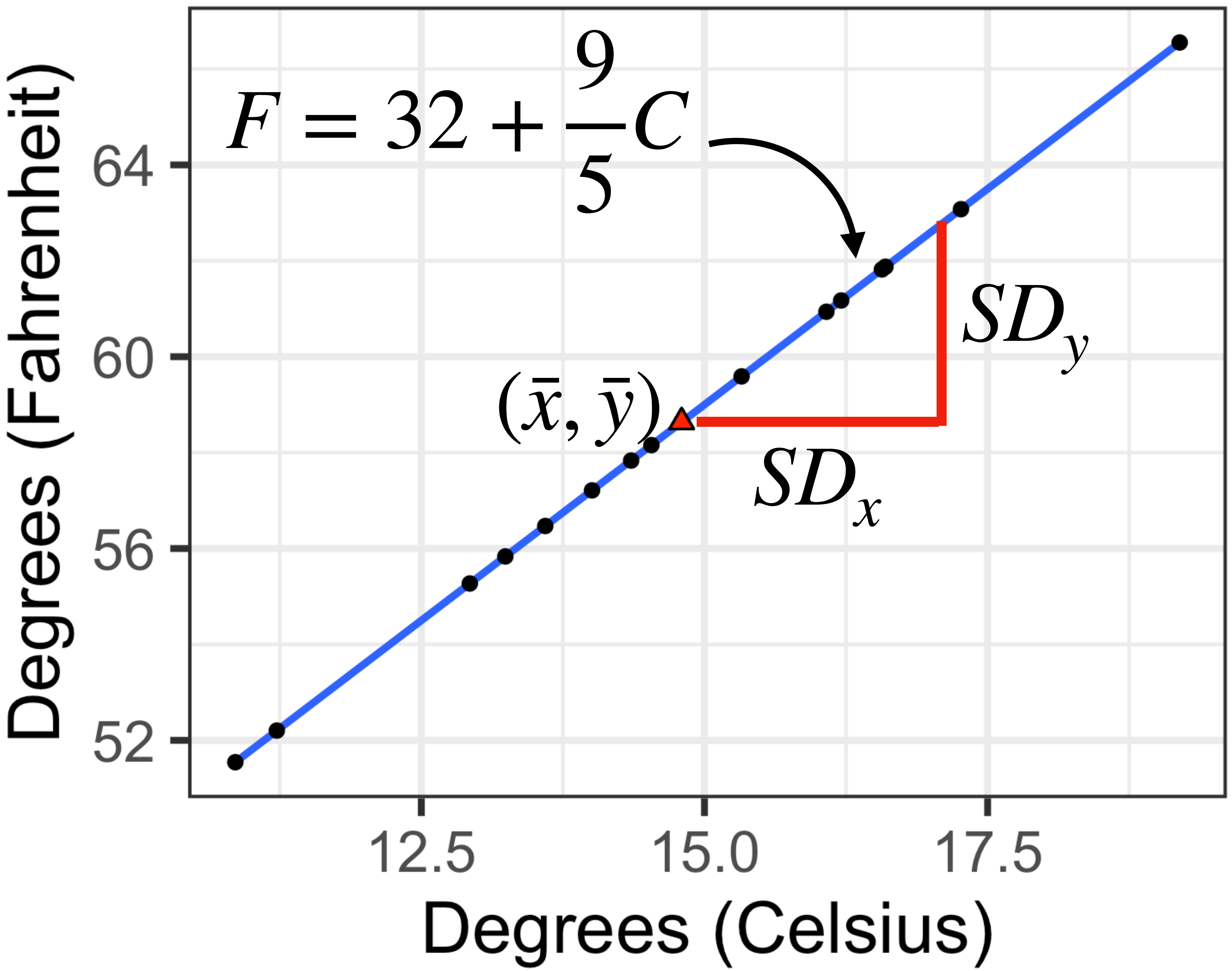
The fitted regression line



The fitted regression line



The fitted regression line

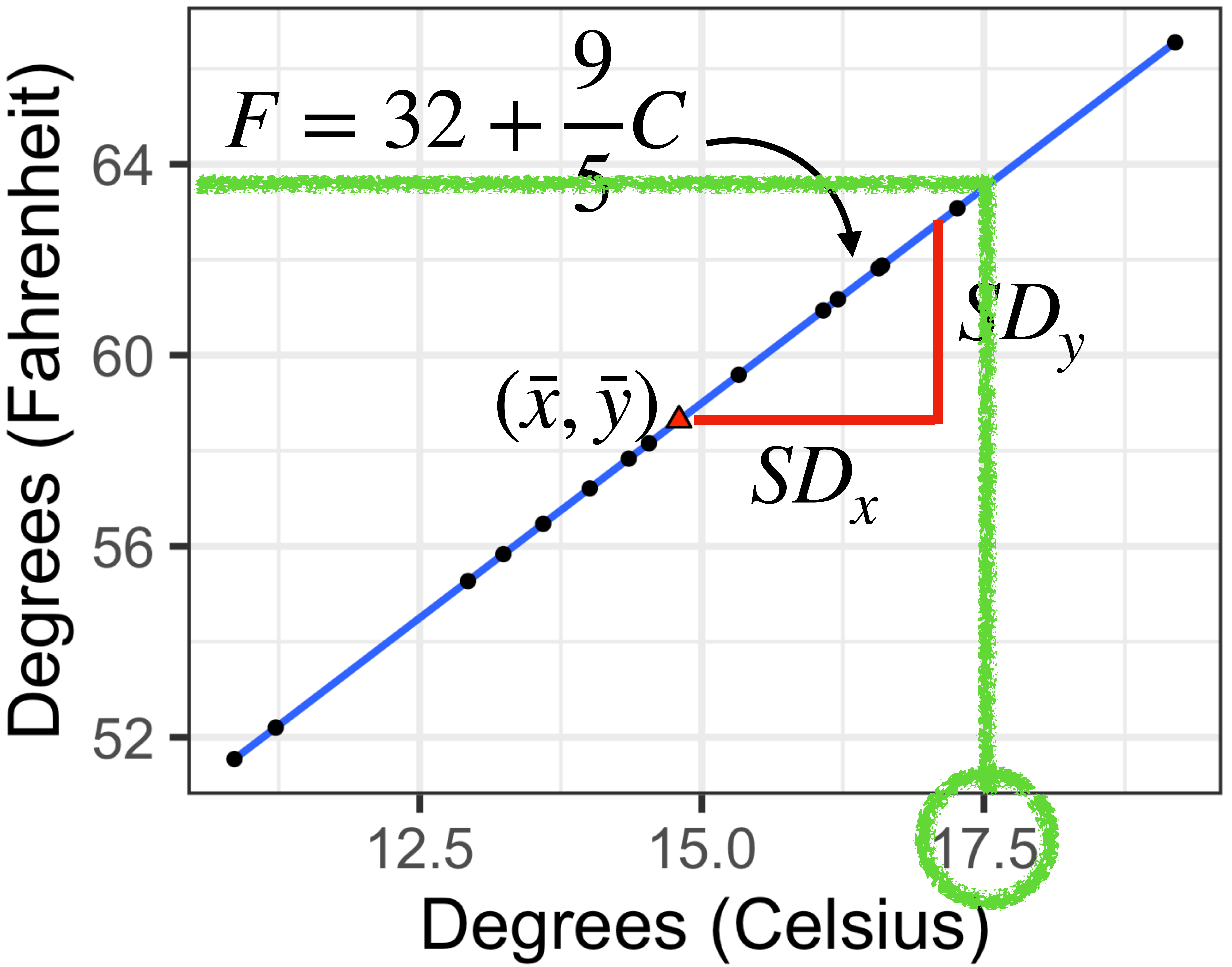


$$\frac{rise}{run} = \frac{s_y}{s_x} = \frac{2.88}{1.60} = 1.80$$

$$\frac{9}{5} = 1.80$$

If two variables have a perfect correlation ($r = \pm 1$), the slope of the line that fits the data exactly will have a slope of $\pm s_y / s_x$

Prediction using the fitted regression line



How many degrees (F) is 17.5 C?

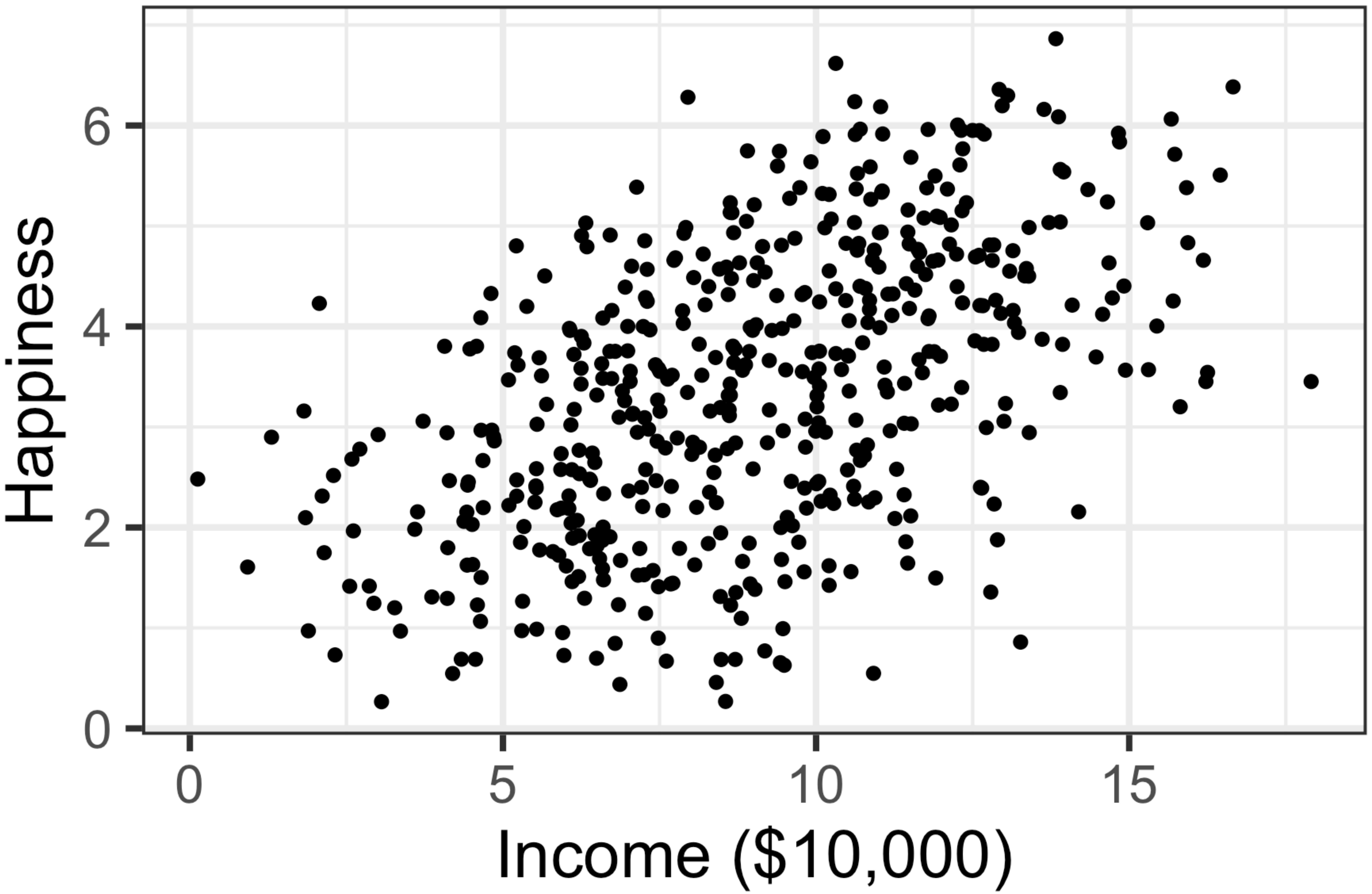
$$F = 32 + \frac{9}{5}C$$

$$F = 32 + \frac{9}{5}(17.5)$$

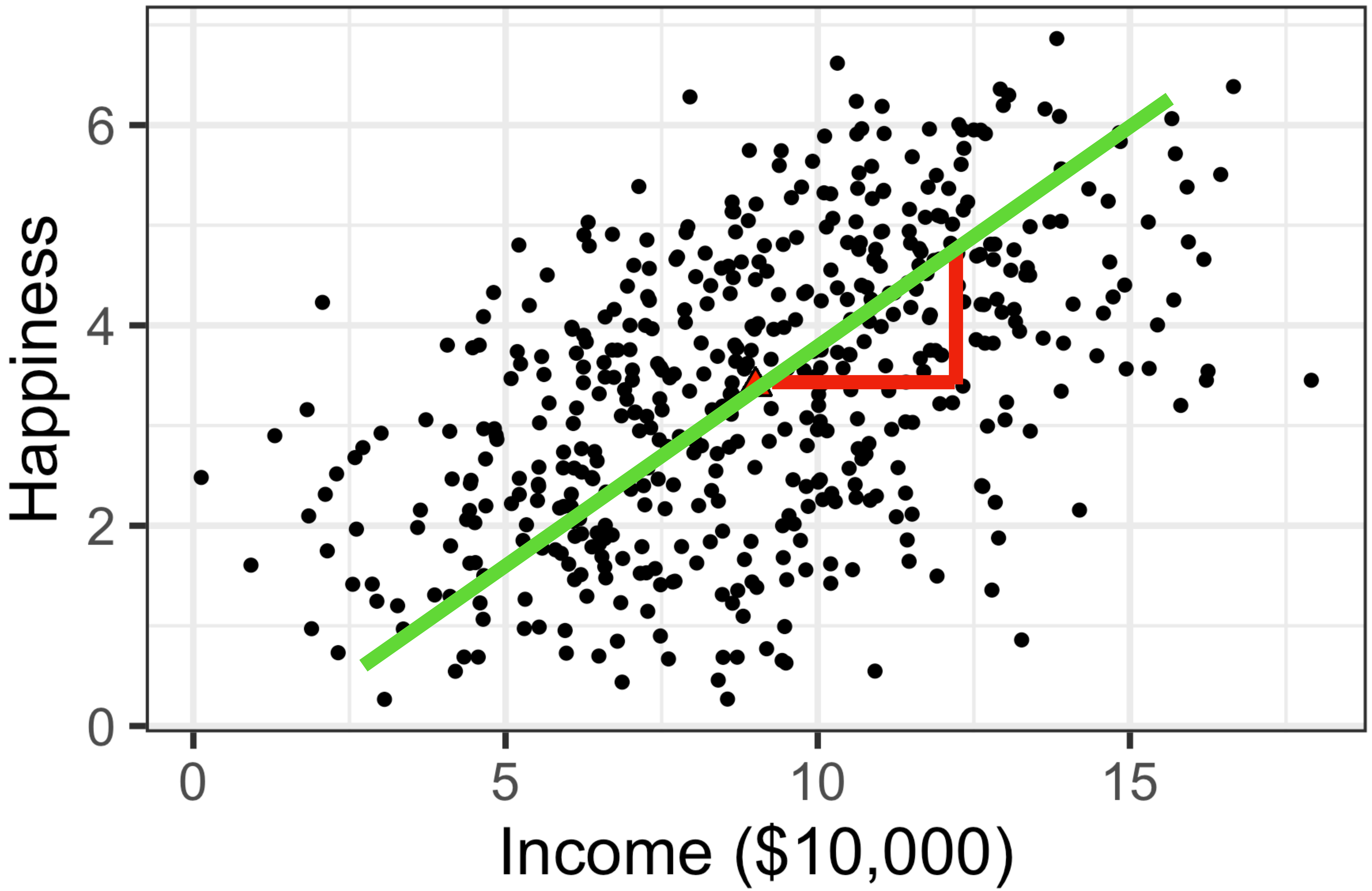
$$F = 63.5$$ ✓

What if our relationship doesn't have a perfect correlation?

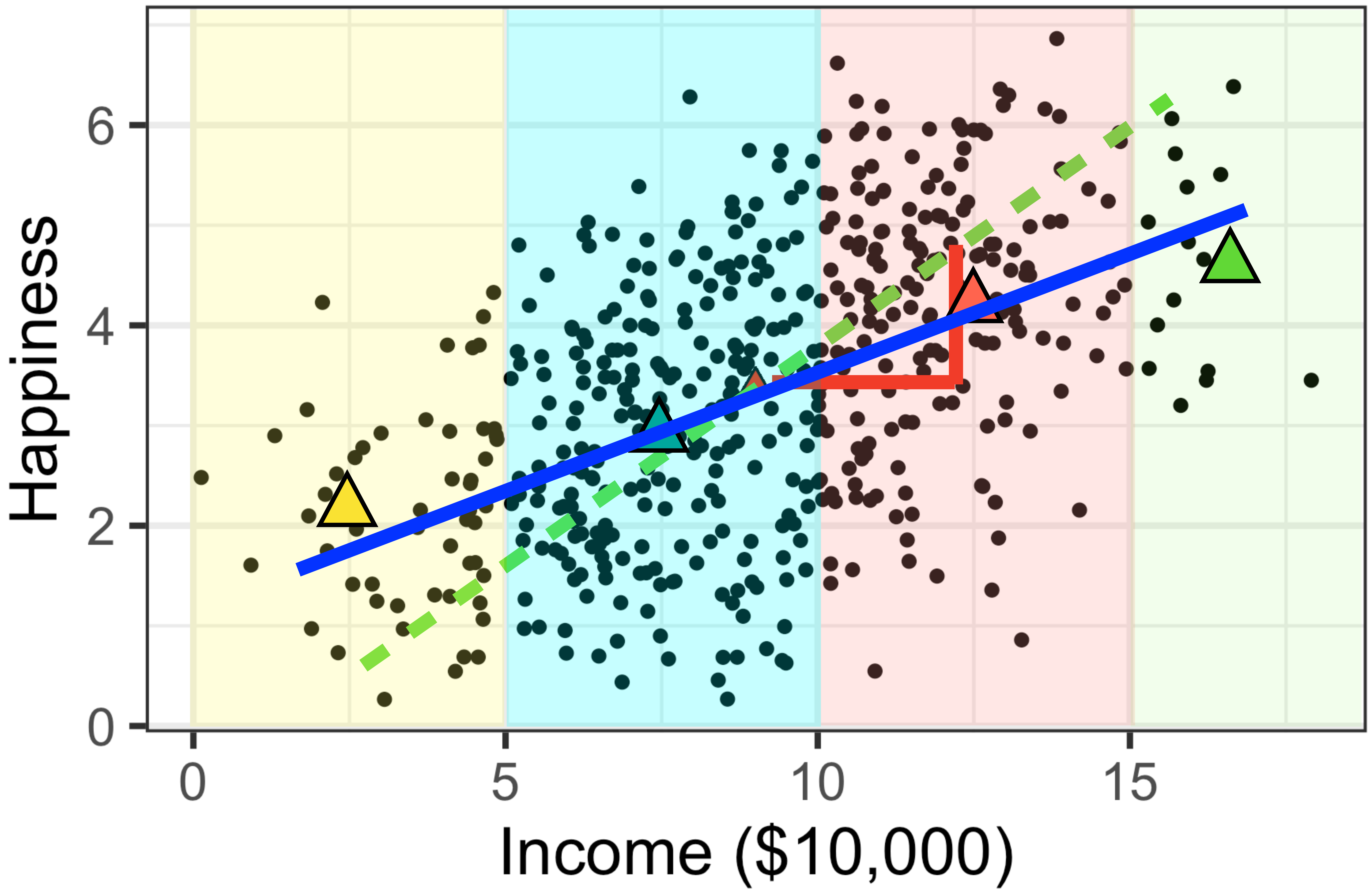
The fitted regression line: $r(s_y/s_x)$



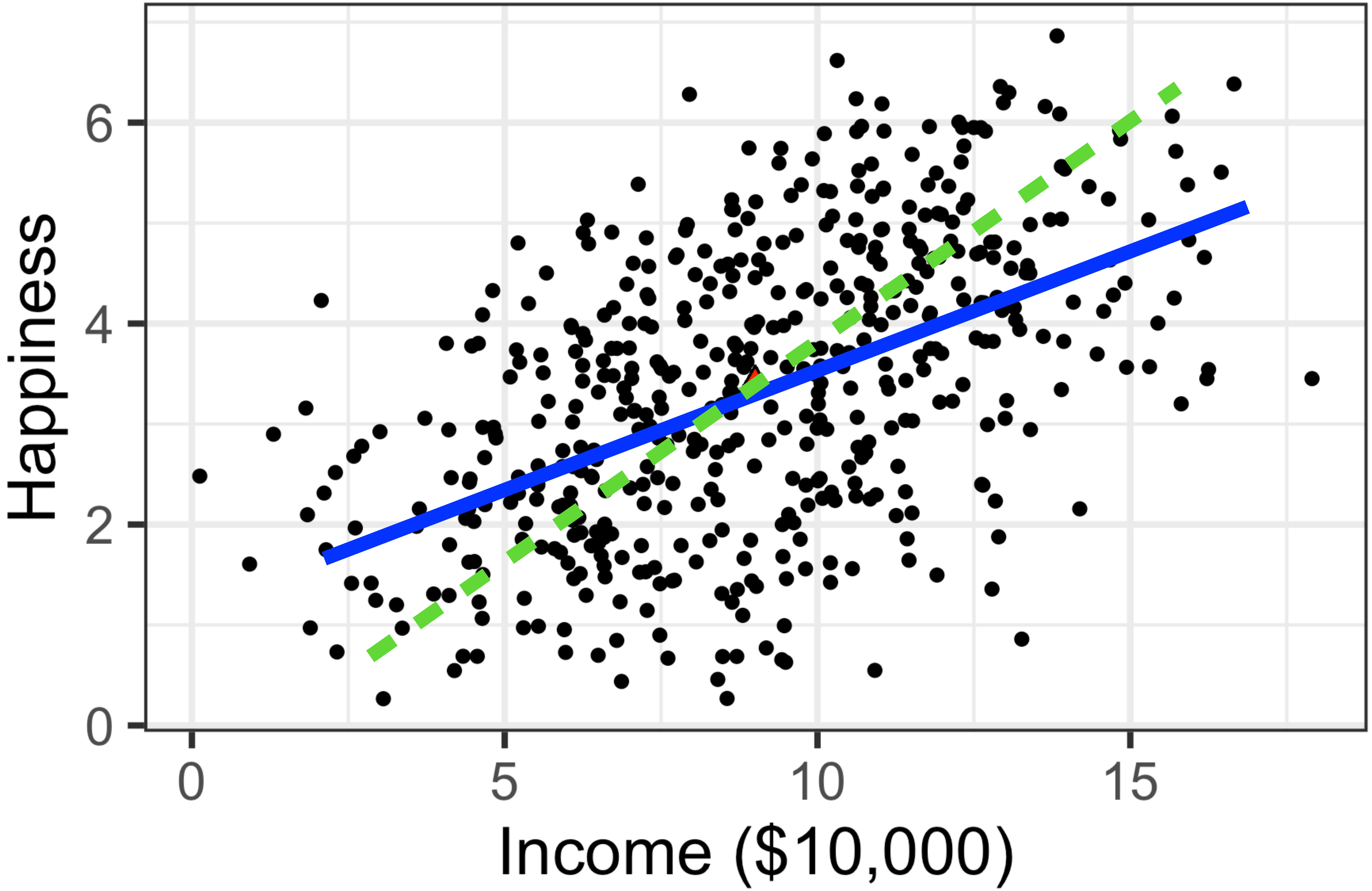
The fitted regression line: $r(s_y/s_x)$



The fitted regression line: $r(s_y/s_x)$

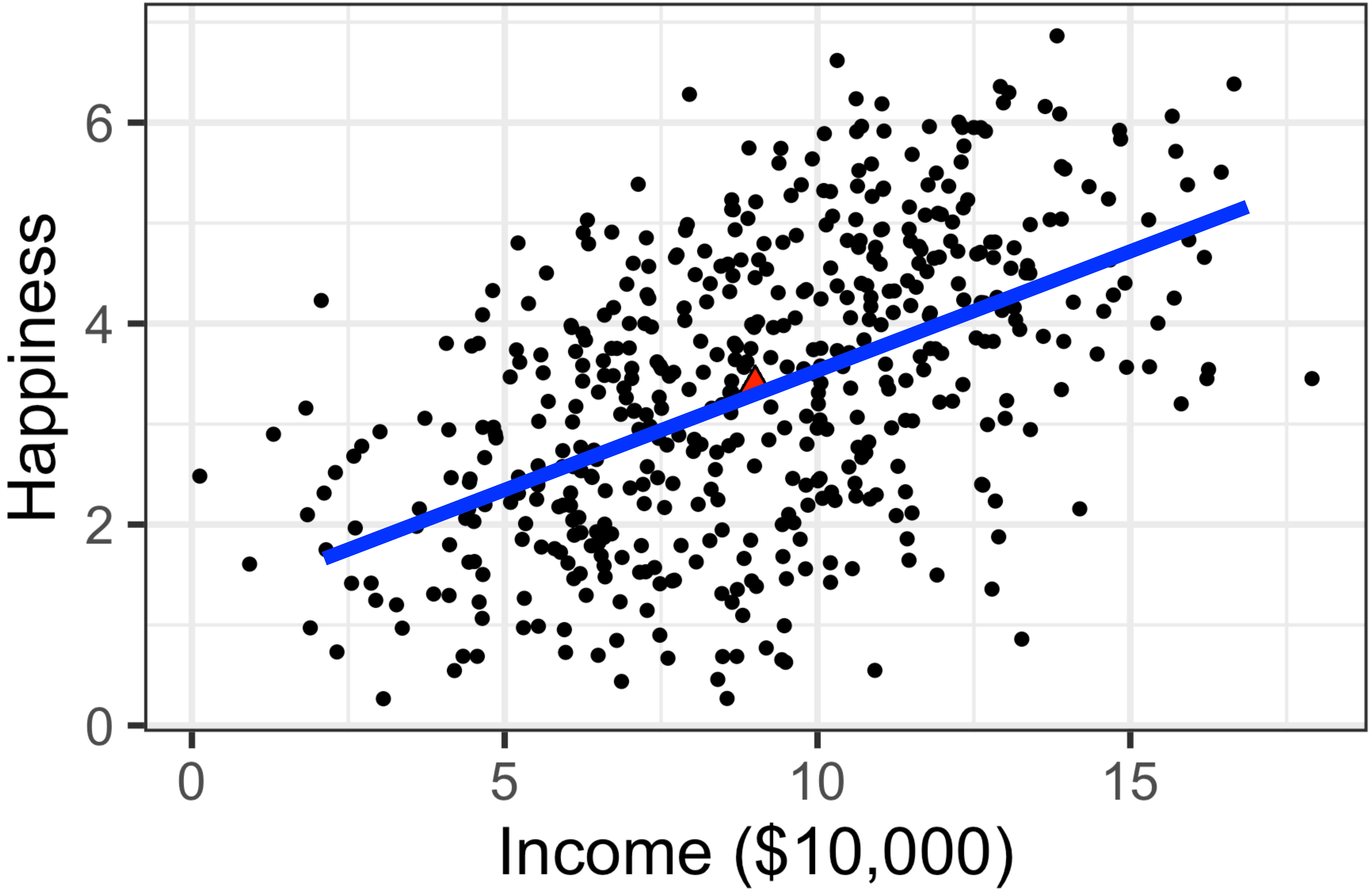


The fitted regression line: $r(s_y/s_x)$



If two variables do NOT have a perfect correlation ($r \neq \pm 1$), the slope of the line that fits the data best (least-squares or fitted regression line) will have a slope of $r(s_y/s_x)$ and passes through the point $(\bar{x}, \bar{y})$

The fitted regression line: $r(s_y/s_x)$




```
cor(income, happiness)
```

[1] 0.5093659

```
sd(income)
```

[1] 3.205921

```
sd(happiness)
```

[1] 1.432813

$$r(s_y/s_x) = 0.51(\frac{1.43}{3.2})$$

$$r(s_y/s_x) = 0.227$$

The fitted regression line: full equation

$$(F = 32 + \frac{9}{5}C)$$

(y = b + mx + e)

$$y = \beta_0 + \beta_1 X + \epsilon$$

Predicted value of
dependent variable
(i.e. happiness)

Intercept
(i.e. predicted happiness
at $0 income)

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

Regression coefficient
(i.e. How much we expect
y to change with x)

$$\beta_1 = r(s_y / s_x)$$

Independent variable
(i.e. income)

Error
(i.e. variation
in the estimate)

The fitted regression line: full equation

(y = b + mx + e)

$$\boxed{y = \beta_0 + \beta_1 X + \epsilon}$$

```
mean(income)
```

```
[1] 9.004844
```

```
mean(happiness)
```

```
[1] 3.392859
```

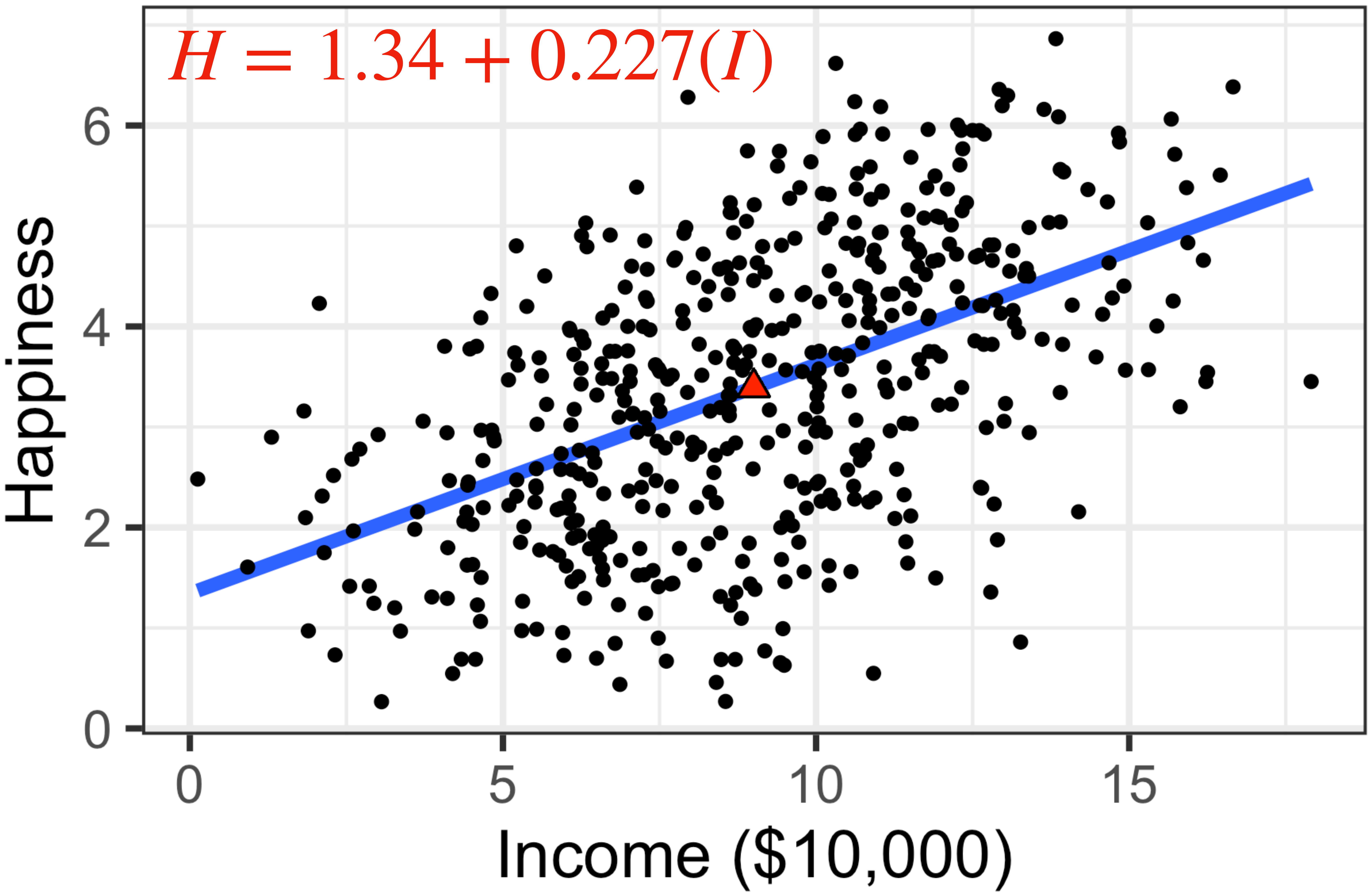
$$\beta_1 = r(s_y/s_x) = 0.227$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_0 = 3.39 - (0.227)(9.0)$$

$$= 1.347$$

The fitted regression line: $Y = \beta_0 + \beta_1 X$



Using R's `lm()` function

```
lm(happiness ~ income, data = income_data)
```

$$y = \beta_0 + \beta_1 X + \epsilon$$

```
Call:
lm(formula = happiness ~ income)

Coefficients:
(Intercept)        income
     1.3429        0.2276
```

For every 1 unit increase in income, there is a 0.227 unit increase in happiness

Using R's `lm()` function

```
lm(happiness ~ income, data = income_data)
```

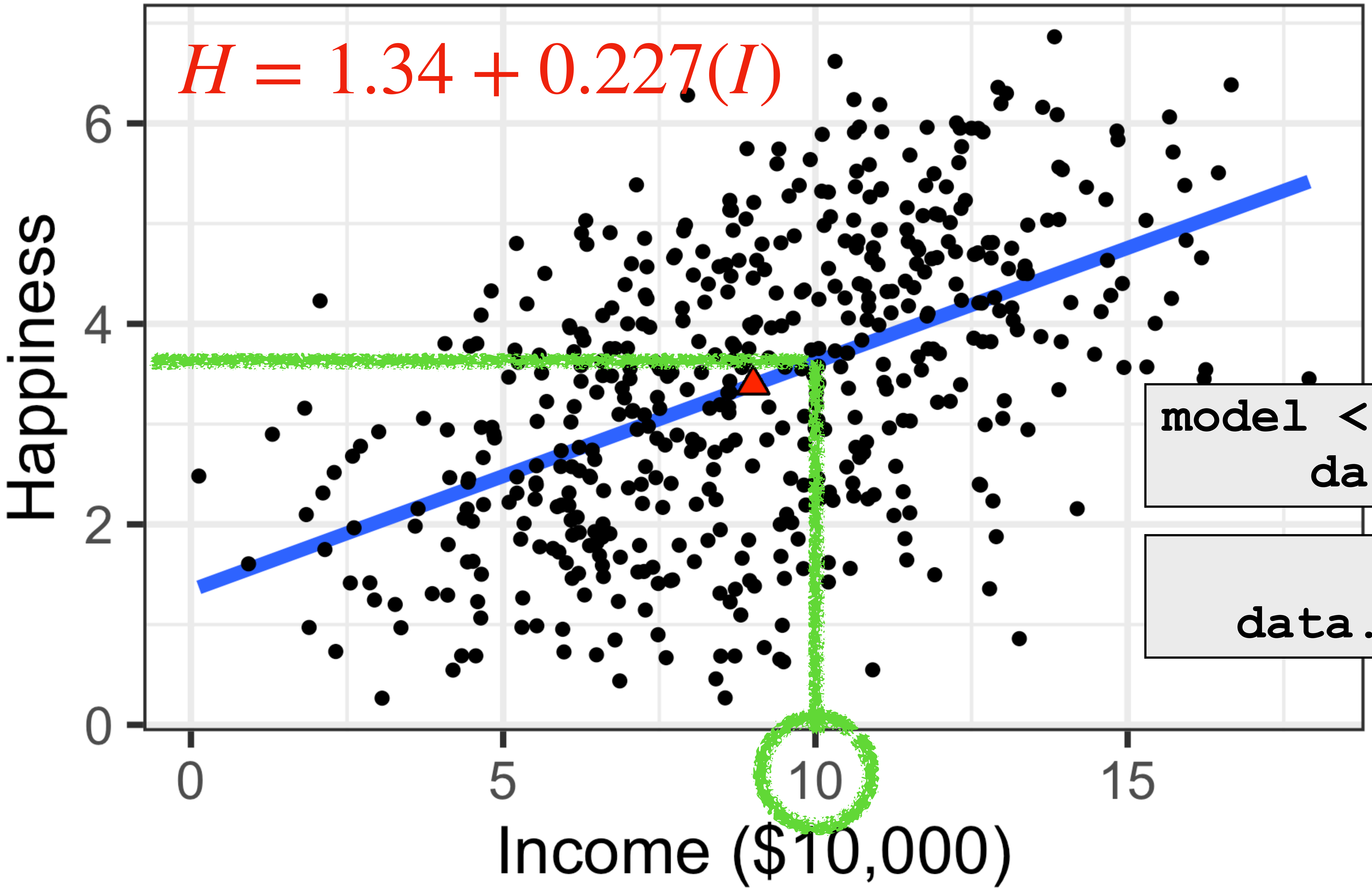
$$happiness = 1.34 + 0.227(income)$$

```
Call:
lm(formula = happiness ~ income)

Coefficients:
(Intercept)       income
     1.3429       0.2276
```

For every 1 unit increase in income, there is a 0.227 unit increase in happiness

Prediction using the fitted regression line



$$H = 1.34 + 0.227(I)$$

$$H = 1.34 + 0.227(10)$$

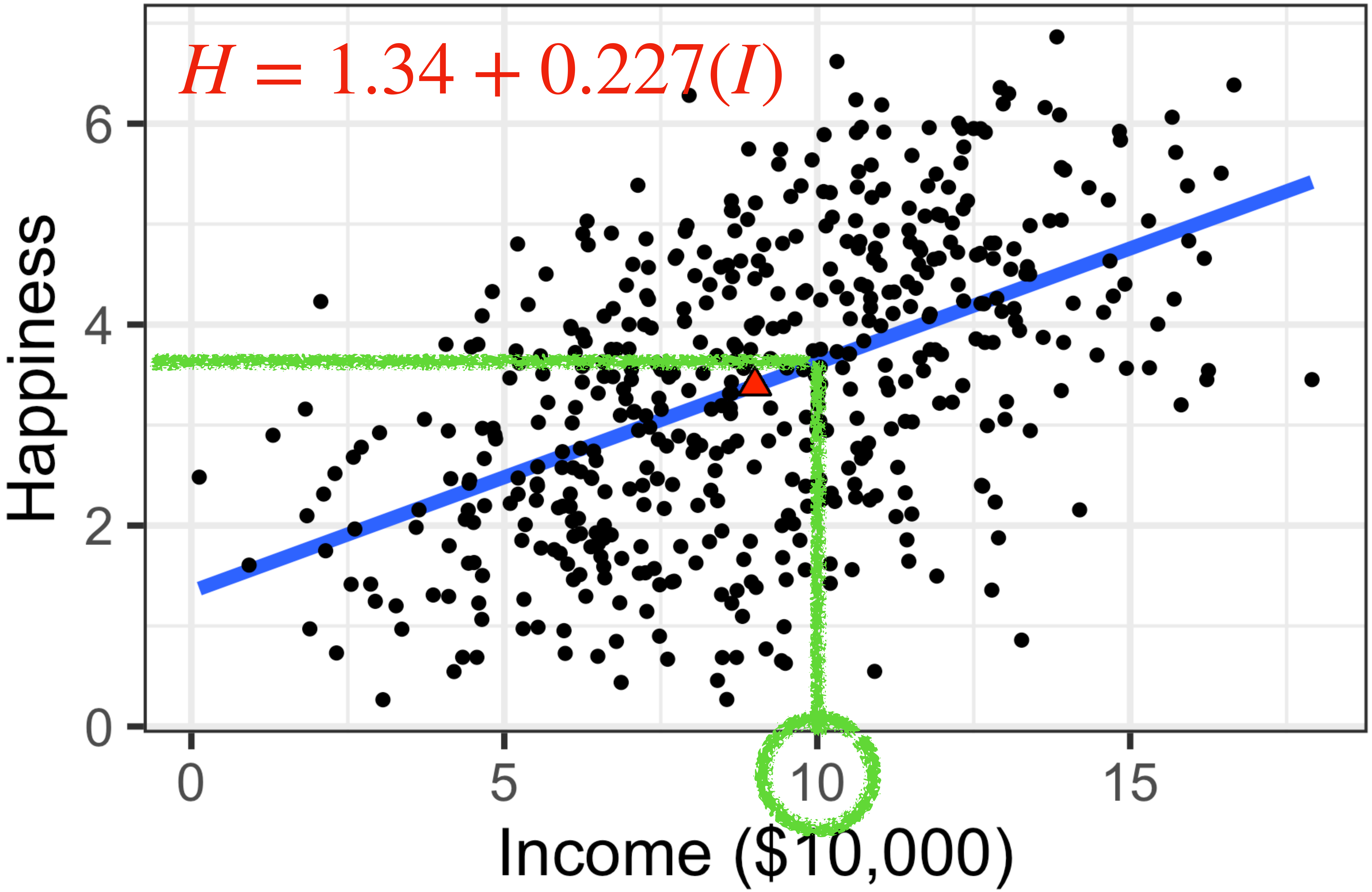
$$H = 3.61$$

```
model <- lm(happiness~income,
  data = income_data)
```

```
predict(model,
data.frame(income = 10))
```

3.619406

Prediction using the fitted regression line

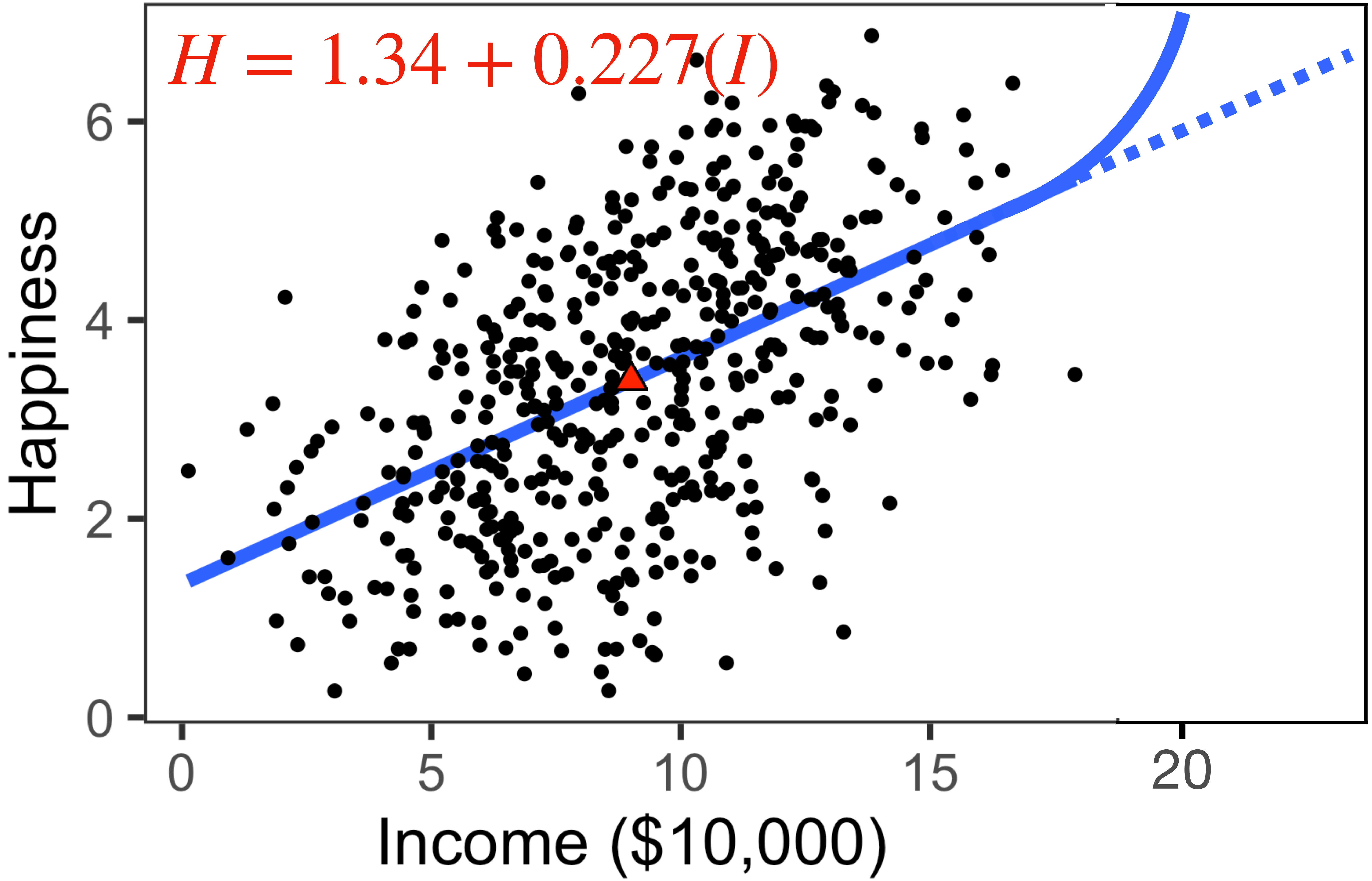


A note on prediction

Interpolation:
Predicting Y for an X within the range of the data

Extrapolation:
Predicting Y for an X outside the range of the data

Prediction using the fitted regression line



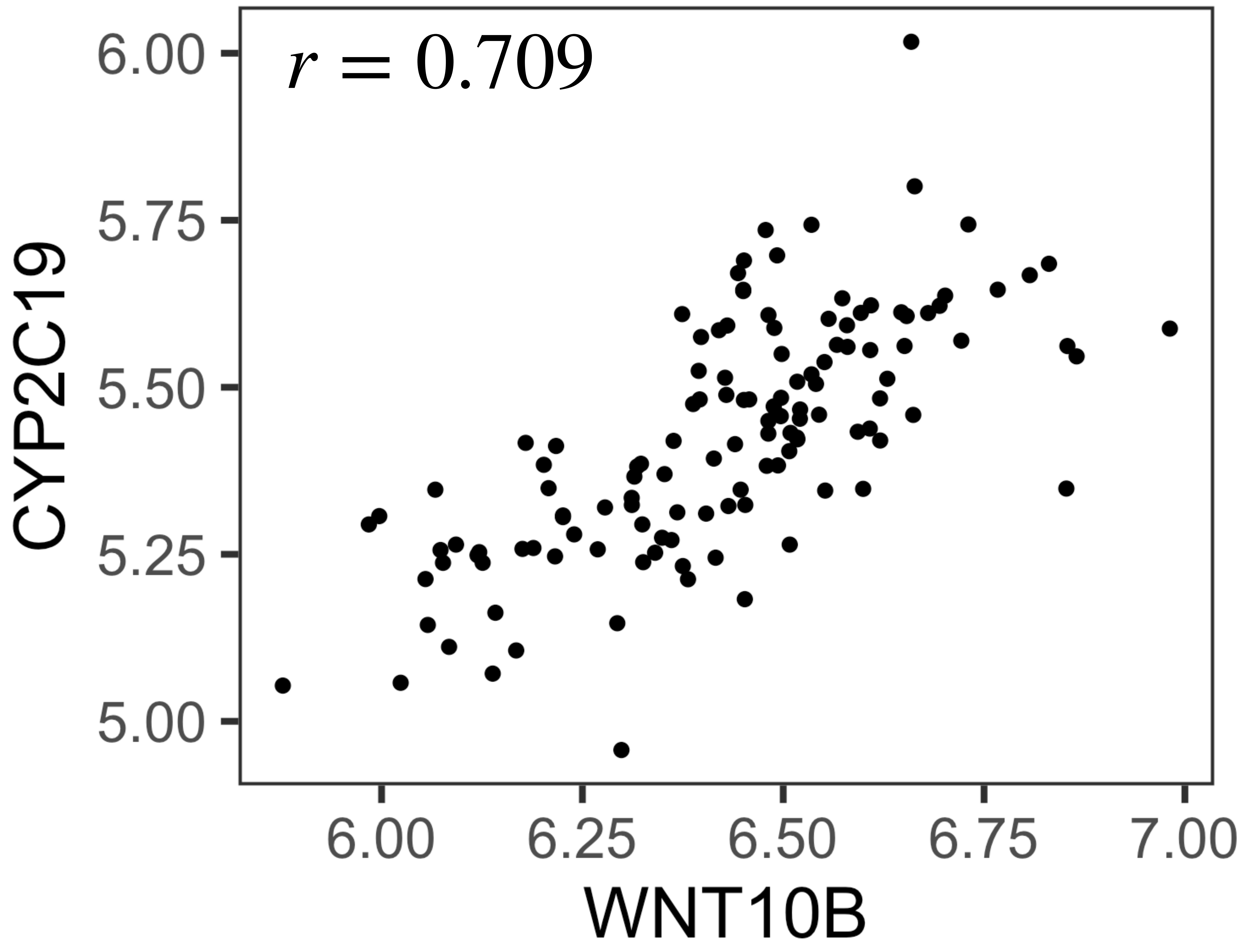
A note on prediction

Interpolation:
Predicting Y for an X within the range of the data

Extrapolation:
Predicting Y for an X outside the range of the data

No assurance relationship remains linear outside range

Gene expression example




```
> mean(WNT10B)
[1] 6.428509
```

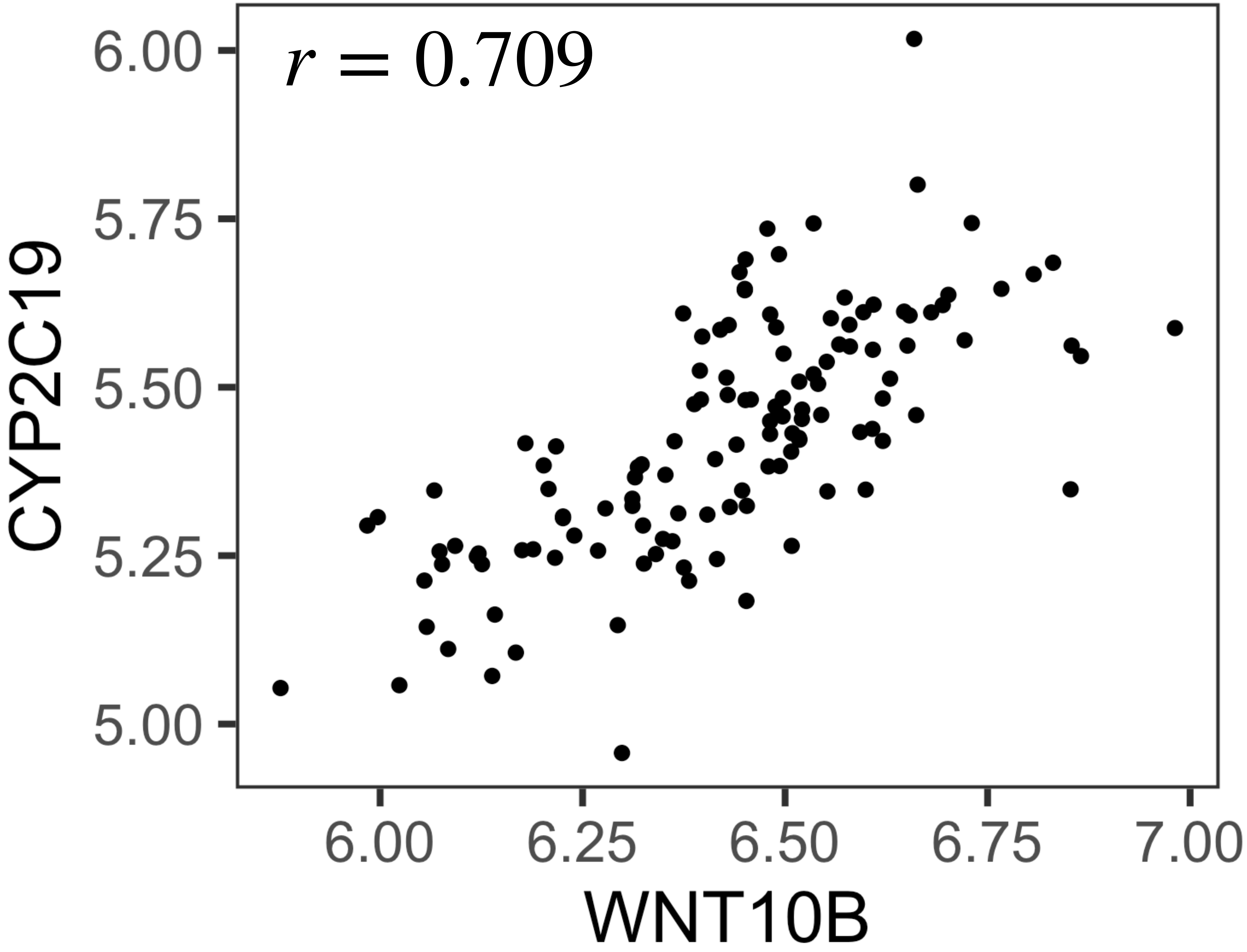
```
> sd(WNT10B)
[1] 0.2085725
```

```
> mean(CYP2C19)
[1] 5.427981
```

```
> sd(CYP2C19)
[1] 0.1771954
```

Calculate the regression line.

Gene expression example




```
> mean(WNT10B)
[1] 6.428509
```

```
> sd(WNT10B)
[1] 0.2085725
```

```
> mean(CYP2C19)
[1] 5.427981
```

```
> sd(CYP2C19)
[1] 0.1771954
```

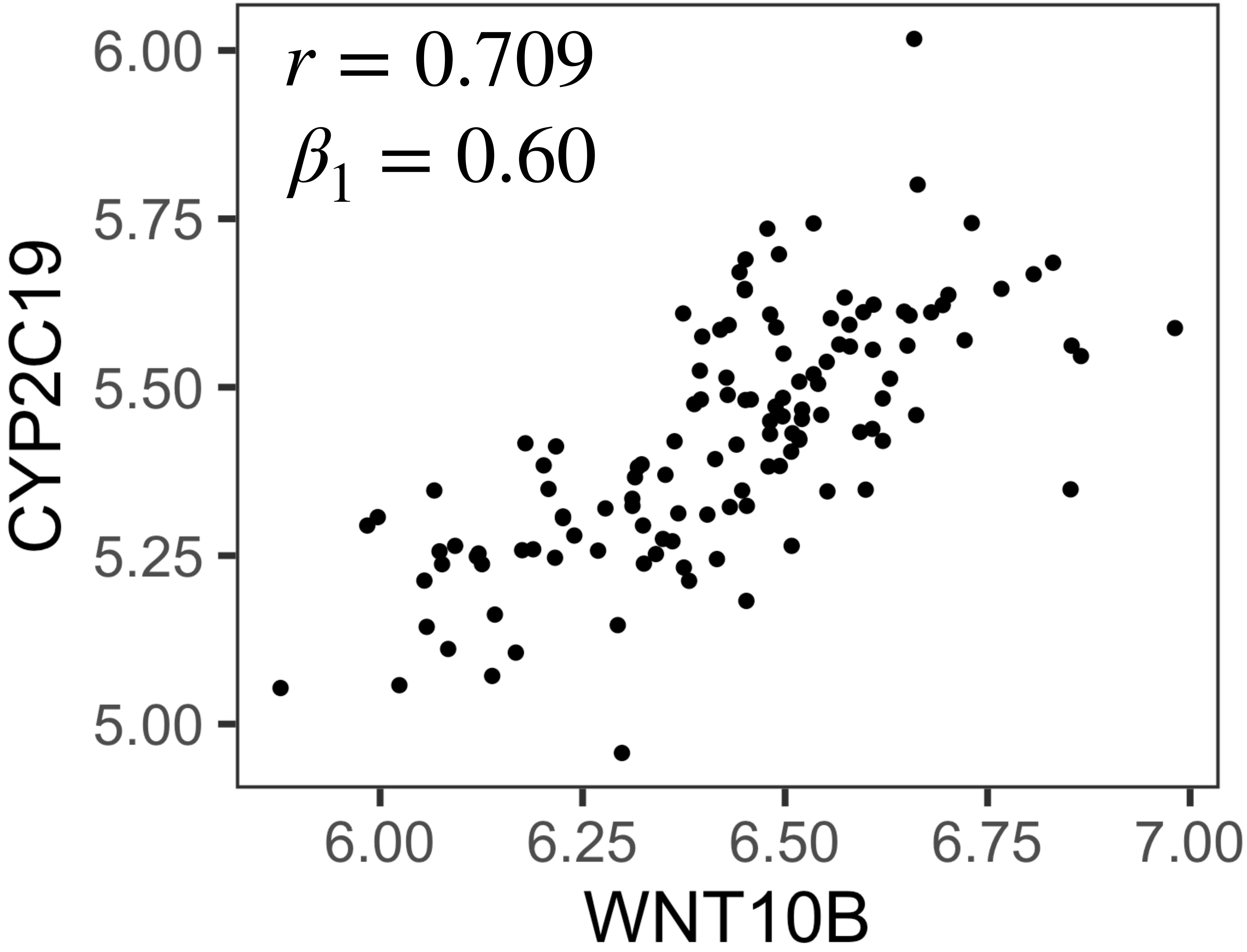
Calculate the regression line.

$$\beta_1 = r(s_y/s_x)$$

$$\beta_1 = (0.709)(0.177/0.209)$$

$$= 0.60$$

Gene expression example




```
> mean(WNT10B)
[1] 6.428509
```

```
> sd(WNT10B)
[1] 0.2085725
```

```
> mean(CYP2C19)
[1] 5.427981
```

```
> sd(CYP2C19)
[1] 0.1771954
```

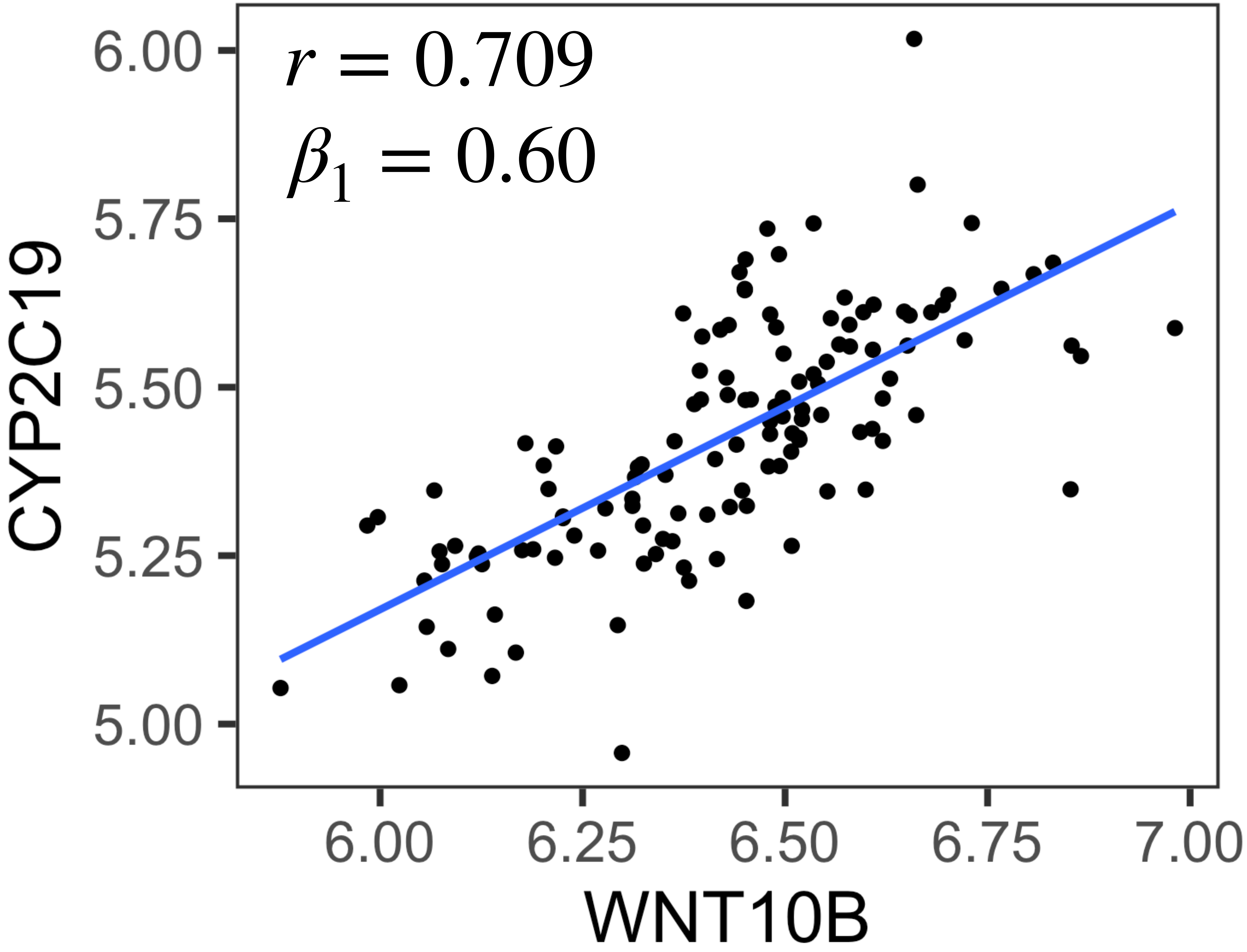
Calculate the regression line.

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

$$\beta_1 = 5.34 - (0.60)(6.43)$$

$$= 1.482$$

Gene expression example




```
> mean(WNT10B)
[1] 6.428509
```

```
> sd(WNT10B)
[1] 0.2085725
```

```
> mean(CYP2C19)
[1] 5.427981
```

```
> sd(CYP2C19)
[1] 0.1771954
```

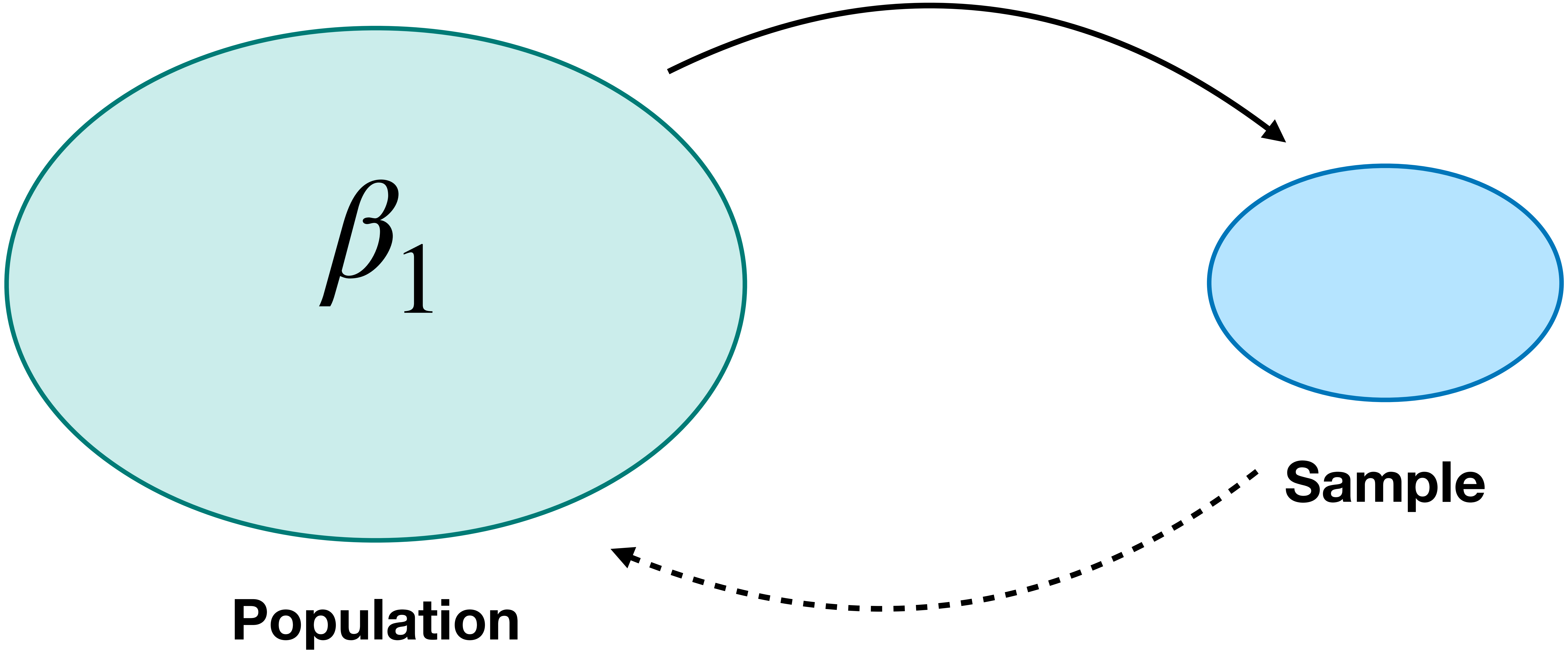
Calculate the regression line.

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

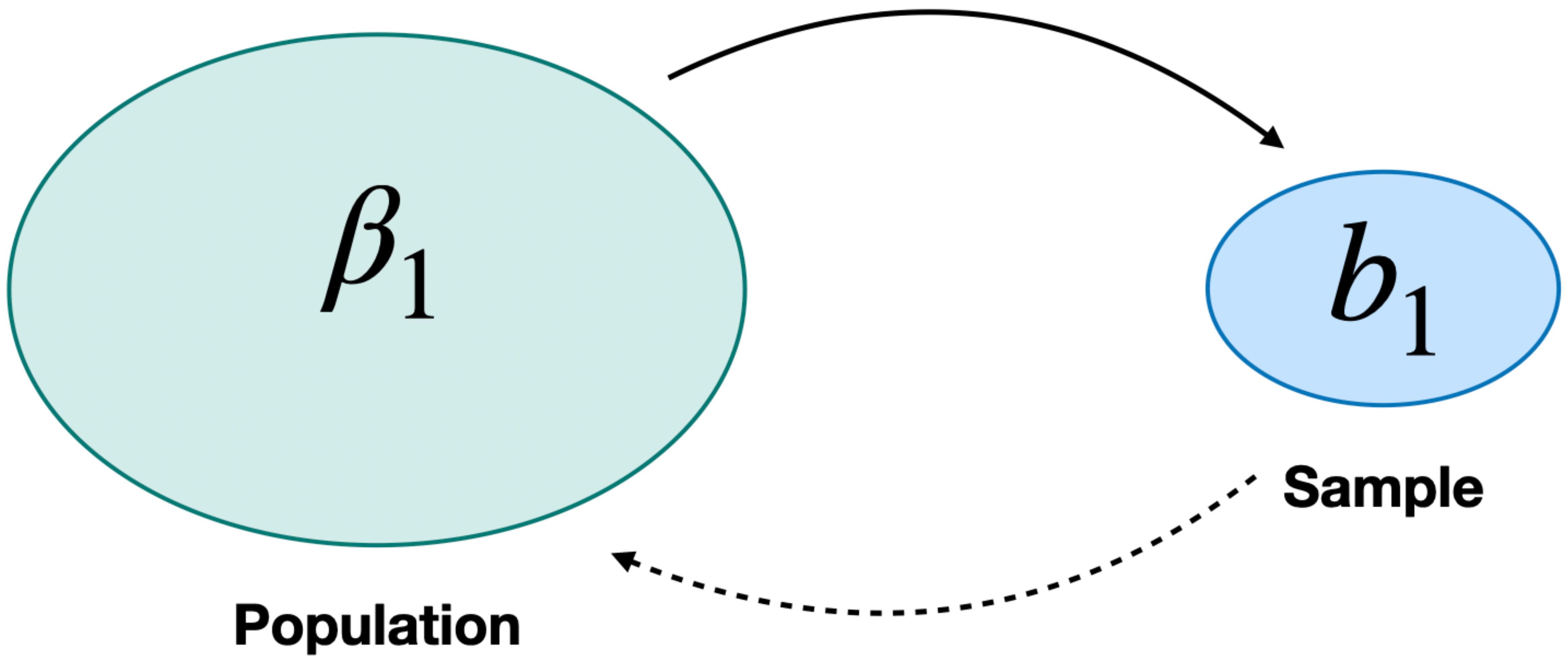
$$\beta_1 = 5.34 - (0.60)(6.43)$$

$$= 1.482$$

Interpreting the linear model



Interpreting the linear model



- **Estimate**

  ($\bar{y}$ for $\mu \rightarrow \bar{b}_1$ for β_1)

- **Error of the estimate**

  ($SE_{\bar{y}} \rightarrow SE_{\bar{b}_1}$)

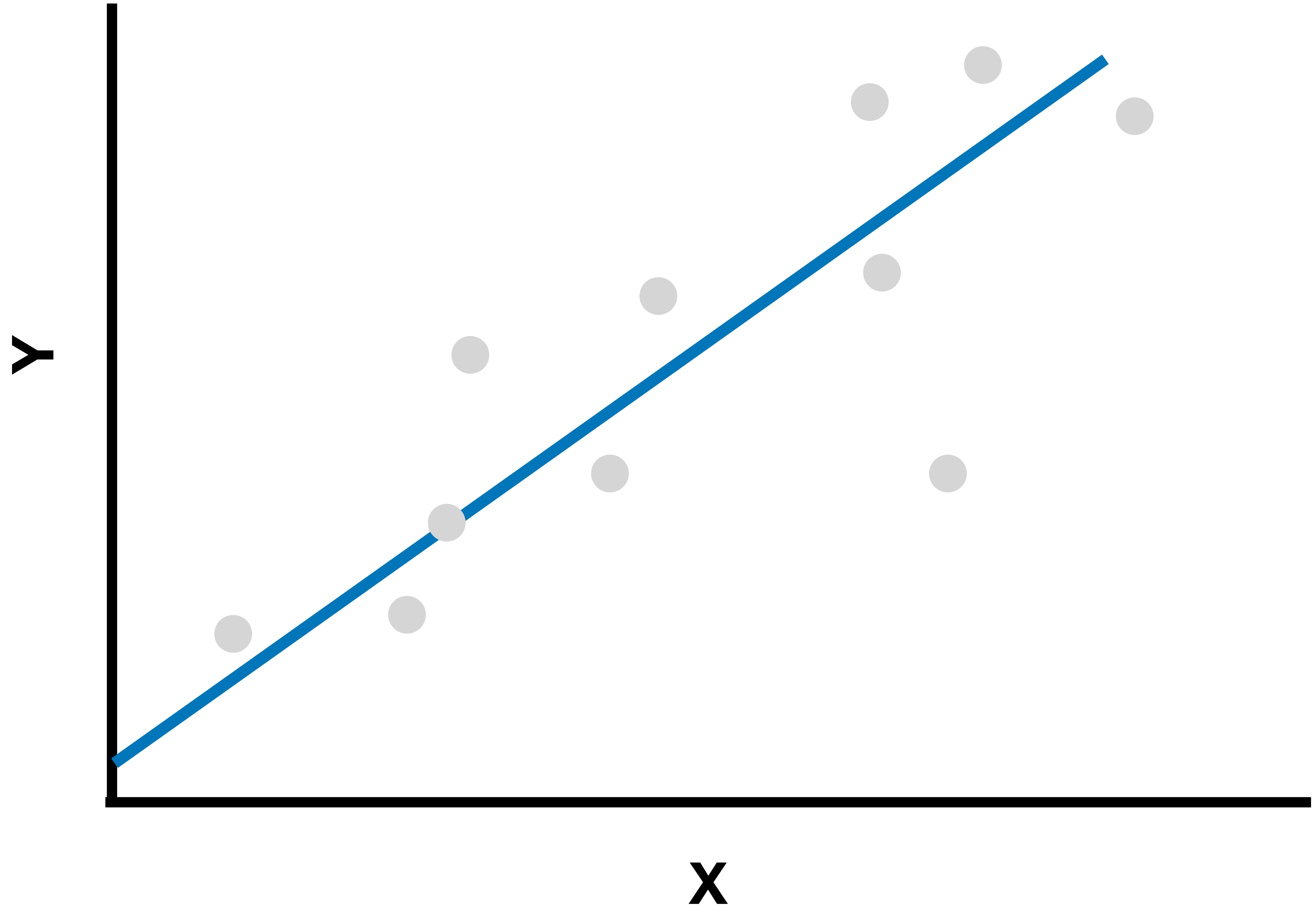
- **Confidence interval**

  ($\bar{y} \pm t_{0.025} SE_{\bar{y}} \rightarrow \bar{b}_1 \pm t_{0.025} SE_{\bar{b}_1}$)

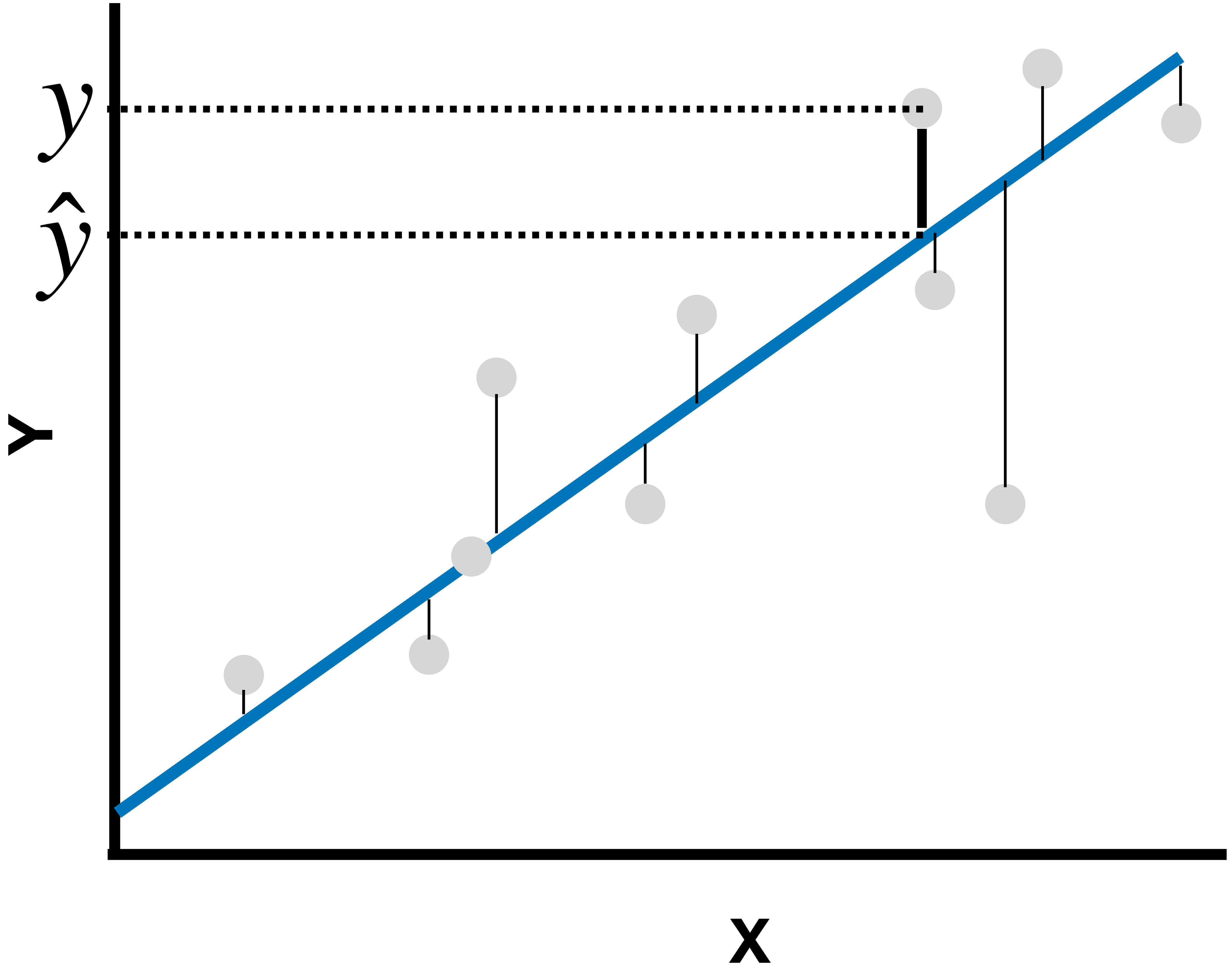
- **Hypothesis testing**

  ($H_0 : \mu = 0 \rightarrow H_0 : \beta_1 = 0$)

Error of the regression coefficient



Error of the regression coefficient



residual = observed - fitted

$$e_i = y_i - \hat{y}_i$$

Residual sum of squares:

$$SS(resid) = \sum e_i^2$$

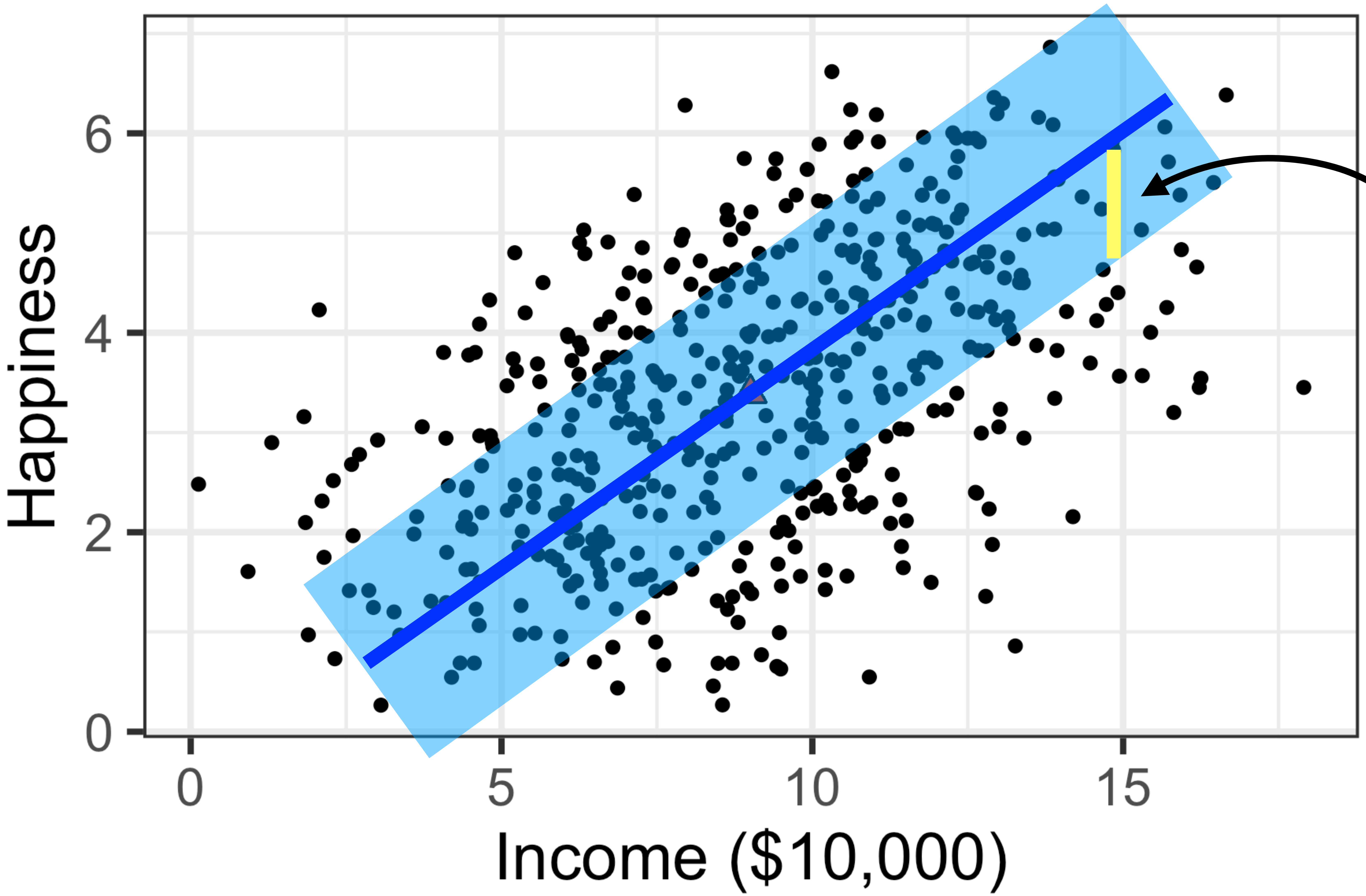
The residual sum of squares will be small if the data points all lie very close to the line.

Error of the regression coefficient

“Variance” around the best fit line

$$SS(resid) = 0 \quad < \quad SS(resid) = 7.98 \quad < \quad SS(resid) = 83.1$$

The "best fit" line **minimizes the SS(resid)**



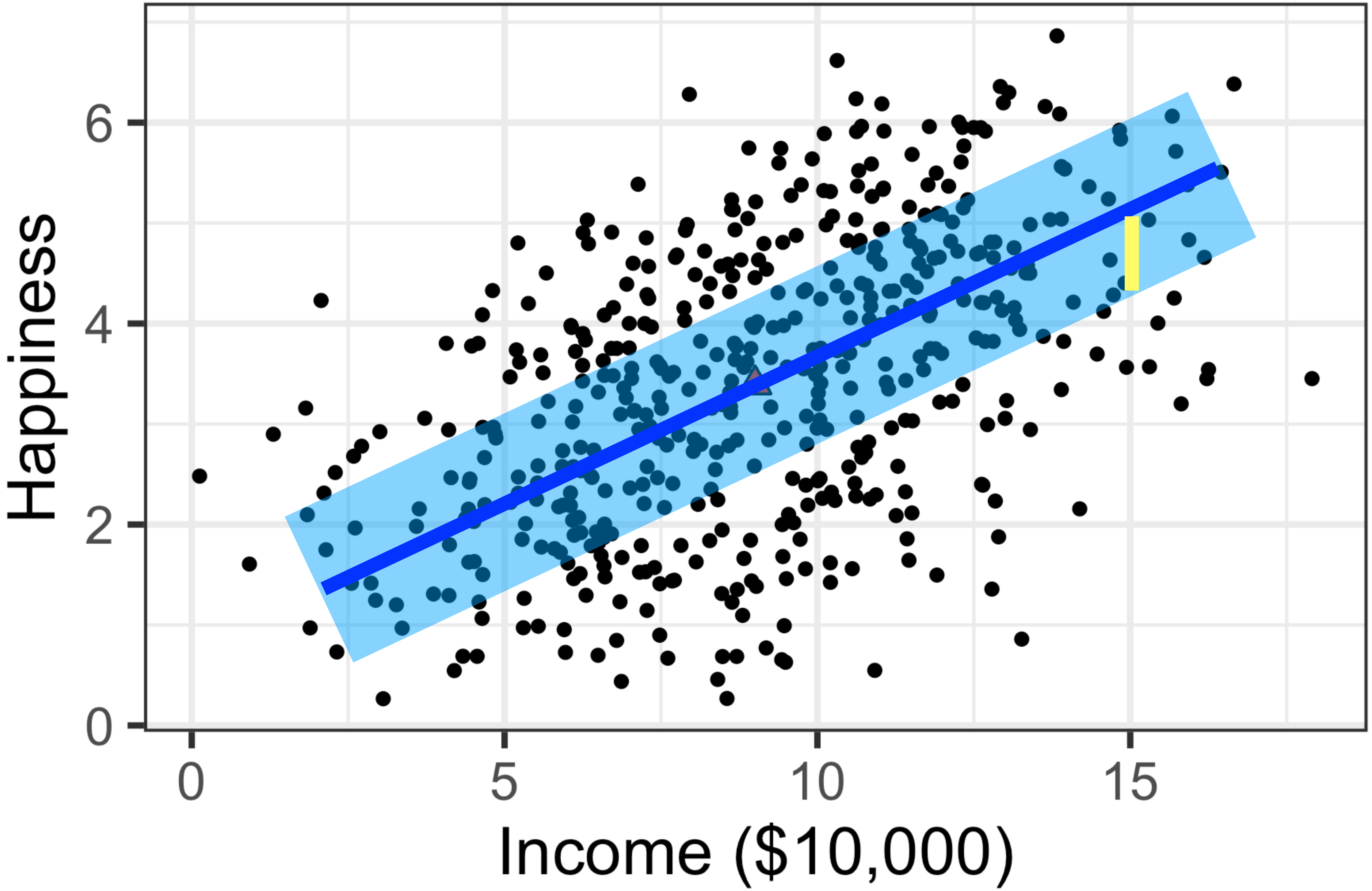
Residual SD:

$$s_e = \sqrt{\frac{\sum e_i^2}{n-2}}$$

Degrees of freedom

Residual SD specifies how far off predictions made using the regression model tend to be.

The "best fit" line **minimizes the SS(resid)**



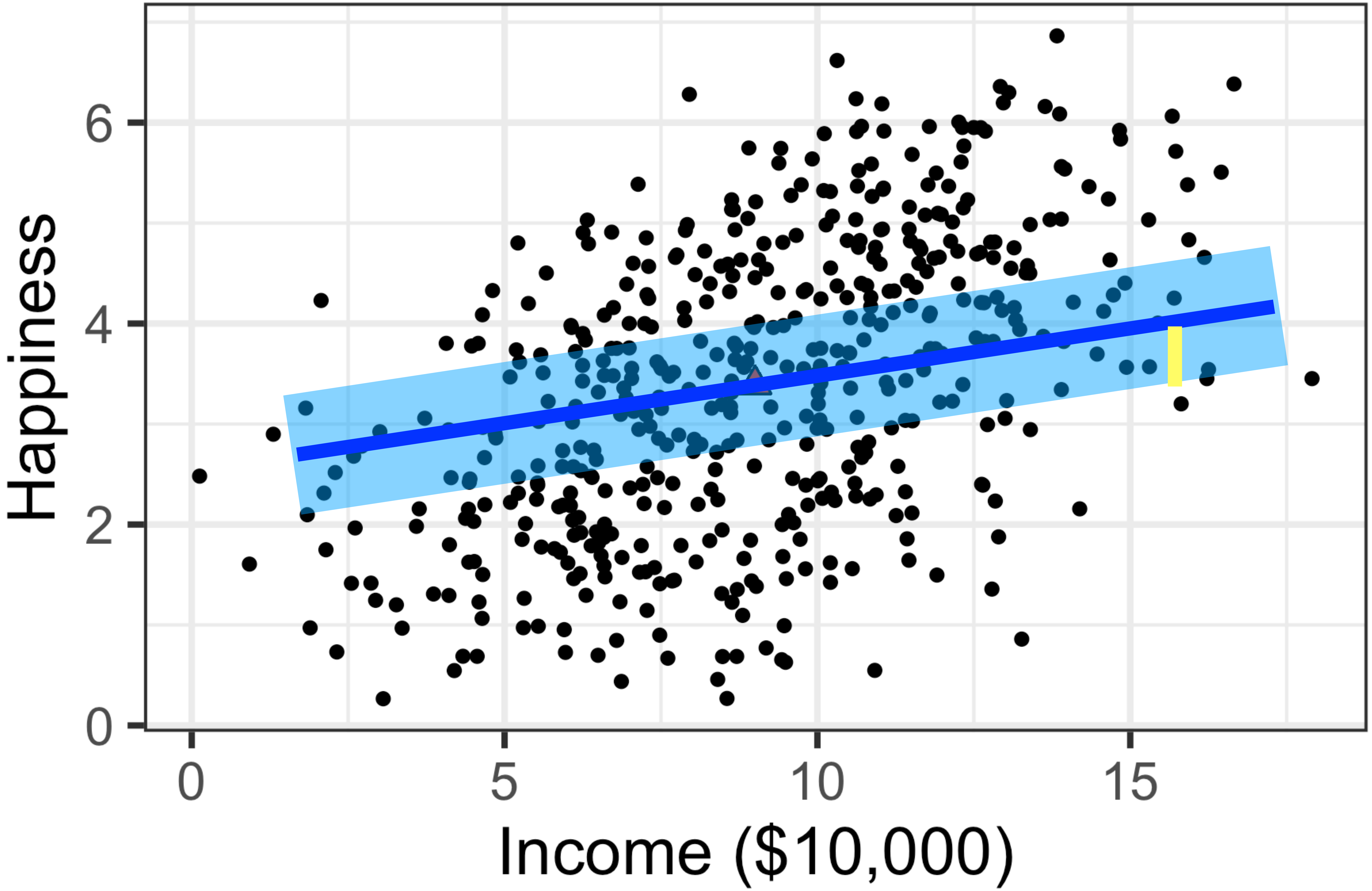
Residual SD:

$$s_e = \sqrt{\frac{\sum e_i^2}{n-2}}$$

Degrees of freedom

Residual SD specifies how far off predictions made using the regression model tend to be.

The “best fit” line **minimizes the SS(resid)**



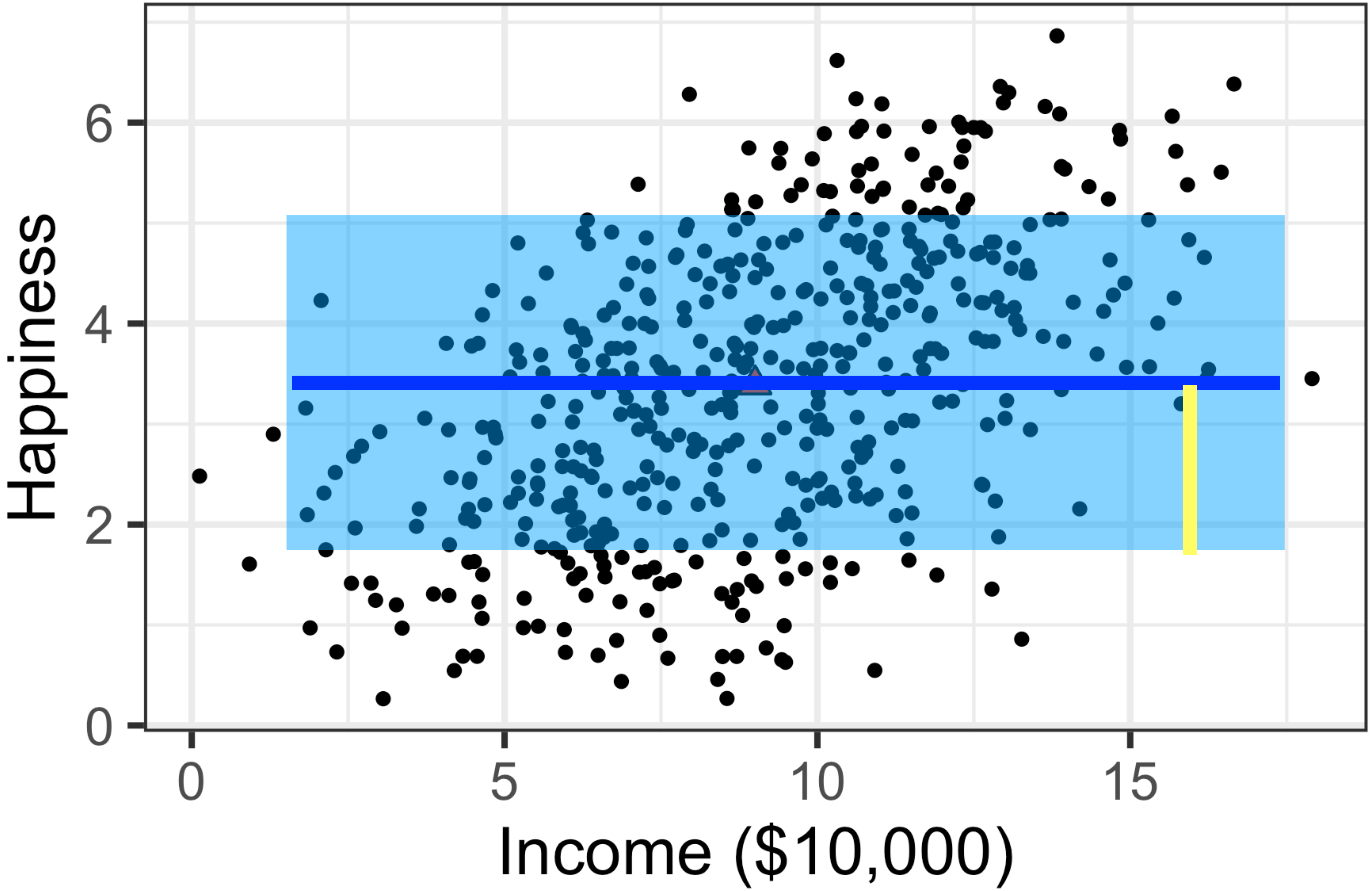
Residual SD:

$$s_e = \sqrt{\frac{\sum e_i^2}{n-2}}$$

Degrees of freedom

Residual SD specifies how far off predictions made using the regression model tend to be.

The “best fit” line **minimizes the SS(resid)**



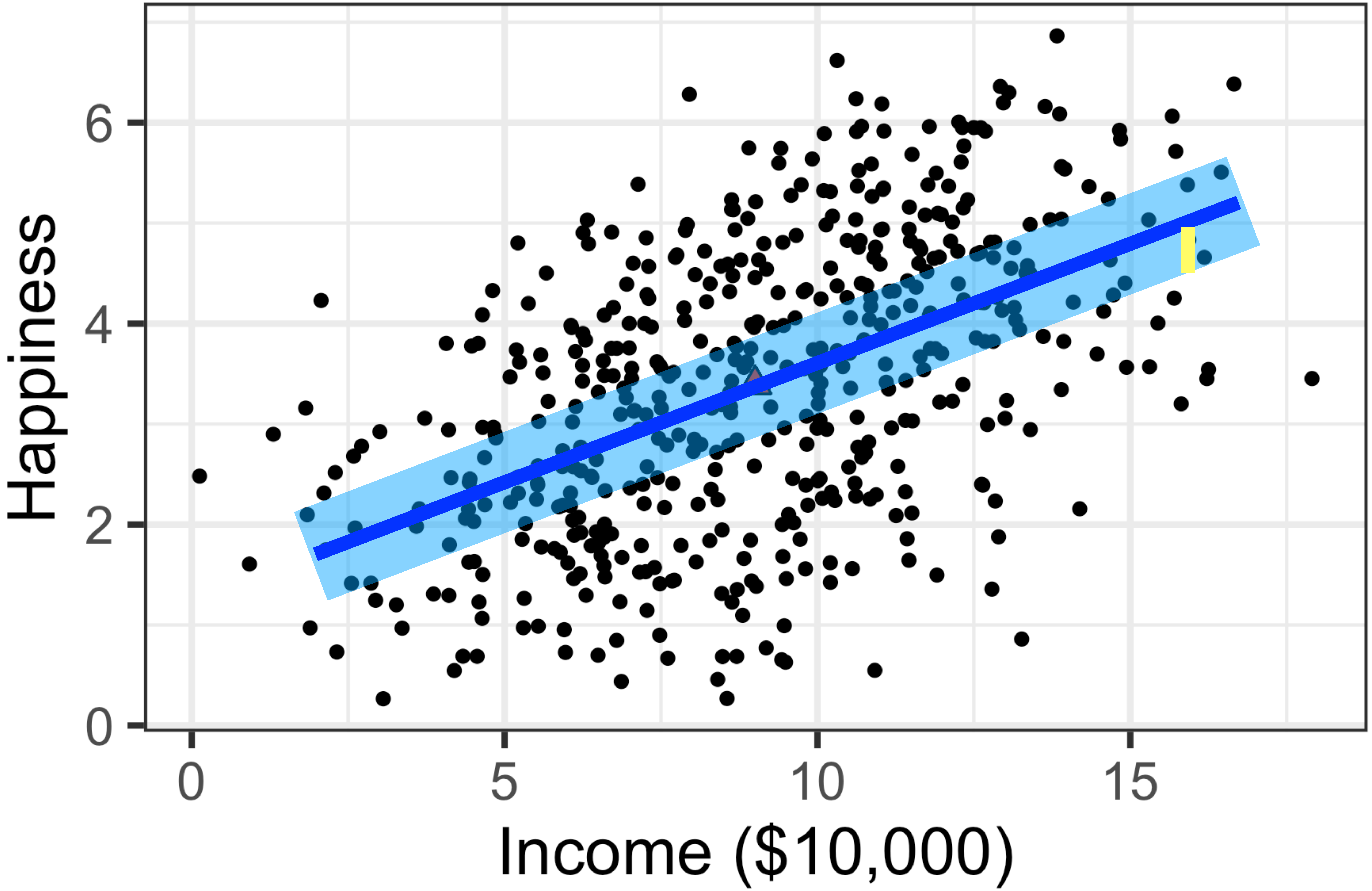
Residual SD:

$$s_e = \sqrt{\frac{\sum e_i^2}{n-2}}$$

Degrees of freedom

Residual SD specifies how far off predictions made using the regression model tend to be.

The “best fit” line **minimizes the SS(resid)**



Residual SD:

$$s_e = \sqrt{\frac{\sum e_i^2}{n-2}}$$

Degrees of freedom

Residual SD specifies how far off predictions made using the regression model tend to be.

The "best fit" line **minimizes the SS(resid)**

SD:

$$s_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n - 1}}$$

Degrees of freedom

Residual SD:

$$s_e = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - 2}}$$

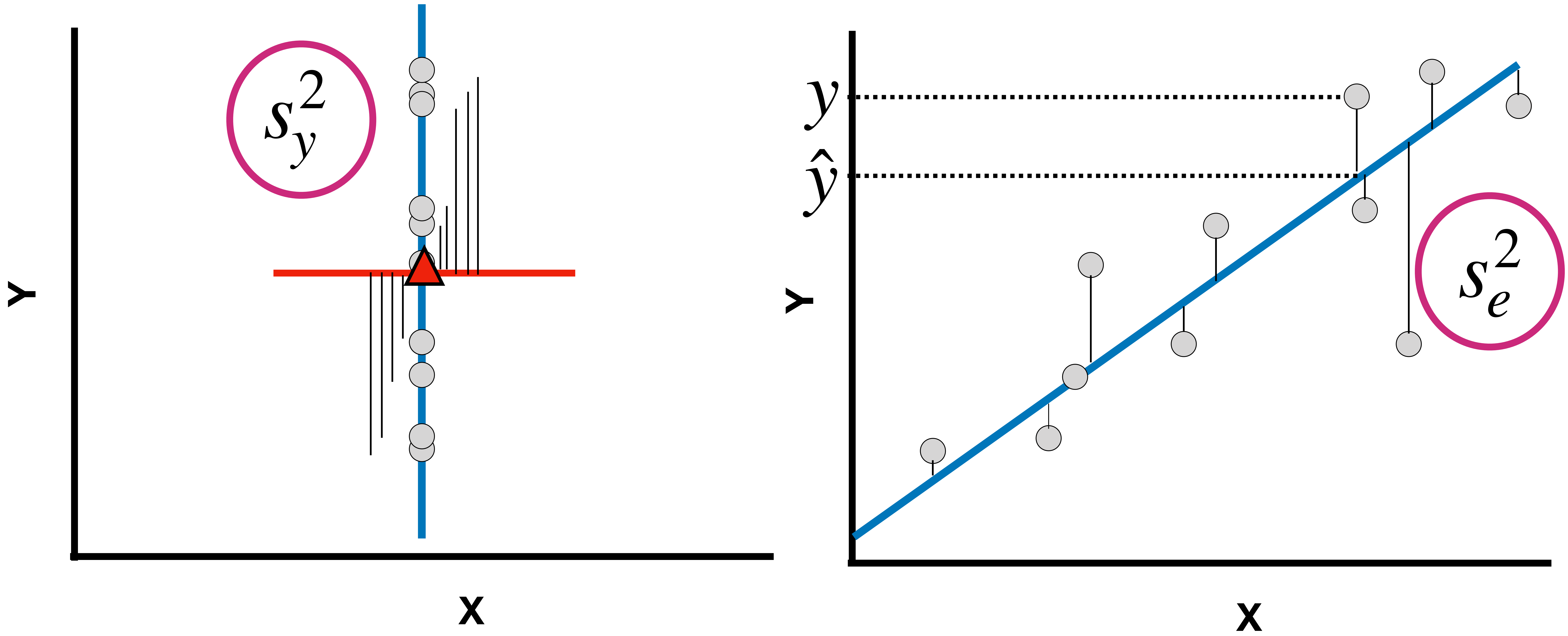
Degrees of freedom

SD measures variability around the *mean*

Residual SD measures variability around the *regression line*

The coefficient of determination, r^2

How much of the total variance in Y can be explained by X?



The coefficient of determination, r^2

The proportion of variance in Y that is explained by the linear relationship between X and Y

$$r^2 \approx \frac{s_y^2 - s_e^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$$

Also the square of correlation coefficient, *r*

0 (i.e. data fits regression line very well)

$r^2 = 1 - 0 = 1$

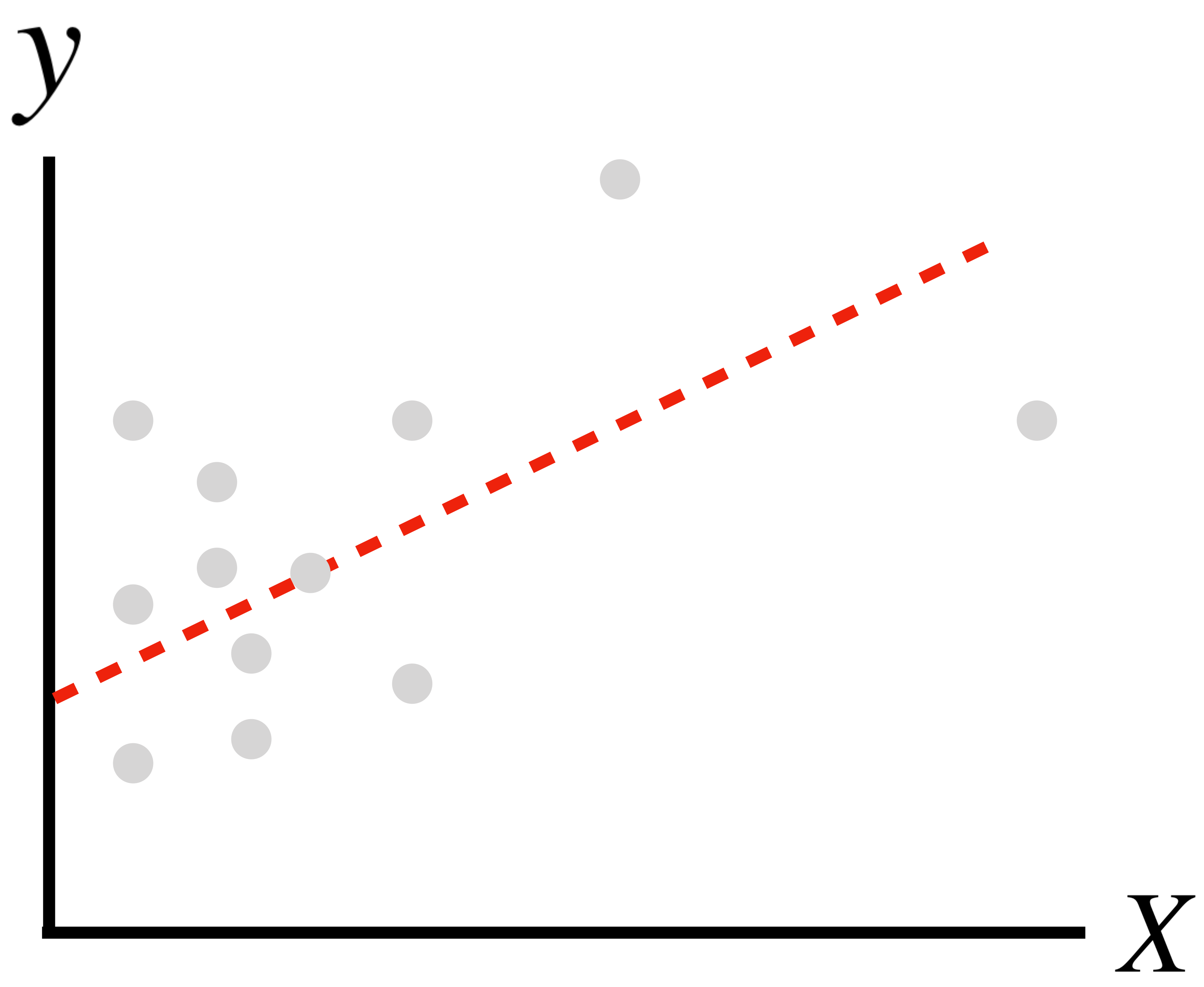
s_y (i.e. data fits regression line very poorly)

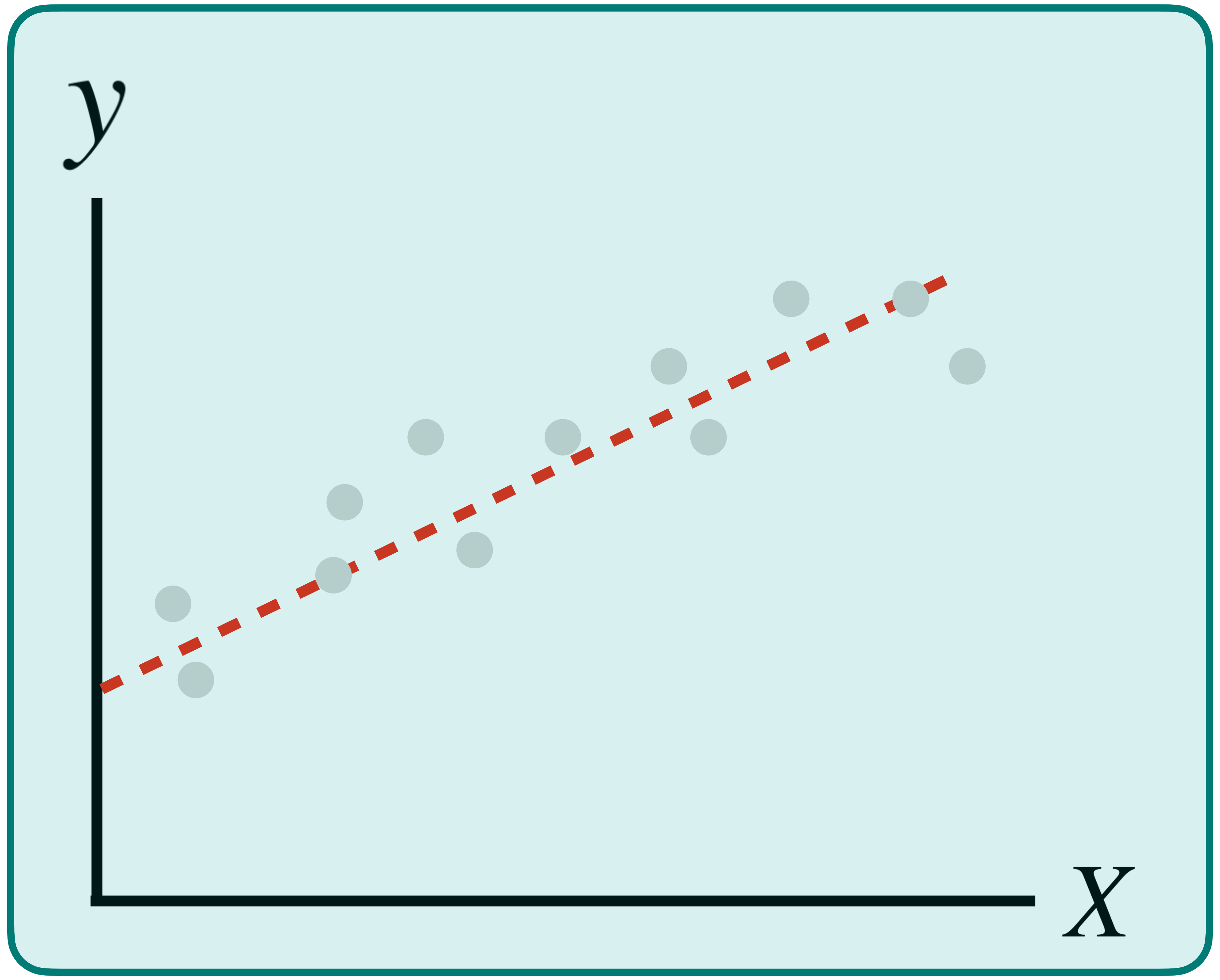
$r^2 = 1 - 1 = 0$

The coefficient of determination, r^2

The proportion of variance in Y that is explained by the linear relationship between X and Y

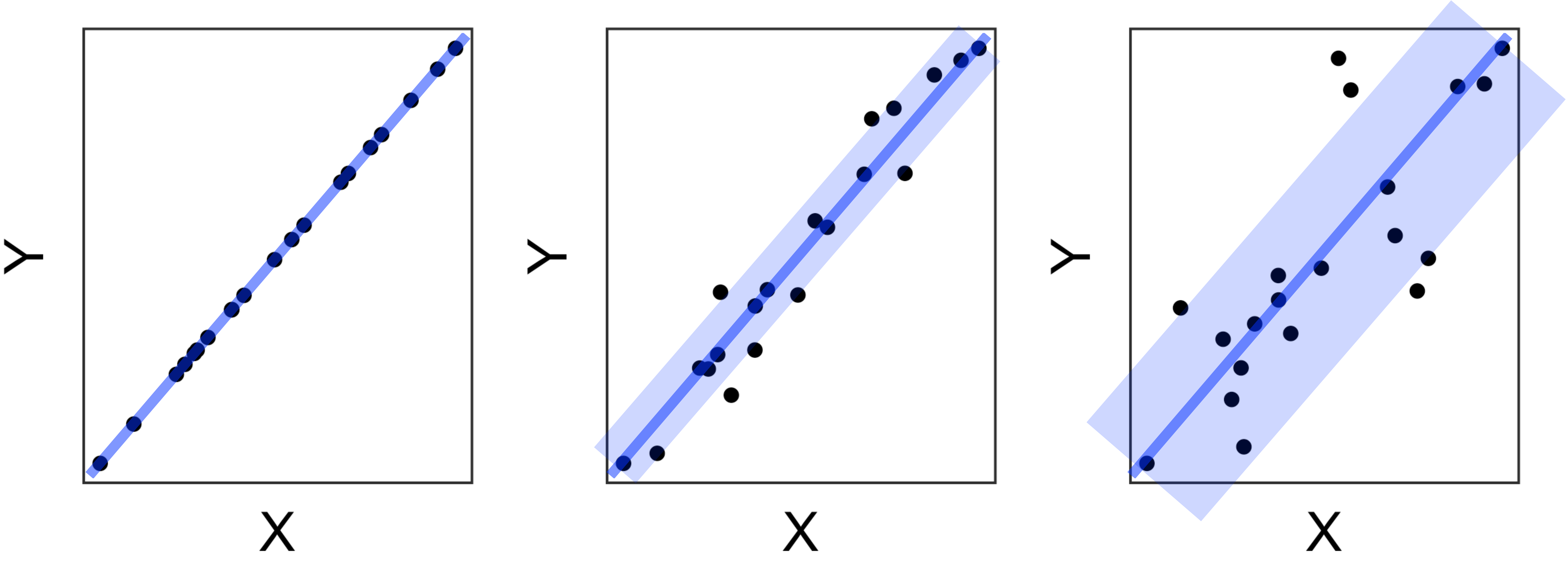
Which model has the higher r^2 value?





$$s_e = \sqrt{\frac{SS(resid)}{n-2}}$$

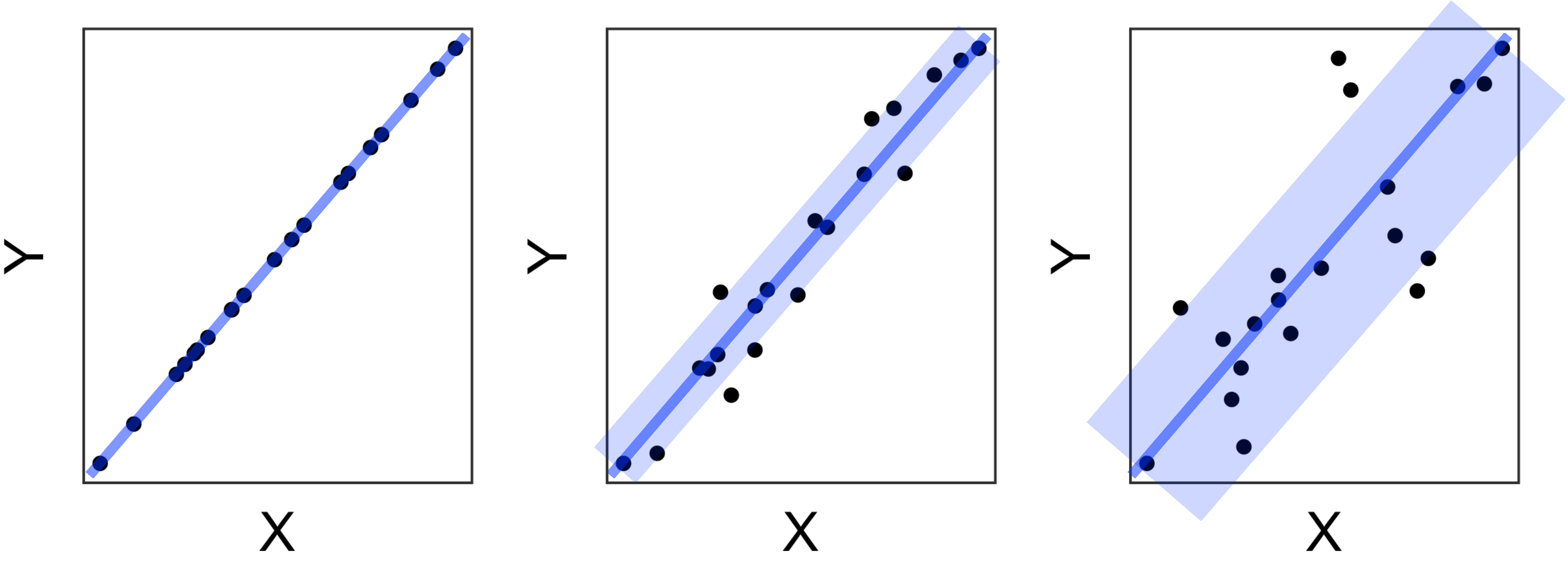
$$r^2 = 1 - \frac{s_e^2}{s_y^2}$$



$SS(resid)$	0	7.98	83.1
n	20	20	20
s_y	2.89	2.85	3.45
s_e			
r^2			

$$s_e = \sqrt{\frac{SS(resid)}{n-2}}$$

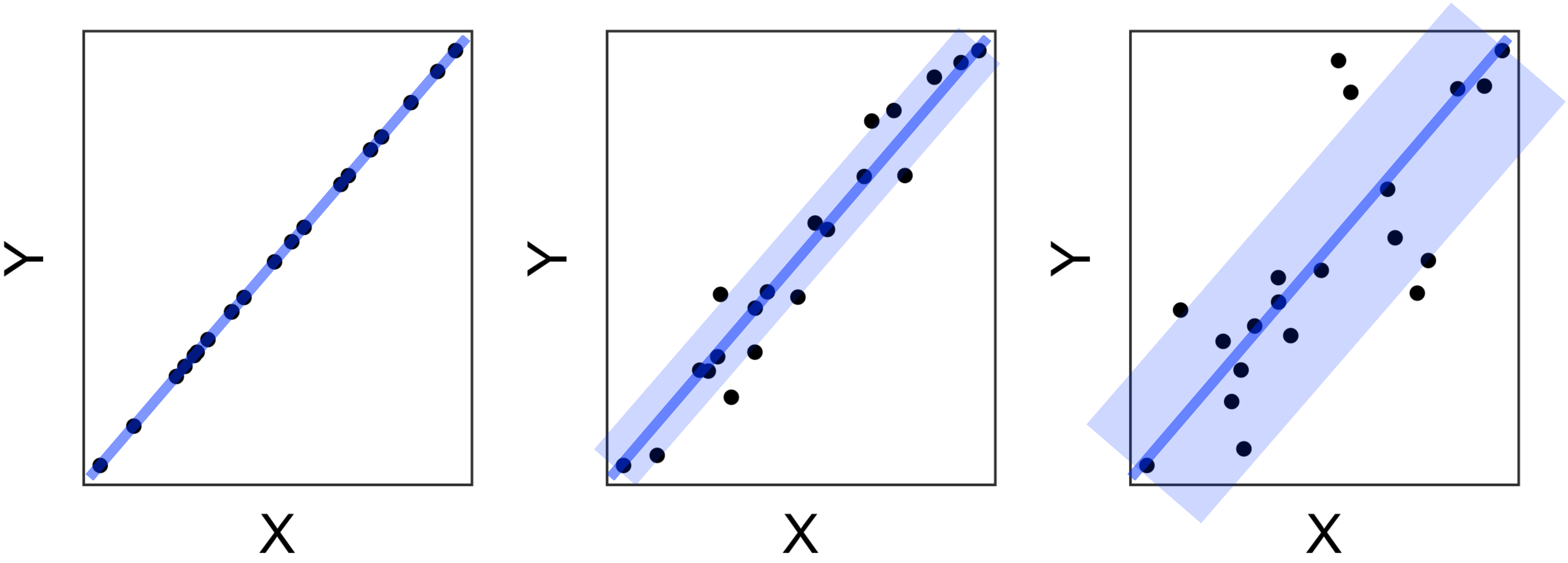
$$r^2 = 1 - \frac{s_e^2}{s_y^2}$$



$SS(resid)$	0	7.98	83.1
n	20	20	20
s_y	2.89	2.85	3.45
s_e			
r^2			

$$s_e = \sqrt{\frac{SS(resid)}{n-2}}$$

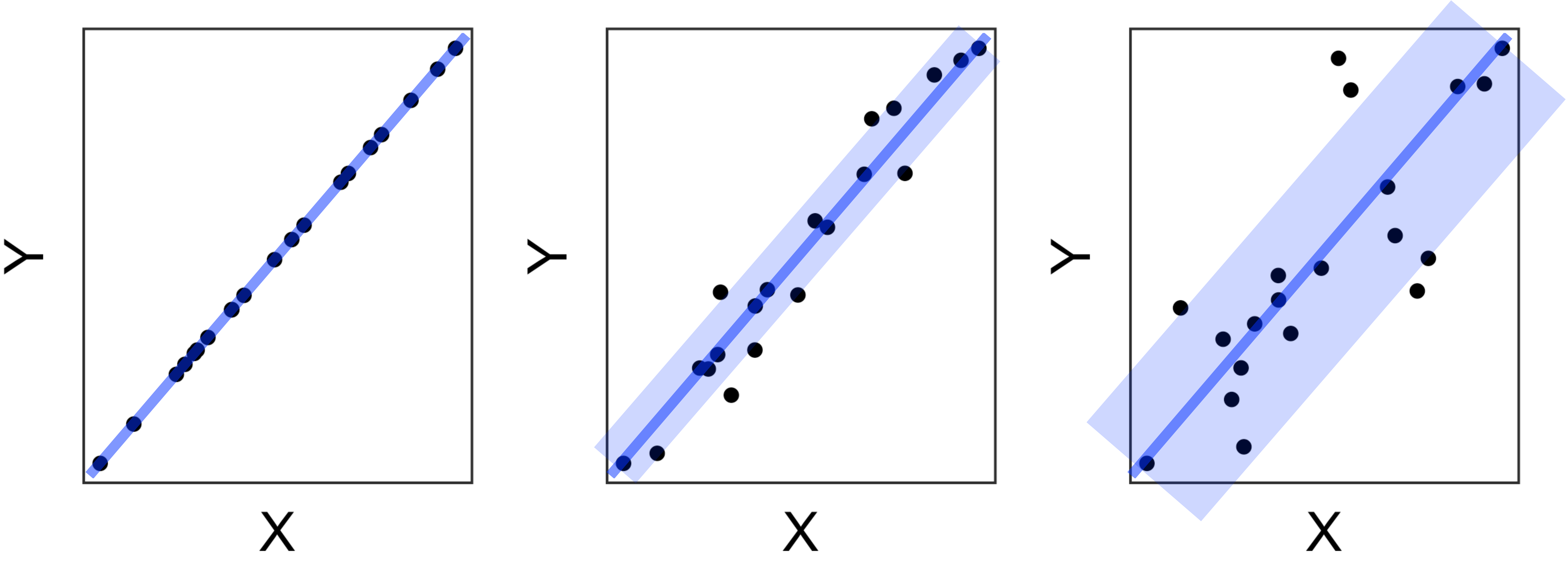
$$r^2 = 1 - \frac{s_e^2}{s_y^2}$$



$SS(resid)$	0	7.98	83.1
n	20	20	20
s_y	2.89	2.85	3.45
s_e	0	0.631	2.04
r^2			

$$s_e = \sqrt{\frac{SS(resid)}{n-2}}$$

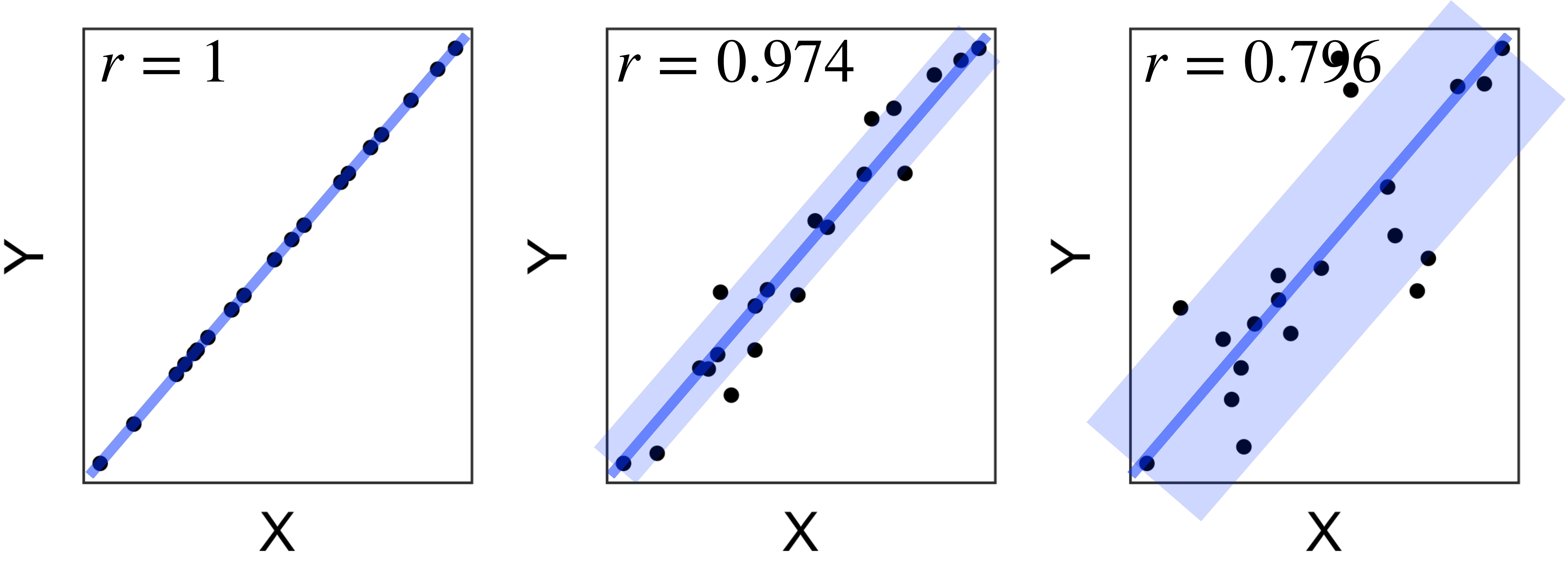
$$r^2 = 1 - \frac{s_e^2}{s_y^2}$$



$SS(resid)$	0	7.98	83.1
n	20	20	20
s_y	2.89	2.85	3.45
s_e	0	0.631	2.04
r^2			

$$s_e = \sqrt{\frac{SS(resid)}{n-2}}$$

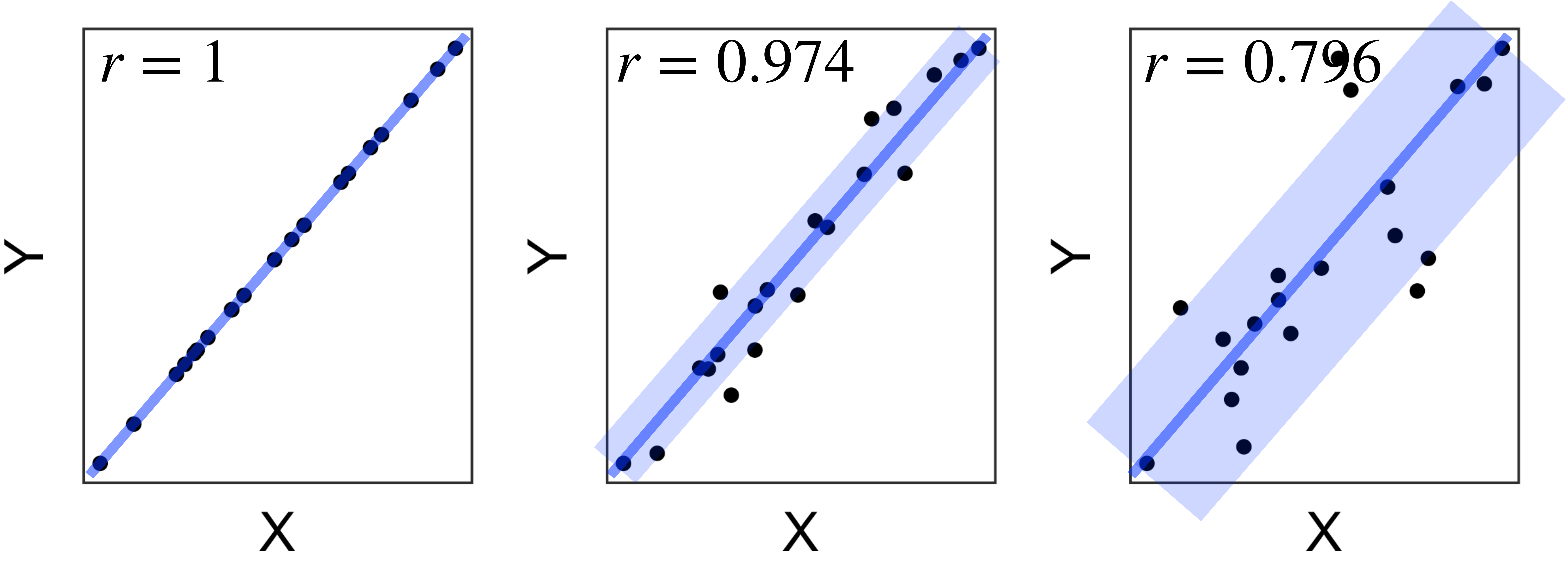
$$r^2 = 1 - \frac{s_e^2}{s_y^2}$$



$SS(resid)$	0	7.98	83.1
n	20	20	20
s_y	2.89	2.85	3.45
s_e	0	0.631	2.04
r^2	1	0.95	0.65

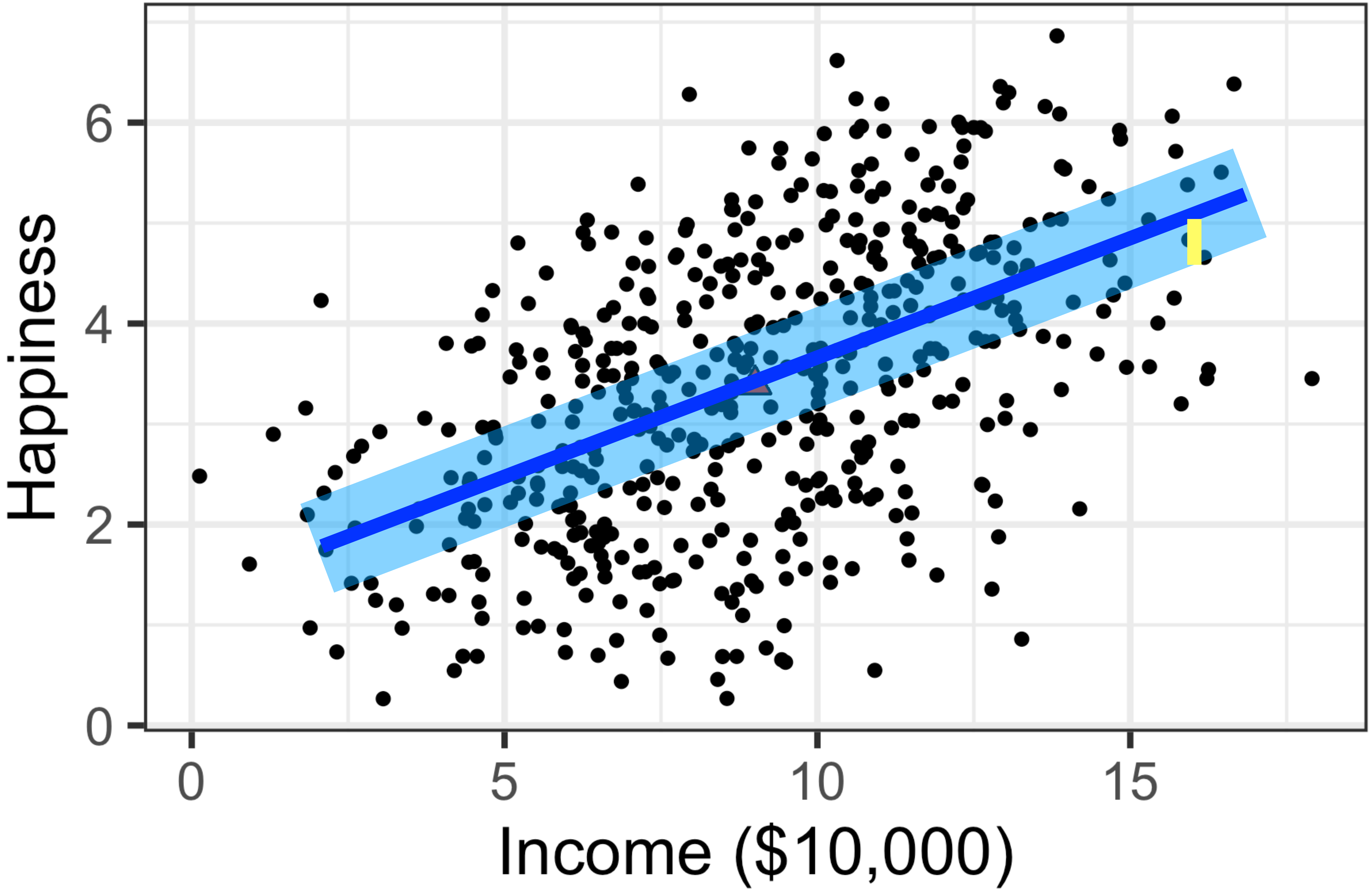
$$s_e = \sqrt{\frac{SS(resid)}{n-2}}$$

$$r^2 = 1 - \frac{s_e^2}{s_y^2}$$



$SS(resid)$	0	7.98	83.1
n	20	20	20
s_y	2.89	2.85	3.45
s_e	0	0.631	2.04
r^2	1 (1)	0.95 (0.949)	0.65 (0.634)

The coefficient of determination, r^2




```
cor(income, happiness)
```

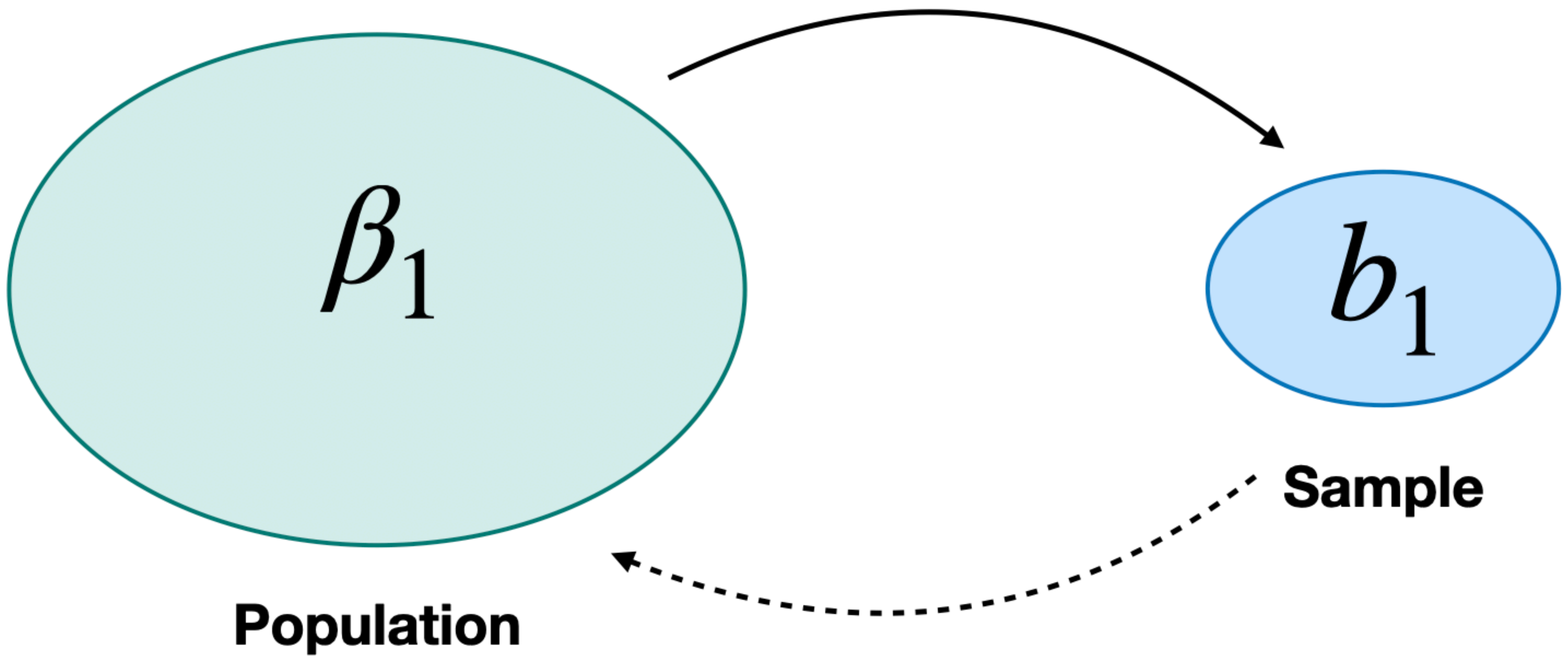
```
[1] 0.5093659
```

$$r^2 = 0.5093^2$$

$$r^2 = 0.259$$

25.9% of the variance in happiness is explained by the relationship between happiness and income

Interpreting the linear model



- **Estimate**

  ($\bar{y}$ for $\mu \rightarrow \bar{b}_1$ for β_1)

- **Error of the estimate**

  $(SE_{\bar{y}} \rightarrow SE_{\bar{b}_1})$

- **Confidence interval**

  $(\bar{y} \pm t_{0.025} SE_{\bar{y}} \rightarrow \bar{b}_1 \pm t_{0.025} SE_{\bar{b}_1})$

- **Hypothesis testing**

  $(H_0 : \mu = 0 \rightarrow H_0 : \beta_1 = 0)$

Testing the hypothesis $H_0 : \beta_1 = 0$

- We can calculate all this by hand, but we are not going to… instead we will take advantage of the power of R and focus on how to interpret its output

Test statistic:

$$t_s = \frac{b_1 - 0}{SE_{b_1}}$$

$$(df = n - 2)$$

Standard error of b_1

$$SE_{b_1} = \frac{s_e}{s_x\sqrt{n-1}}$$

Understand degrees of freedom, but the other math is not essential to memorize!!!

Testing the hypothesis $H_0 : \beta_1 = 0$

```
> summary(lm(happiness ~ income, data = income_data))

Call:
lm(formula = happiness ~ income, data = income_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5033 -0.8873  0.0459  0.9218  3.1279 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.34291    0.16505   8.136  3.3e-15 ***
income       0.22765    0.01727  13.182  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.234 on 496 degrees of freedom
Multiple R-squared:  0.2595,	Adjusted R-squared:  0.258 
F-statistic: 173.8 on 1 and 496 DF,  p-value: < 2.2e-16
```

Testing the hypothesis $H_0 : \beta_1 = 0$

```
> summary(lm(happiness ~ income, data = income_data))

Call:
lm(formula = happiness ~ income, data = income_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5033 -0.8873  0.0459  0.9218  3.1279 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.34291    0.16505   8.136  3.3e-15 ***
income       0.22765    0.01727  13.182  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.234 on 496 degrees of freedom
Multiple R-squared:  0.2595,	Adjusted R-squared:  0.258 
F-statistic: 173.8 on 1 and 496 DF,  p-value: < 2.2e-16
```

Repeat the formula

Testing the hypothesis $H_0 : \beta_1 = 0$

```
> summary(lm(happiness ~ income, data = income_data))

Call:
lm(formula = happiness ~ income, data = income_data)

Residuals:
    Min      1Q  Median      3Q     Max
-3.5033 -0.8873  0.0459  0.9218  3.1279

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.34291    0.16505   8.136  3.3e-15 ***
income       0.22765    0.01727  13.182  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.234 on 496 degrees of freedom
Multiple R-squared:  0.2595,	Adjusted R-squared:  0.258
F-statistic: 173.8 on 1 and 496 DF,  p-value: < 2.2e-16
```

Residuals (come back to)

Testing the hypothesis $H_0 : \beta_1 = 0$

```
> summary(lm(happiness ~ income, data = income_data))

Call:
lm(formula = happiness ~ income, data = income_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5033 -0.8873  0.0459  0.9218  3.1279 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.34291    0.16505   8.136  3.3e-15 ***
income       0.22765    0.01727  13.182  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.234 on 496 degrees of freedom
Multiple R-squared:  0.2595,    Adjusted R-squared:  0.258 
F-statistic: 173.8 on 1 and 496 DF,  p-value: < 2.2e-16
```

Slope and intercept

β_0

β_1

$H = 1.34 + 0.227(I)$

Testing the hypothesis $H_0 : \beta_1 = 0$

```
> summary(lm(happiness ~ income, data = income_data))

Call:
lm(formula = happiness ~ income, data = income_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5033 -0.8873  0.0459  0.9218  3.1279 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.34291    0.16505   8.136  3.3e-15 ***
income       0.22765    0.01727  13.182  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.234 on 496 degrees of freedom
Multiple R-squared:  0.2595,	Adjusted R-squared:  0.258 
F-statistic: 173.8 on 1 and 496 DF,  p-value: < 2.2e-16
```

SE_{b_1}

How much variation

β_0

β_1

$H = 1.34 + 0.227(I)$

Testing the hypothesis $H_0 : \beta_1 = 0$

```
> summary(lm(happiness ~ income, data = income_data))

Call:
lm(formula = happiness ~ income, data = income_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5033 -0.8873  0.0459  0.9218  3.1279 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.34291    0.16505   8.136  3.3e-15 ***
income       0.22765    0.01727  13.182  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.234 on 496 degrees of freedom
Multiple R-squared:  0.2595,	Adjusted R-squared:  0.258 
F-statistic: 173.8 on 1 and 496 DF,  p-value: < 2.2e-16
```

SE_{b_1}

Hypothesis testing

β_0

β_1

$H = 1.34 + 0.227(I)$

Not usually interested in testing $H_0 : \beta_0 = 0$

Testing the hypothesis $H_0 : \beta_1 = 0$

```
> summary(lm(happiness ~ income, data = income_data))

Call:
lm(formula = happiness ~ income, data = income_data)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.5033 -0.8873  0.0459  0.9218  3.1279 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.34291    0.16505   8.136  3.3e-15 ***
income       0.22765    0.01727  13.182  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.234 on 496 degrees of freedom
Multiple R-squared:  0.2595,	Adjusted R-squared:  0.258 
F-statistic: 173.8 on 1 and 496 DF,  p-value: < 2.2e-16
```

s_e

$(cor)^2$

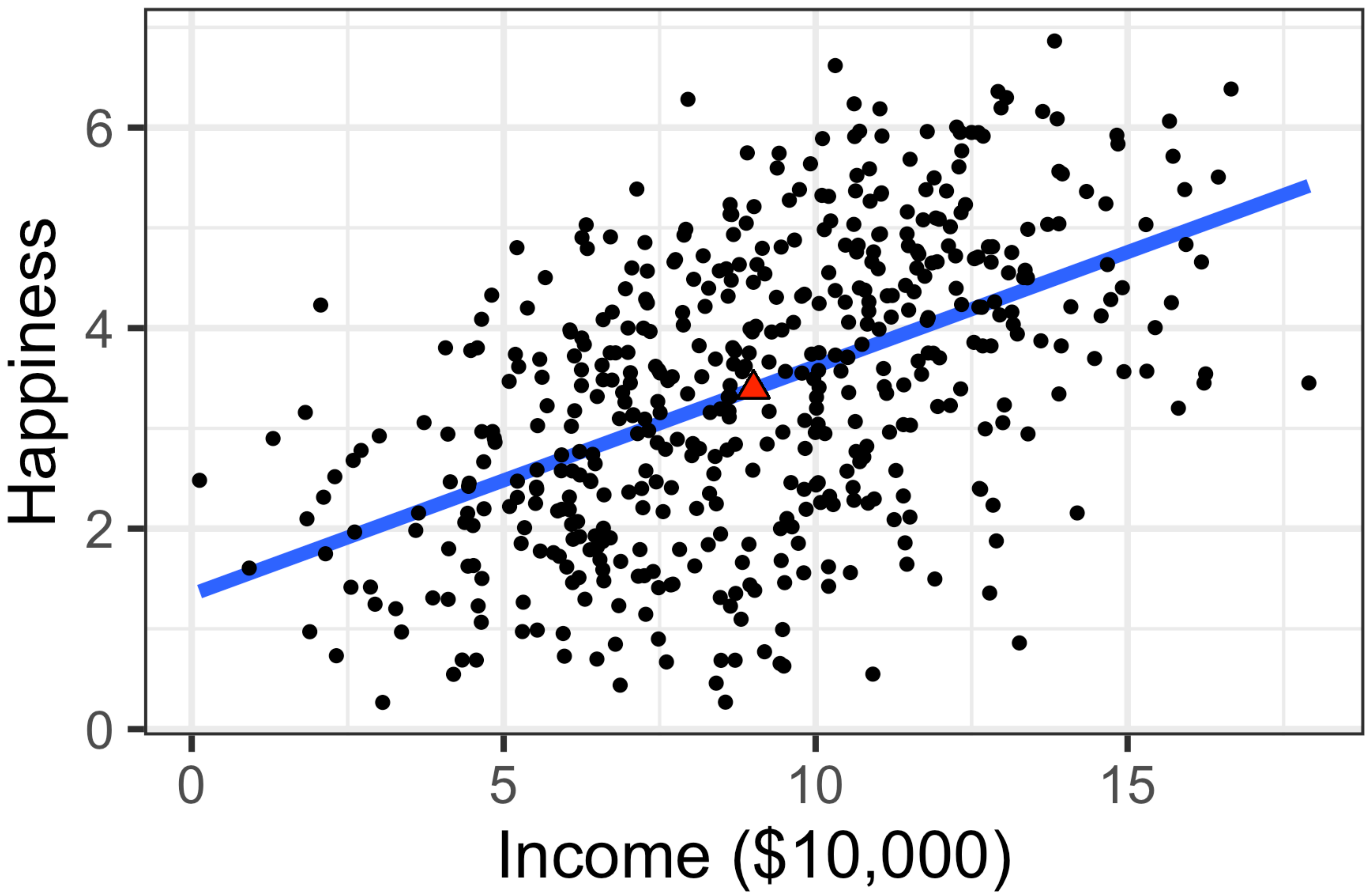
General fit parameters

$1 - \frac{s_e^2}{s_y^2}$

Testing the hypothesis $H_0 : \beta_1 = 0$

- Note: test on β_1 does not ask *whether* the relationship between X and Y is linear, rather *assuming* a linear relationship, is the slope nonzero?
- Directional or non-directional tests
- Beware of curvilinearity, outliers, and influential points (just like with correlation)

Reporting the results

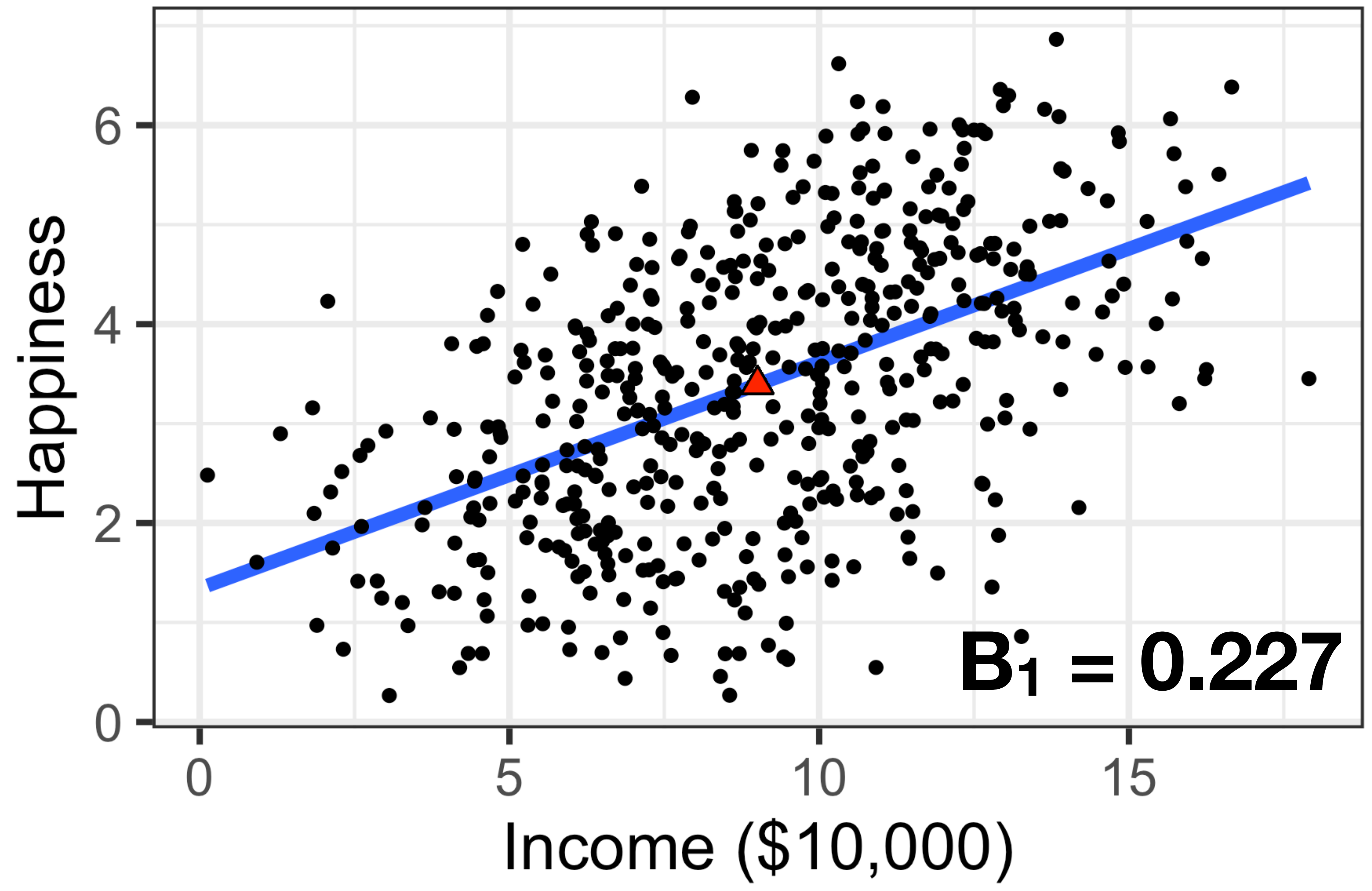


A simple linear regression was performed to test if income significantly predicted happiness. We found that income did significantly predict happiness with a 0.227 unit increase in happiness for every $10,000 increase in income ($B_1 = 0.227 \pm 0.017$, p-value < 2e-16). The model ($Y = 0.227X + 1.34$) explained 25.8% of the total variation in happiness in this study.

Regression vs. correlation

Regression: Describes how one variable (x) affects another variable (y)

Correlation: Quantifies the direction and strength of the relationship between two variables (x and y)




```
cor(happiness, income)
```

0.509

$(0.509^2 = R^2 = 0.259)$

Regression vs. correlation

Regression: Describes how one variable (x) affects another variable (y)

Correlation: Quantifies the direction and strength of the relationship between two variables (x and y)

	Correlation	Regression
When to use	When summarizing direct relationship between two variables	To predict or explain responses
Able to quantify direction of relationship?		
Able to quantify strength of relationship?		
Able to show cause and effect?		
Able to predict and optimize?		
X and Y are interchangeable?		

Regression vs. correlation

Regression: Describes how one variable (x) affects another variable (y)

Correlation: Quantifies the direction and strength of the relationship between two variables (x and y)

	Correlation	Regression
When to use	When summarizing direct relationship between two variables	To predict or explain responses
Able to quantify direction of relationship?	Yes ✓	Yes ✓
Able to quantify strength of relationship?		
Able to show cause and effect?		
Able to predict and optimize?		
X and Y are interchangeable?		

Regression vs. correlation

Regression: Describes how one variable (x) affects another variable (y)

Correlation: Quantifies the direction and strength of the relationship between two variables (x and y)

	Correlation	Regression
When to use	When summarizing direct relationship between two variables	To predict or explain responses
Able to quantify direction of relationship?	Yes	Yes
Able to quantify strength of relationship?	Yes	Yes
Able to show cause and effect?		
Able to predict and optimize?		
X and Y are interchangeable?		

Regression vs. correlation

Regression: Describes how one variable (x) affects another variable (y)

Correlation: Quantifies the direction and strength of the relationship between two variables (x and y)

	Correlation	Regression
When to use	When summarizing direct relationship between two variables	To predict or explain responses
Able to quantify direction of relationship?	Yes ✓	Yes ✓
Able to quantify strength of relationship?	Yes ✓	Yes ✓
Able to show cause and effect?	No ✗	Yes ✓
Able to predict and optimize?		
X and Y are interchangeable?		

Regression vs. correlation

Regression: Describes how one variable (x) affects another variable (y)

Correlation: Quantifies the direction and strength of the relationship between two variables (x and y)

	Correlation	Regression
When to use	When summarizing direct relationship between two variables	To predict or explain responses
Able to quantify direction of relationship?	Yes ✓	Yes ✓
Able to quantify strength of relationship?	Yes ✓	Yes ✓
Able to show cause and effect?	No ✗	Yes ✓
Able to predict and optimize?	No ✗	Yes ✓
X and Y are interchangeable?		

Regression vs. correlation

Regression: Describes how one variable (x) affects another variable (y)

Correlation: Quantifies the direction and strength of the relationship between two variables (x and y)

	Correlation	Regression
When to use	When summarizing direct relationship between two variables	To predict or explain responses
Able to quantify direction of relationship?	Yes ✓	Yes ✓
Able to quantify strength of relationship?	Yes ✓	Yes ✓
Able to show cause and effect?	No ✗	Yes ✓
Able to predict and optimize?	No ✗	Yes ✓
X and Y are interchangeable?	Yes ✓	No ✗

R example

regression_example_blank.R

Source on Save Run Source

```r
# Linear model demo

# LOAD AND INSPECT STOPPING DISTANCE DATA





# COMPUTING REGRESSION LINE BY HAND





# plot regression line
# a = intercept, b = slope


# use lm()




# look at summary
```