Winter Workshop:
Basic and Intermediate Statistics with R

Chai Jun Ho, PhD

CHILD. LANGUAGE. LAB.

CHOSUN UNIVERSITY 1946
조선대학교
CHOSUN UNIVERSITY

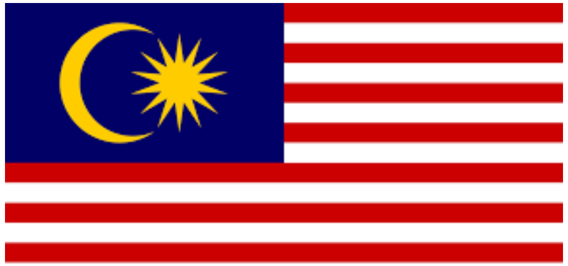
Download R and RStudio

R:

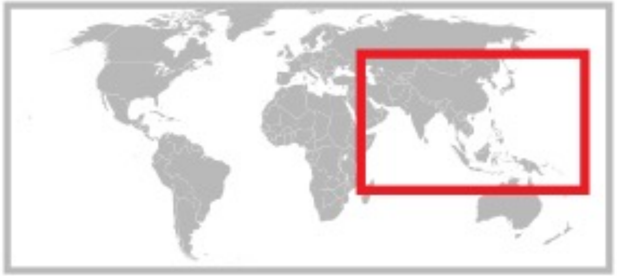
https://cran.r-project.org/bin/windows/base/

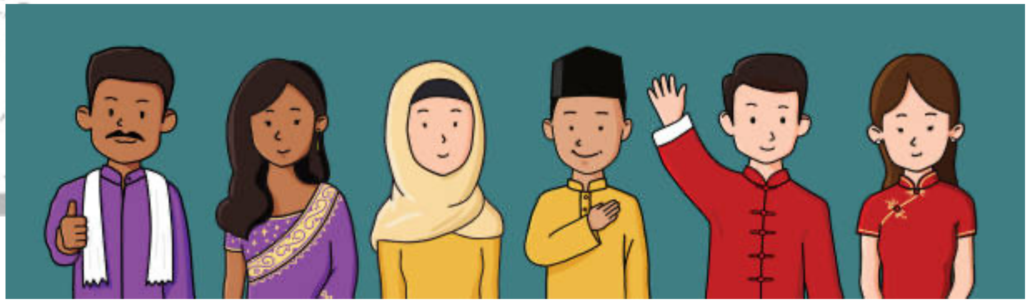
RStudio

https://posit.co/products/open-source/rstudio/




Malaysian Cuisine
fried kway teow
curry puff
nasi lemak
laksa
curry noodle
pao
fish head curry
teh tarik
cendol
roti canai
satay
teh peng
grilled fish
kuih
shumai
mee rojak
chicken rice
kopi-o
ketupat
beef rendang





Schedule

Day 1	Day 2	Day 3
Introduction to Statistics	**Linear Model I**	**Linear Model II**
Introduction to R	**Data Visualization**	**Mixed-Effect Model**

Introduction to Statistics

Is there a relationship between **teaching style** and students' **learning**?

Introduction to Statistics

Is there a relationship between **teaching style** and students' **learning**?

Collect **data**…

Introduction to Statistics

Is there a relationship between **teaching style** and students' **learning**?

Collect **data**…

Conduct **statistical test**

Introduction to Statistics

Is there a relationship between **teaching style** and students' **learning**?

Collect **data**…

Conduct **statistical test**

Visualizing **results**

Introduction to Statistics

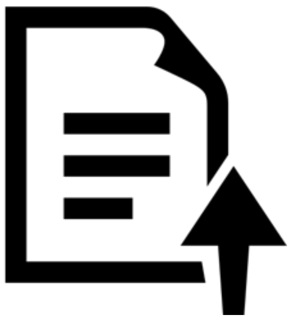
Is there a relationship between **teaching style** and students' **learning**?

Collect **data**…

Conduct **statistical test**

Visualizing **results**

Publish **findings**



Introduction to Statistics

Is there a relationship between **teaching style** and students' **learning**?

Collect **data**…

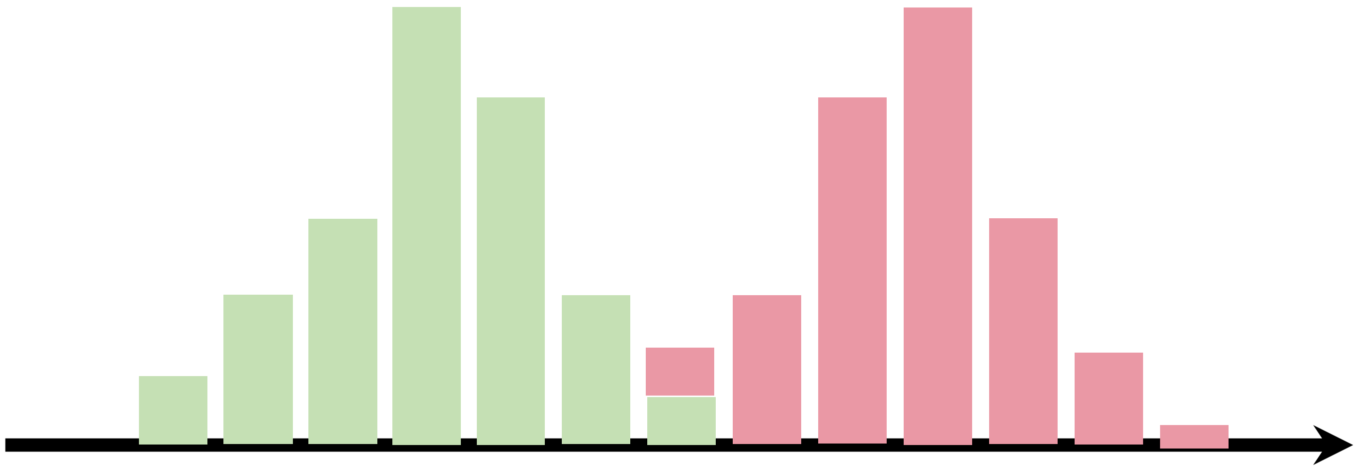
Conduct **statistical test**

Visualizing **results**

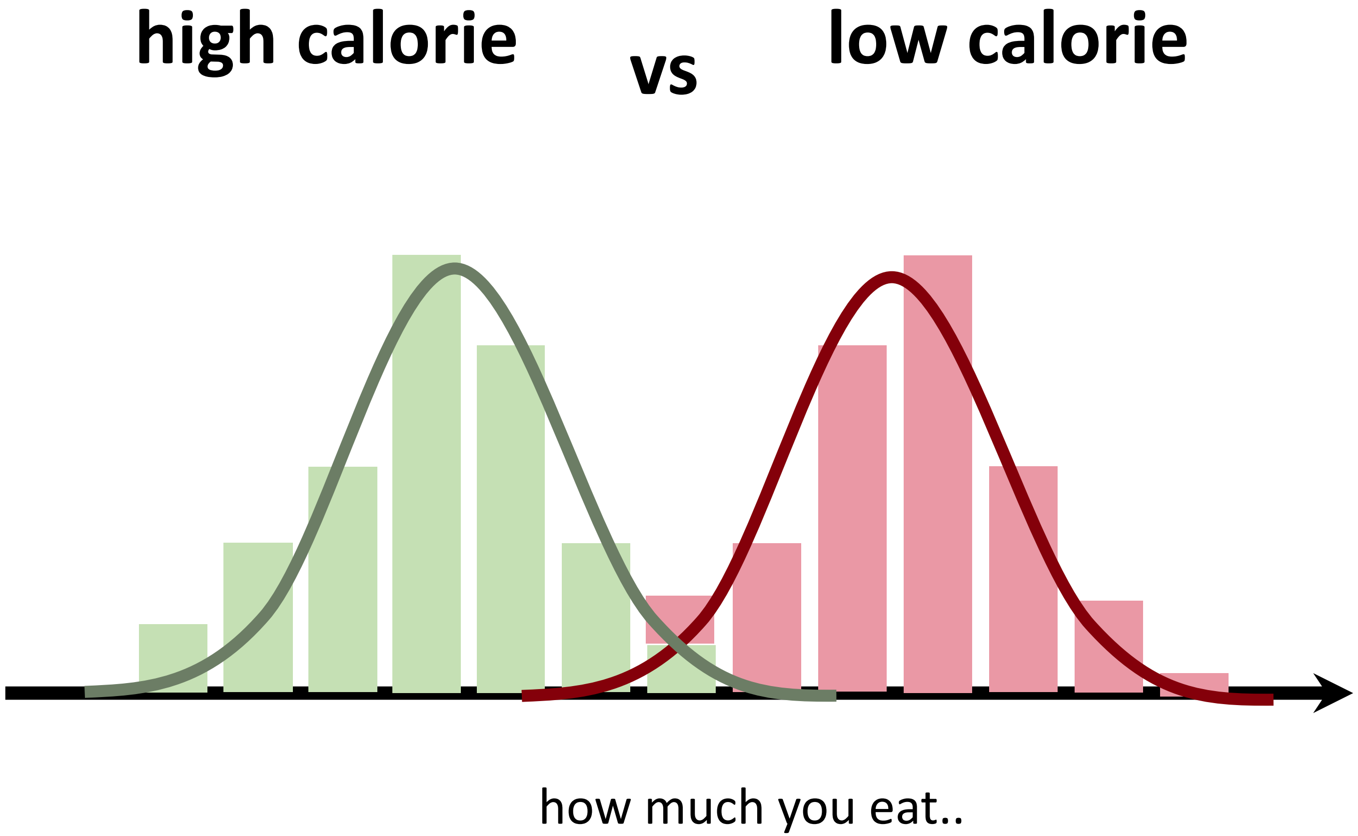
Publish **findings**

high calorie **vs** **low calorie**

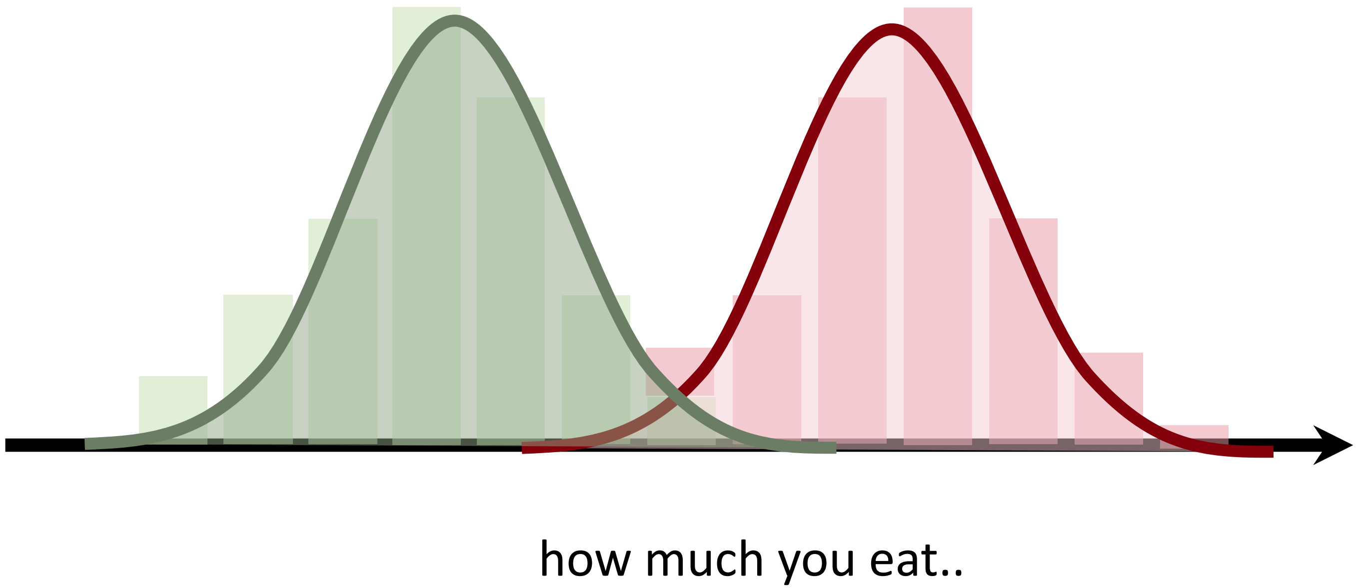
high calorie **vs** **low calorie**



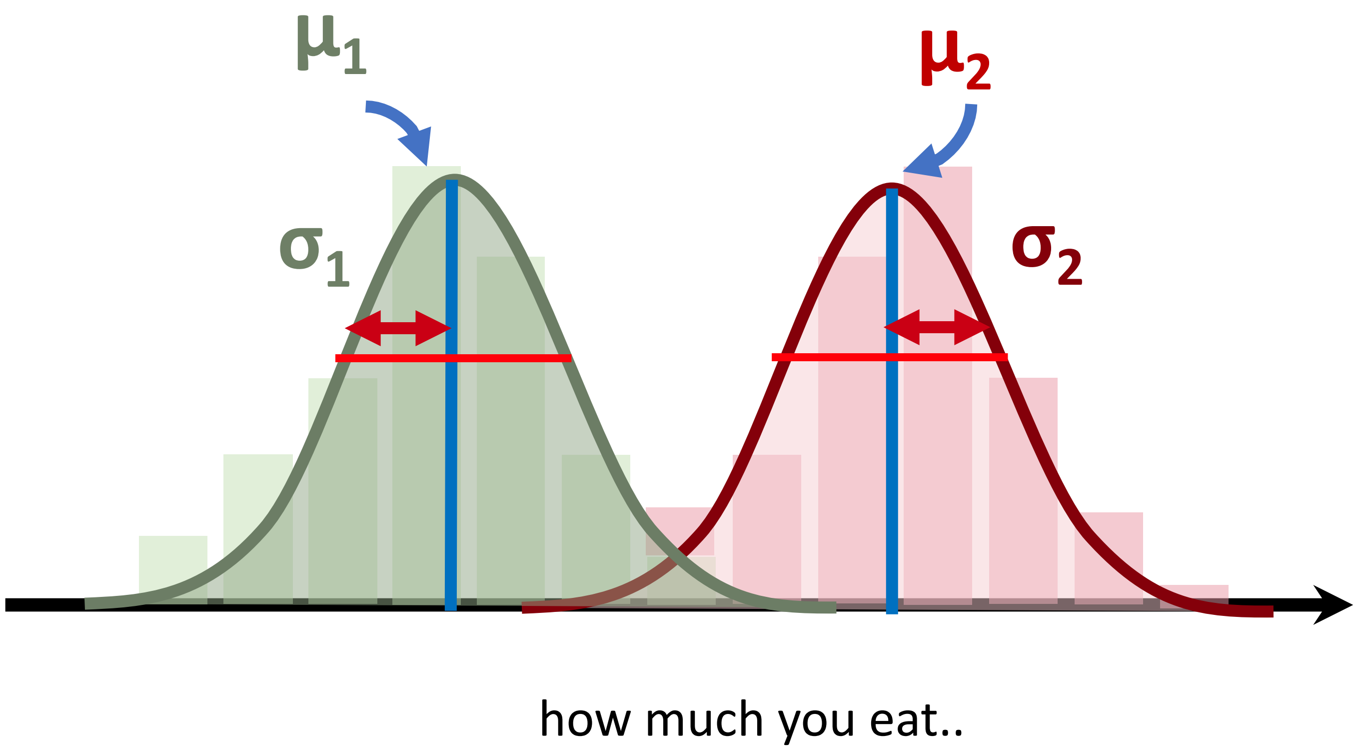
how much you eat..


high calorie
vs
low calorie
how much you eat..

high calorie vs low calorie

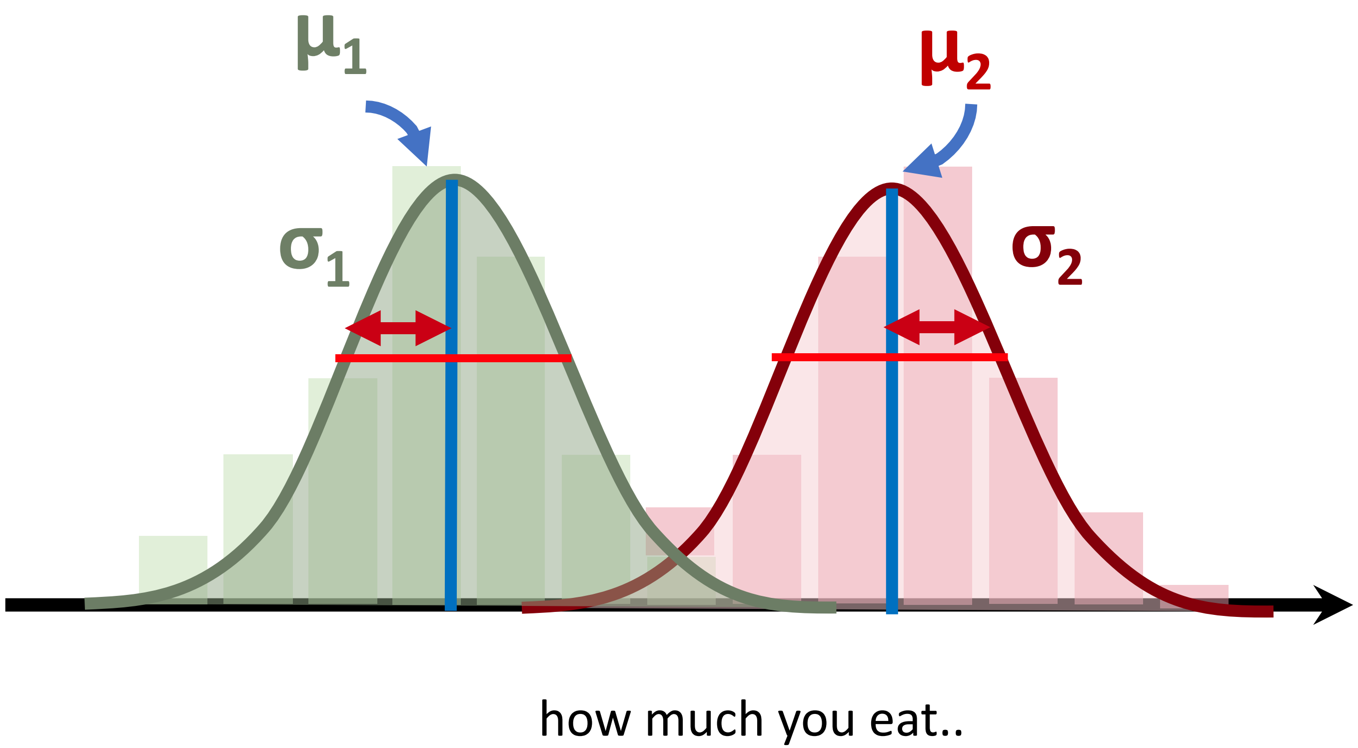


- **mean μ**
- **standard deviation σ**



Null Hypothesis, H_0: $\mu_1 = \mu_2$

Alternative Hypothesis, H_a: $\neg H_0$

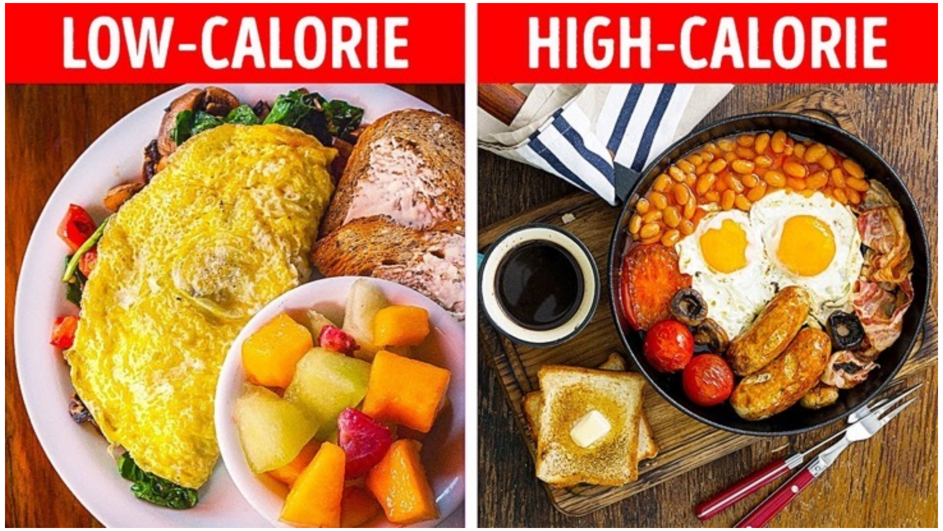


H_a: consumed different amount of food

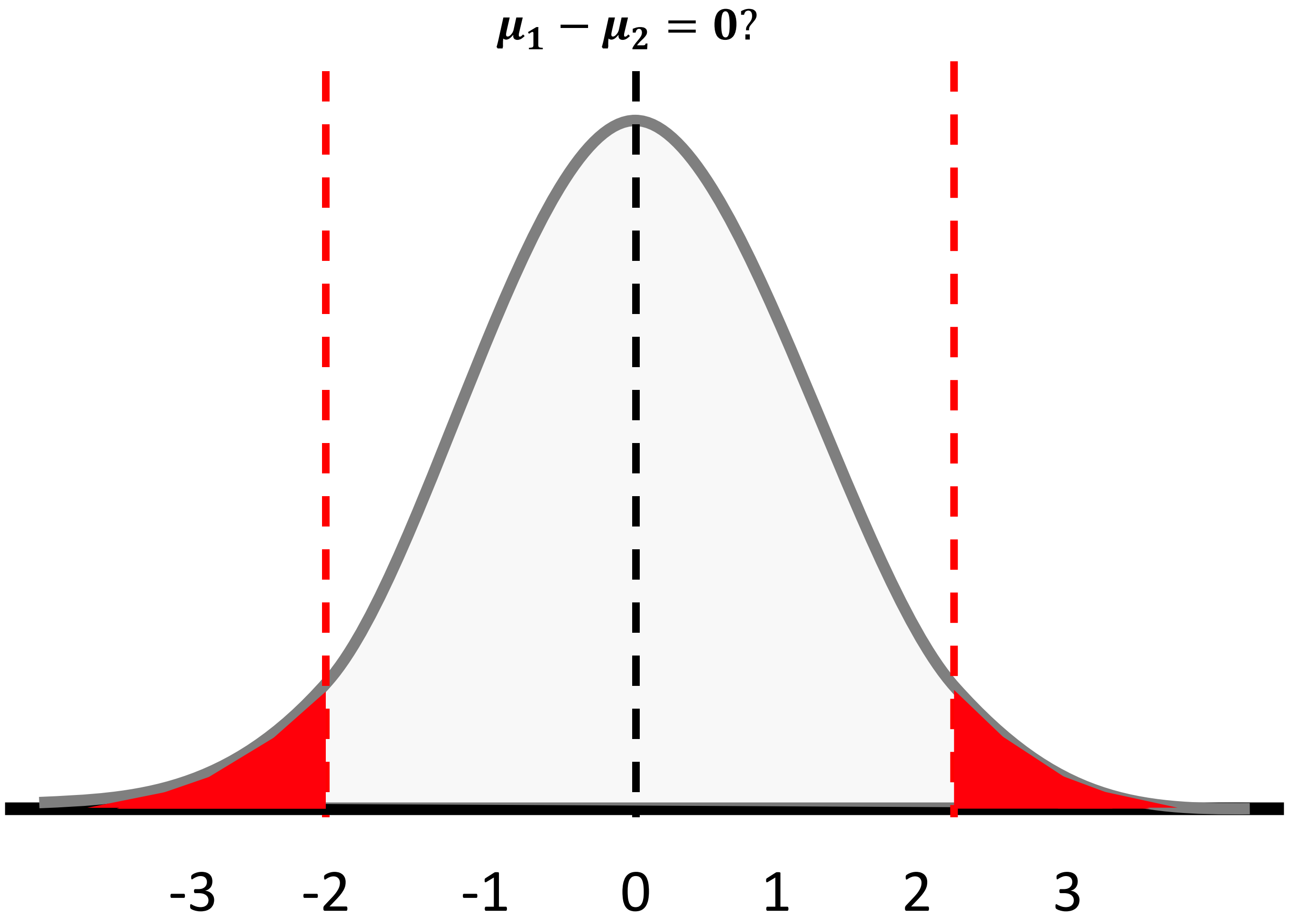
H_a: low calorie > high calorie

H_0: consumed similar amount of food

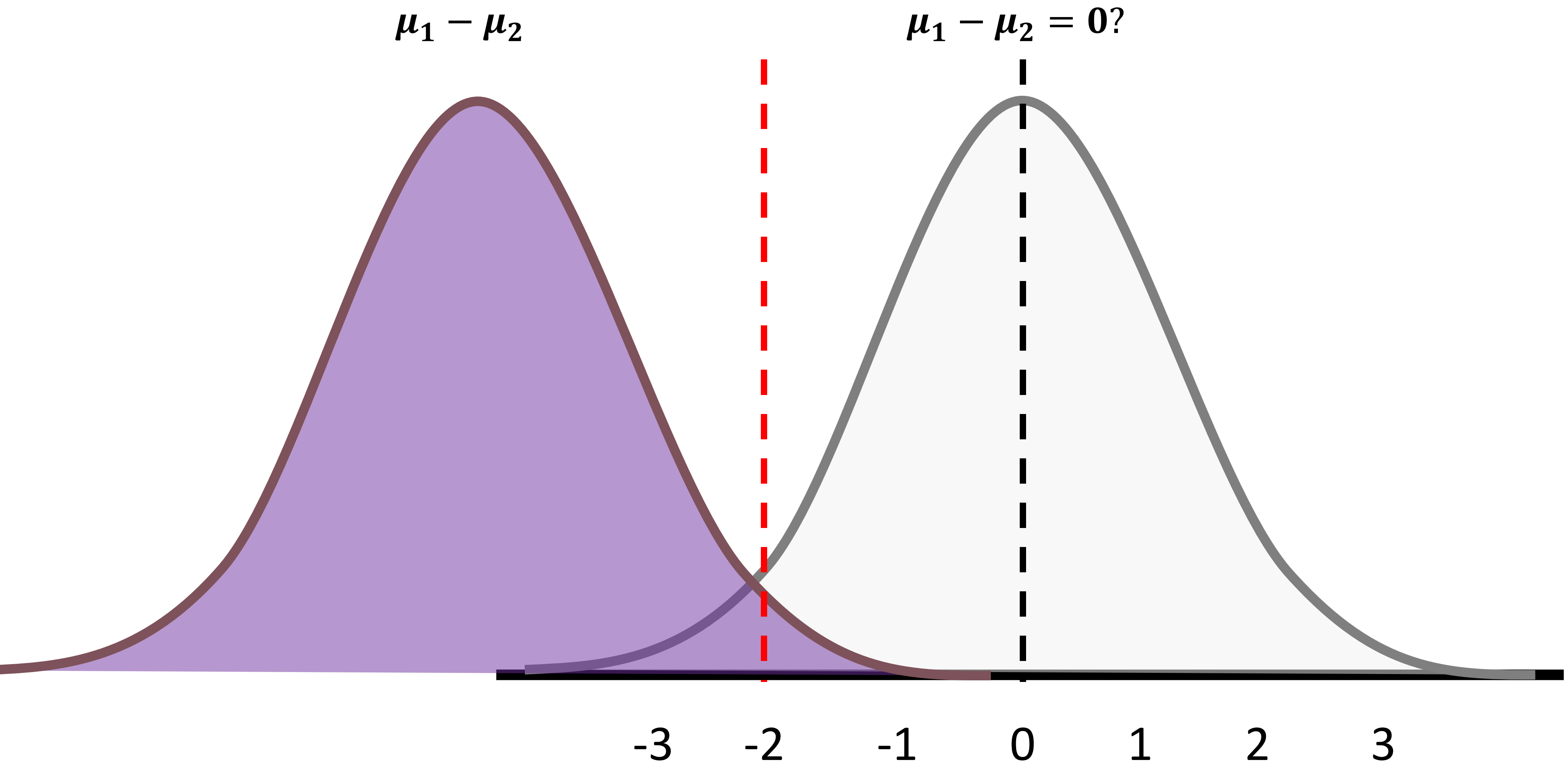
H_0: high calorie = low calorie

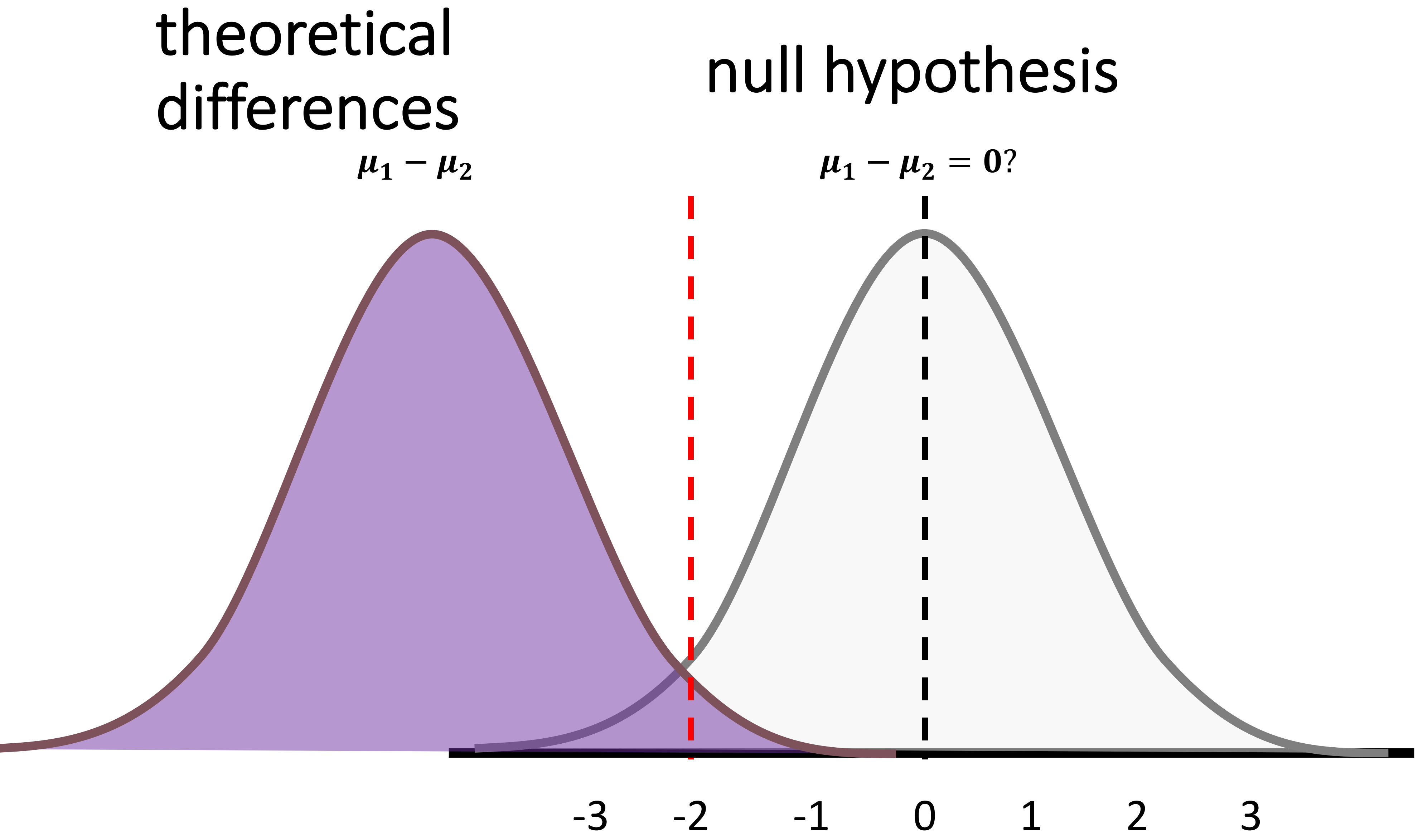


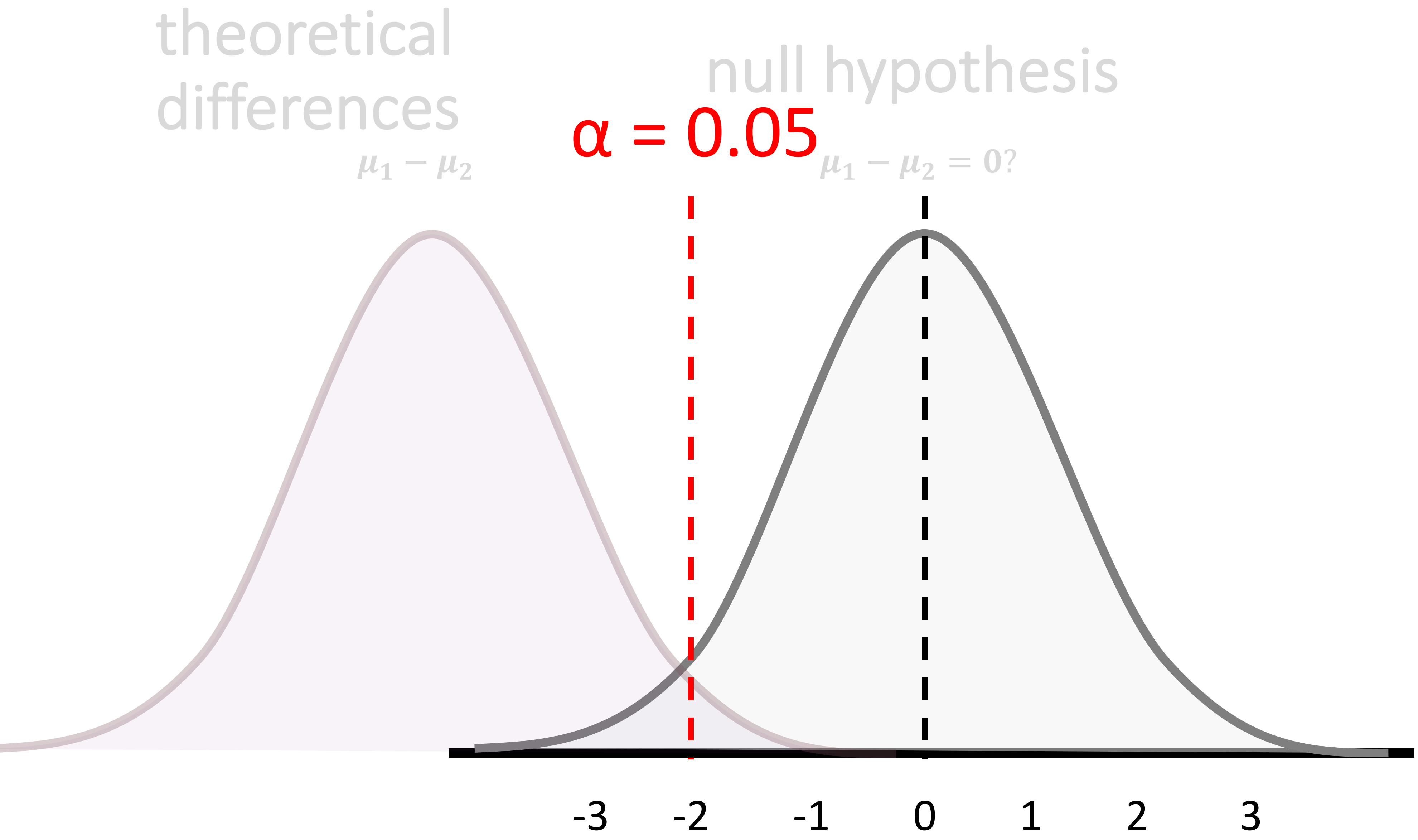
null hypothesis

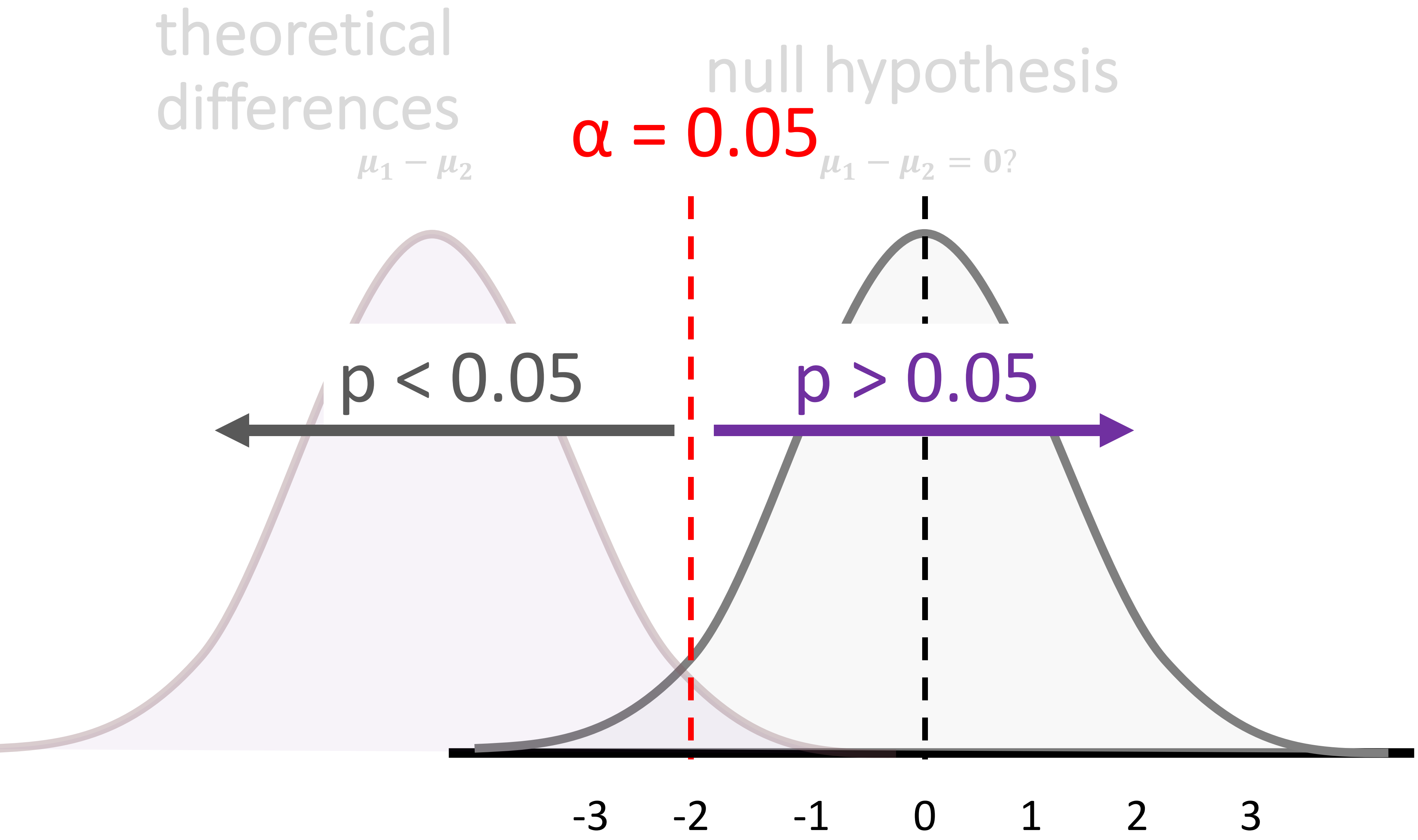


null hypothesis

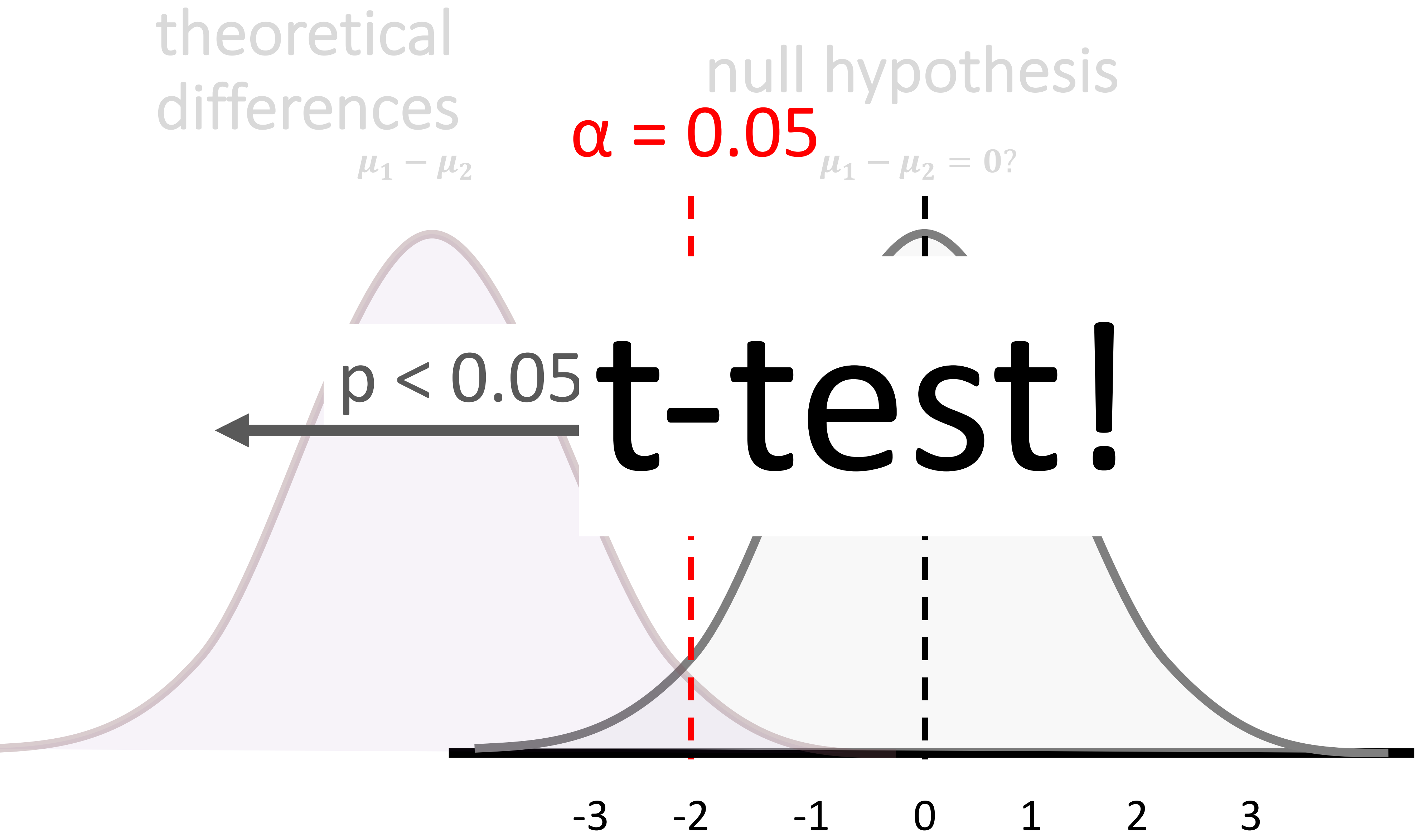



theoretical differences
null hypothesis
$\mu_1 - \mu_2$
$\mu_1 - \mu_2 = 0?$
-3
-2
-1
0
1
2
3


theoretical differences
null hypothesis
α = 0.05
$\mu_1 - \mu_2$
$\mu_1 - \mu_2 = 0?$
-3
-2
-1
0
1
2
3



theoretical differences
null hypothesis
α = 0.05
$\mu_1 - \mu_2$
$\mu_1 - \mu_2 = 0?$
p < 0.05
p > 0.05
-3
-2
-1
0
1
2
3



theoretical
differences
null hypothesis
α = 0.05
$\mu_1 - \mu_2$
$\mu_1 - \mu_2 = 0?$
p < 0.05
t-test!
-3
-2
-1
0
1
2
3

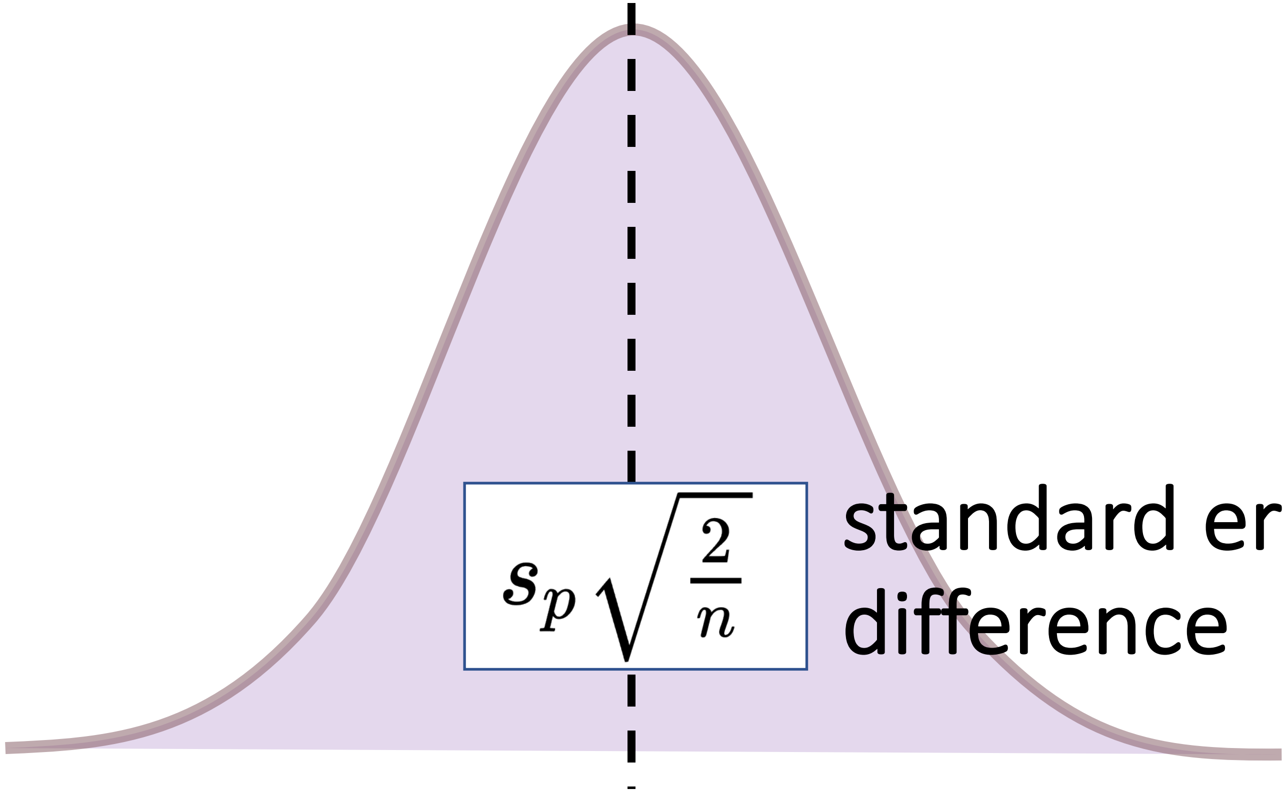
$\boldsymbol{\mu_1 - \mu_2}$?

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\boldsymbol{s}_p \sqrt{\frac{2}{n}}}$$

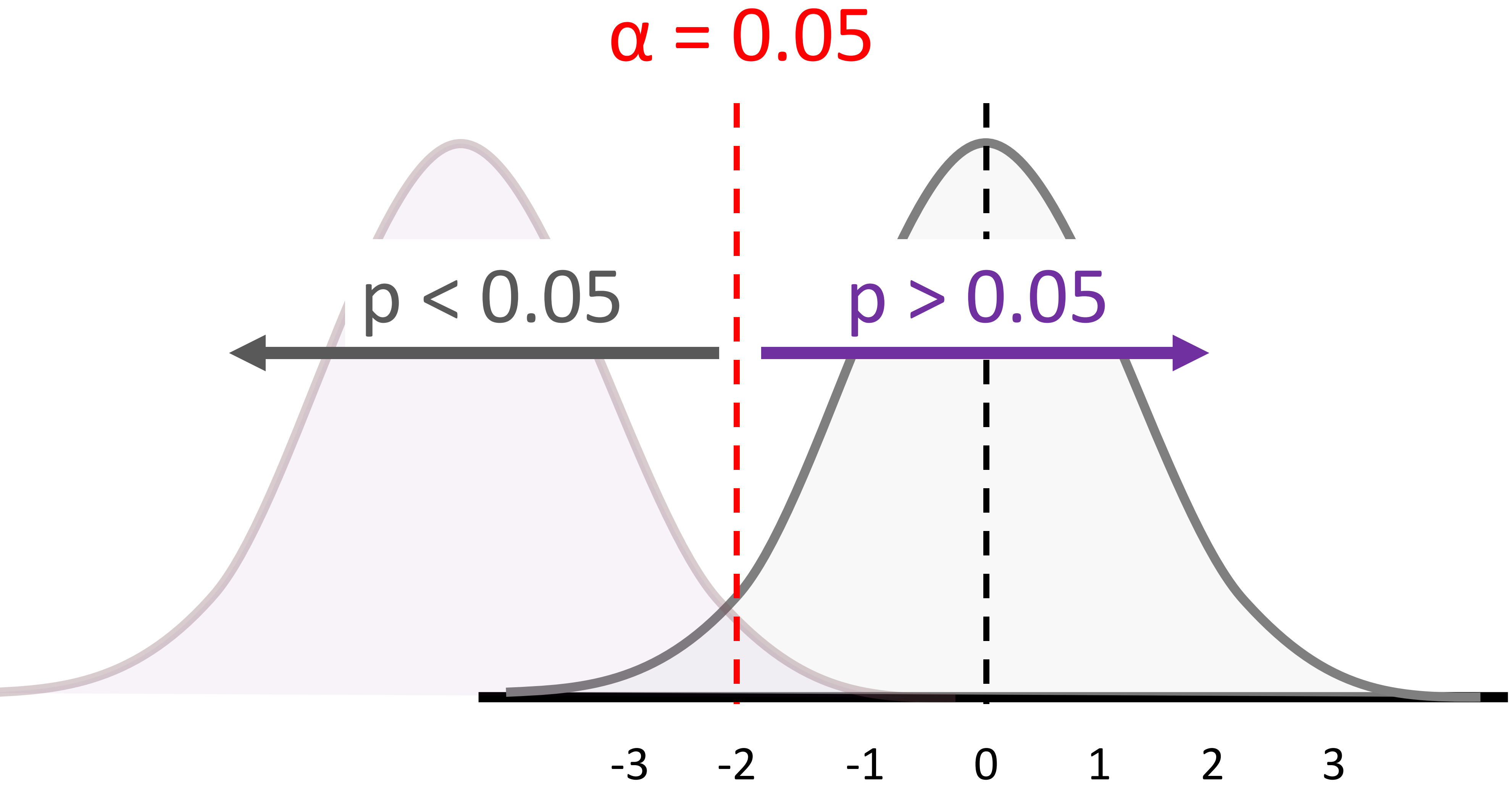
$\bar{X}$: sample mean

$\boldsymbol{s}_p \sqrt{\frac{2}{n}}$: standard error of the difference between two means observed differences

$$\bar{X}_1 - \bar{X}_2$$



standard error of the difference



α = 0.05
p < 0.05
p > 0.05
-3
-2
-1
0
1
2
3

TYPE I & II ERRORS

		Reality	
		H_0 is true	H_0 is false
Decision	Reject H_0	Wrong Type I Error	Correct
	Fail to reject H_0	Correct	Wrong Type II Error

TYPE I & II ERRORS

		Reality	
		H_0 is true	H_0 is false
Decision	Reject H_0		Correct
	Fail to reject H_0		

TYPE I & II ERRORS

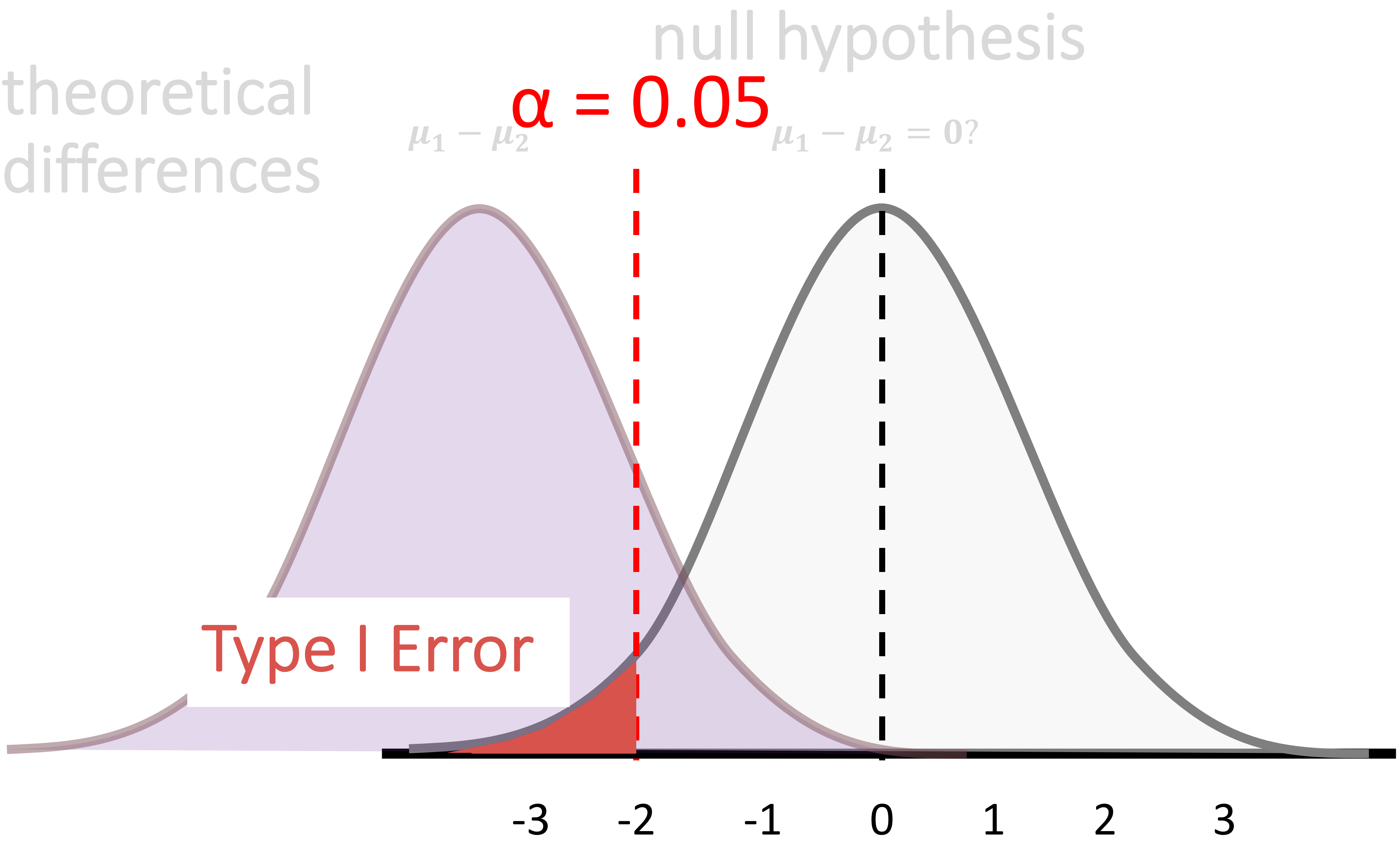
		Reality	
		H_0 is true	H_0 is false
Decision	Reject H_0	Wrong Type I Error	
	Fail to reject H_0		

TYPE I & II ERRORS

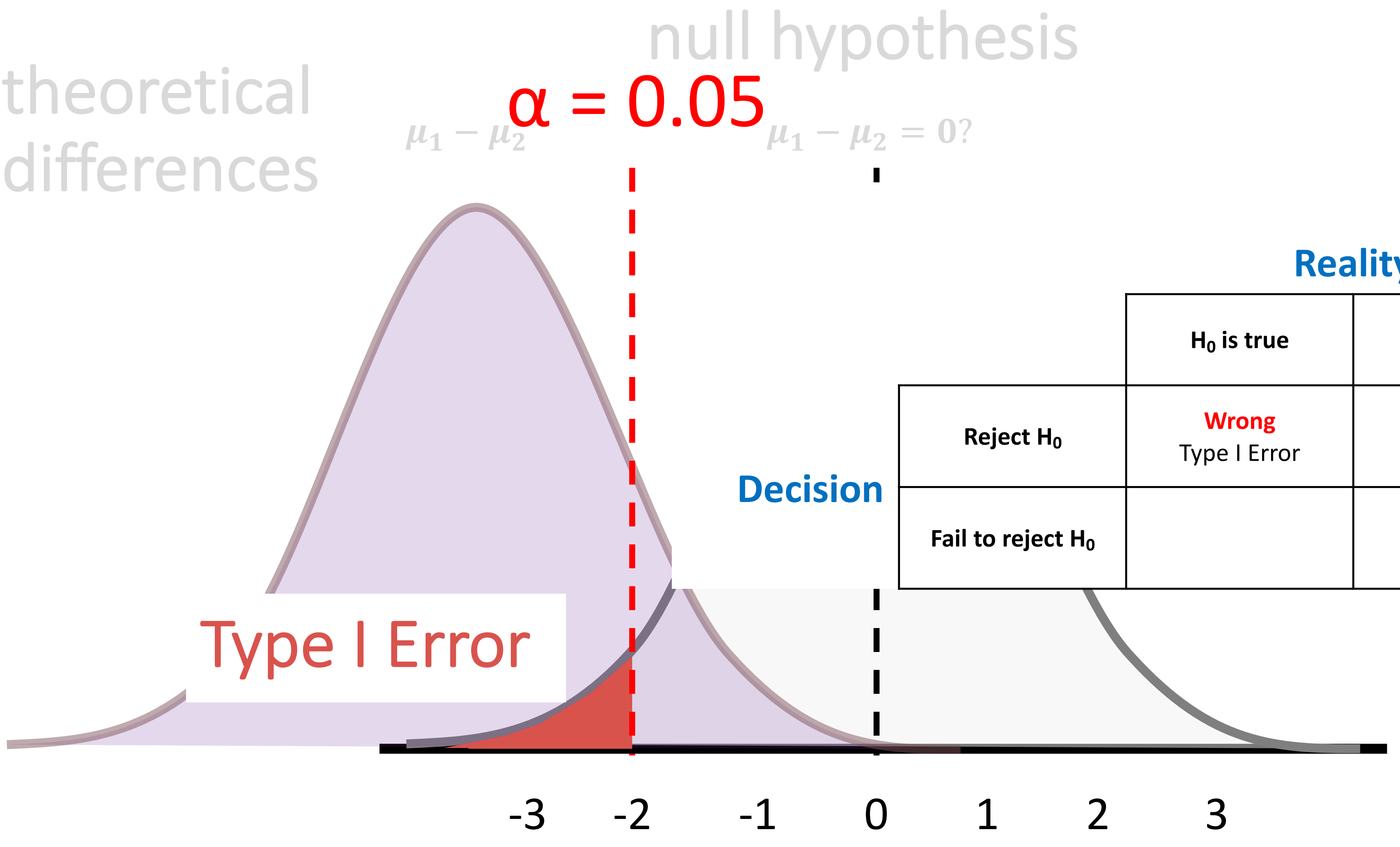
		Reality	
		H_0 is true	H_0 is false
Decision	Reject H_0		
	Fail to reject H_0	Correct	

TYPE I & II ERRORS

		Reality	
		H_0 is true	**H_0 is false**
Decision	**Reject H_0**		
	Fail to reject H_0		**Wrong** Type II Error



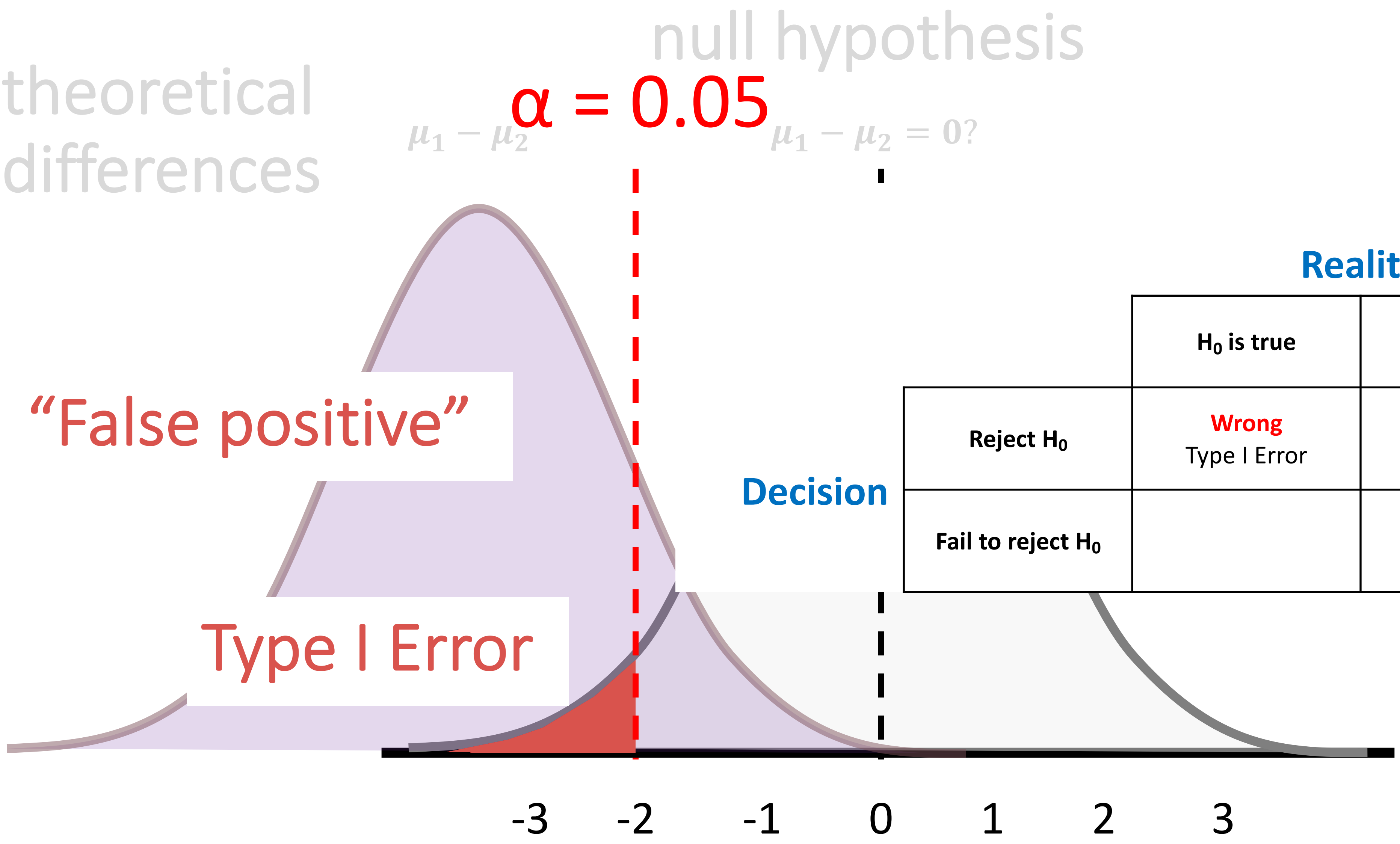
null hypothesis
theoretical differences
α = 0.05
$\mu_1 - \mu_2$
$\mu_1 - \mu_2 = 0?$
Type I Error
-3
-2
-1
0
1
2
3



Reality

Decision

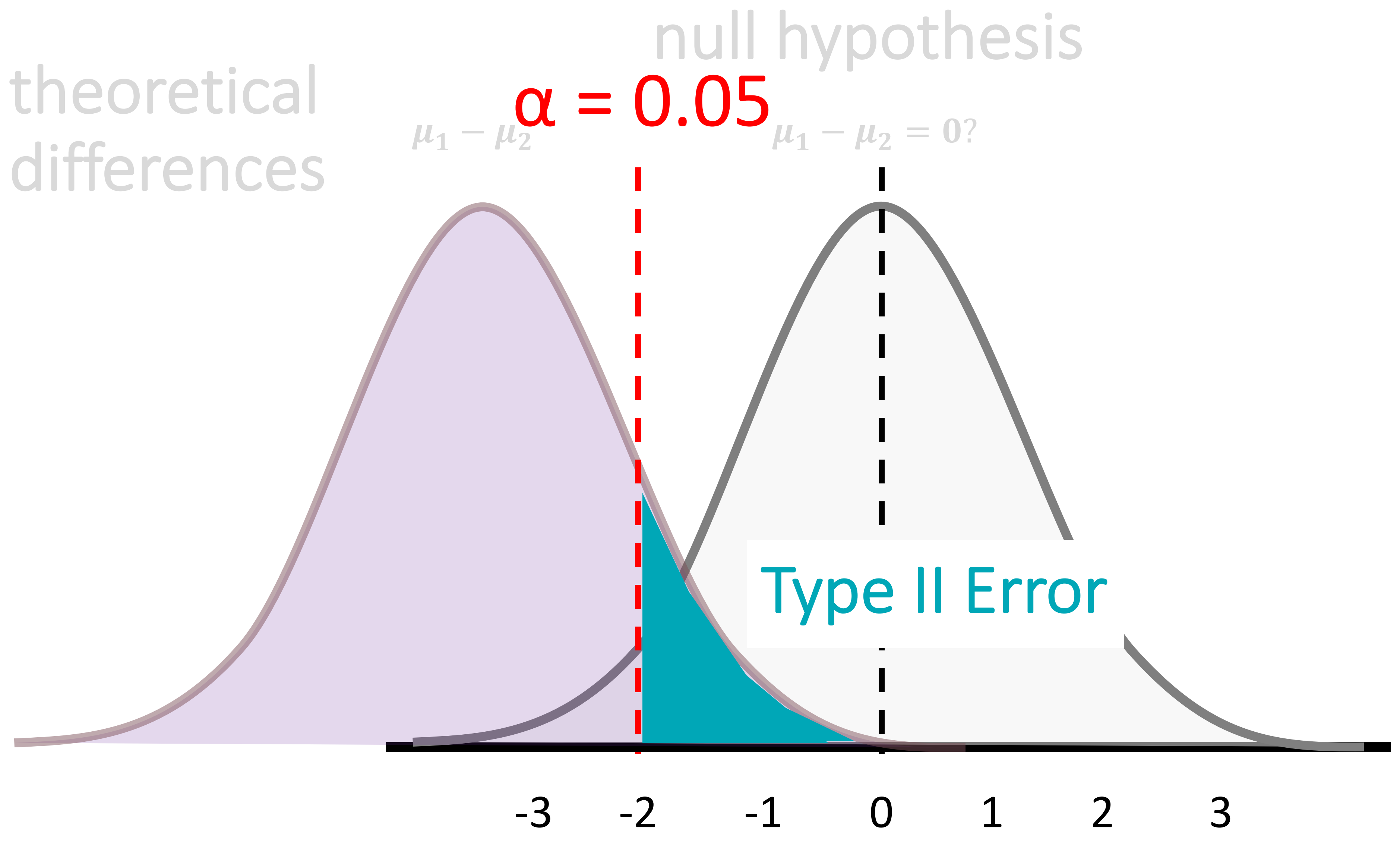
	H_0 is true	H_0 is false
Reject H_0	**Wrong** Type I Error	
Fail to reject H_0		



Reality

Decision

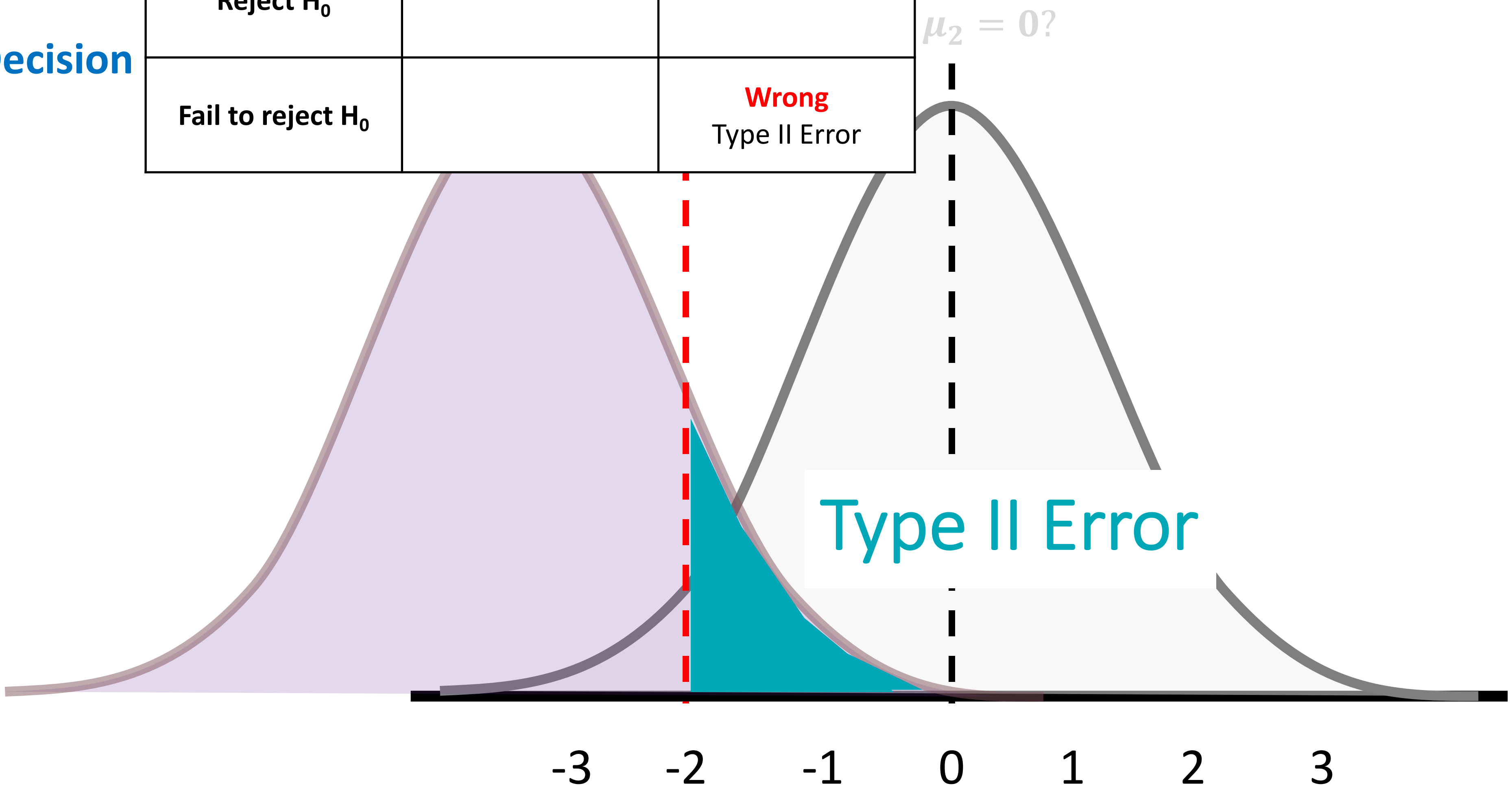
	H_0 is true	H_0 is false
Reject H_0	**Wrong** Type I Error	
Fail to reject H_0		


theoretical differences
$\mu_1 - \mu_2$
null hypothesis
$\mu_1 - \mu_2 = 0?$
α = 0.05
Type II Error
-3
-2
-1
0
1
2
3

Reality

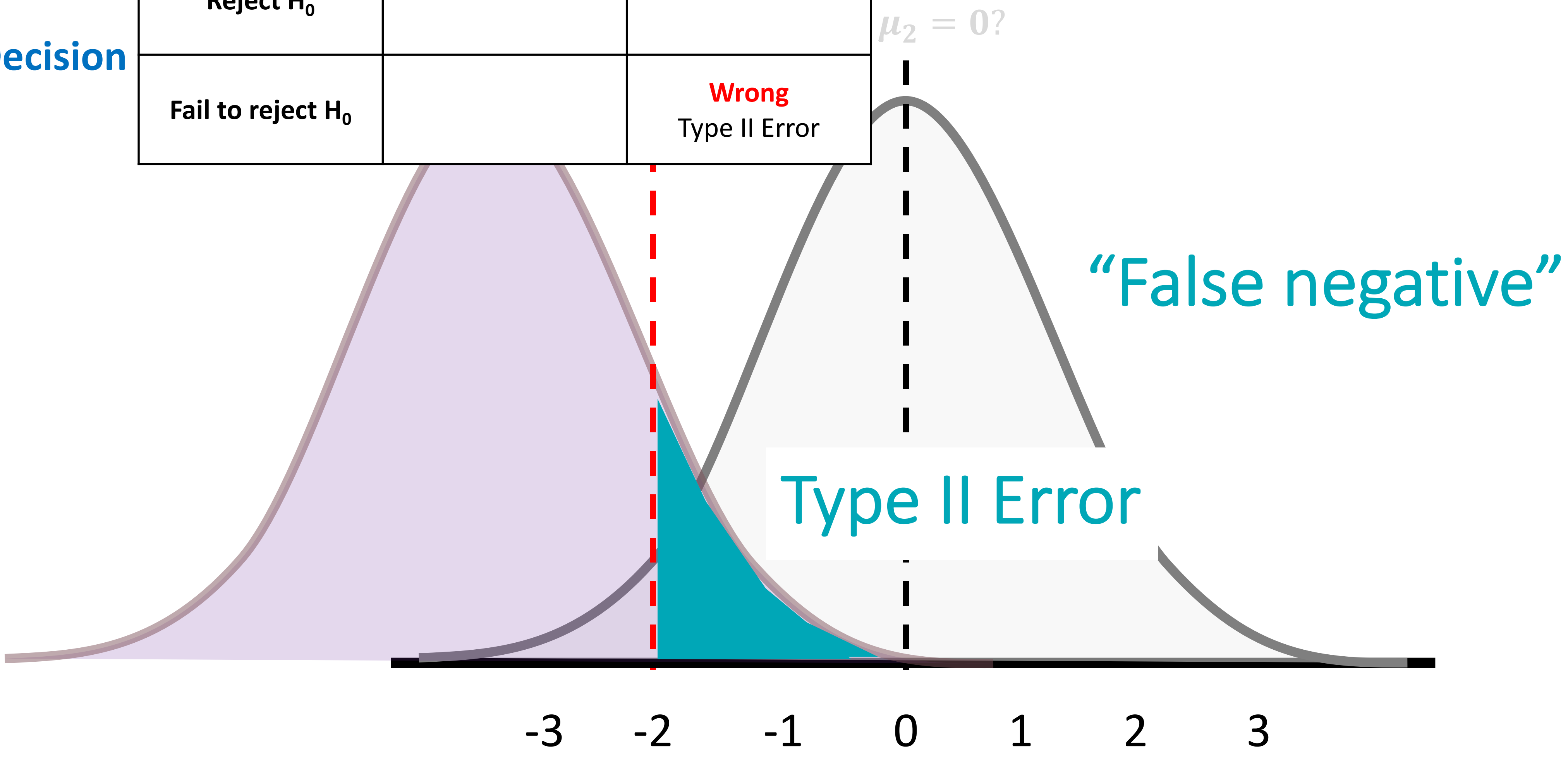
Decision

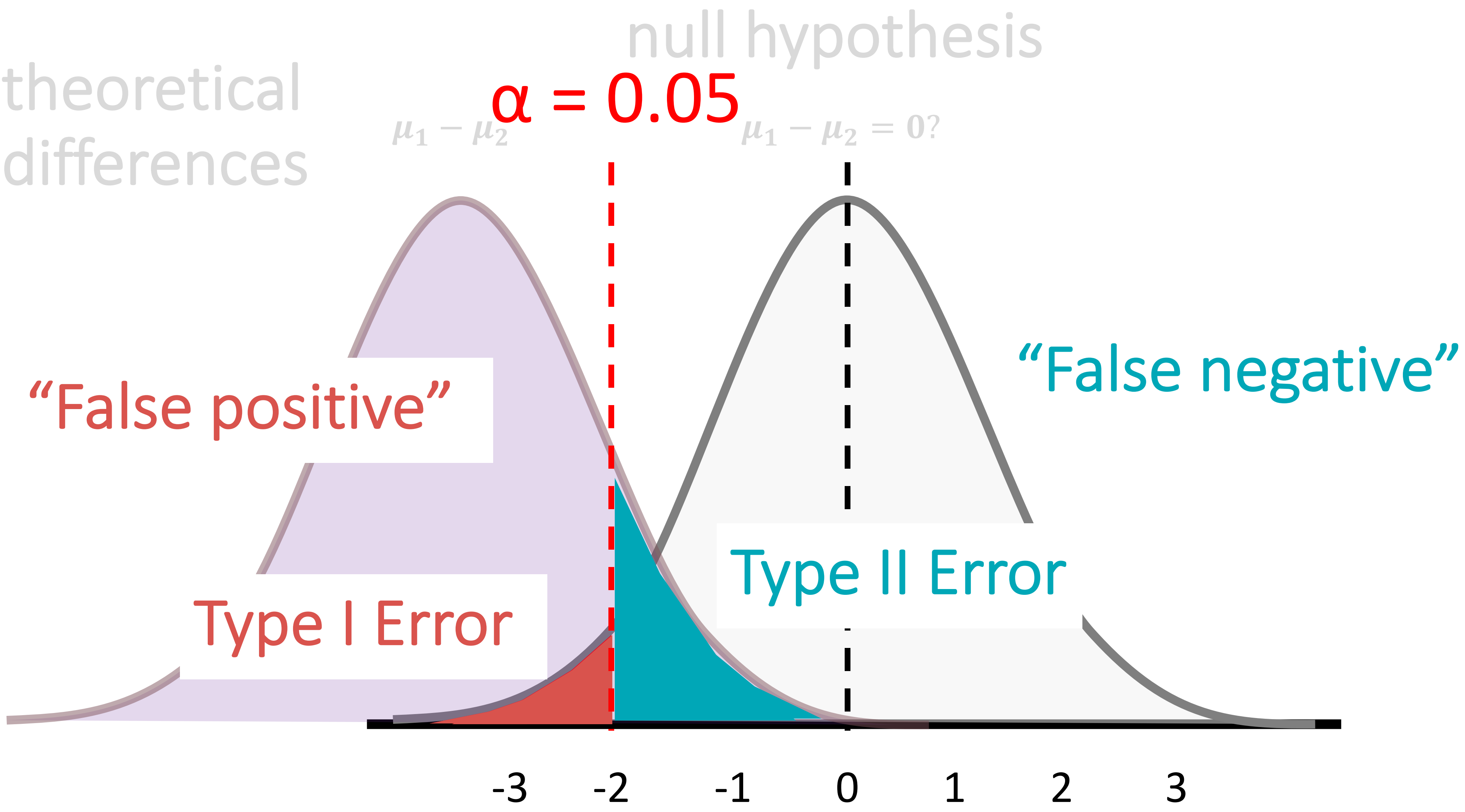
	H_0 is true	**H_0 is false**
Reject H_0		
Fail to reject H_0		**Wrong** Type II Error



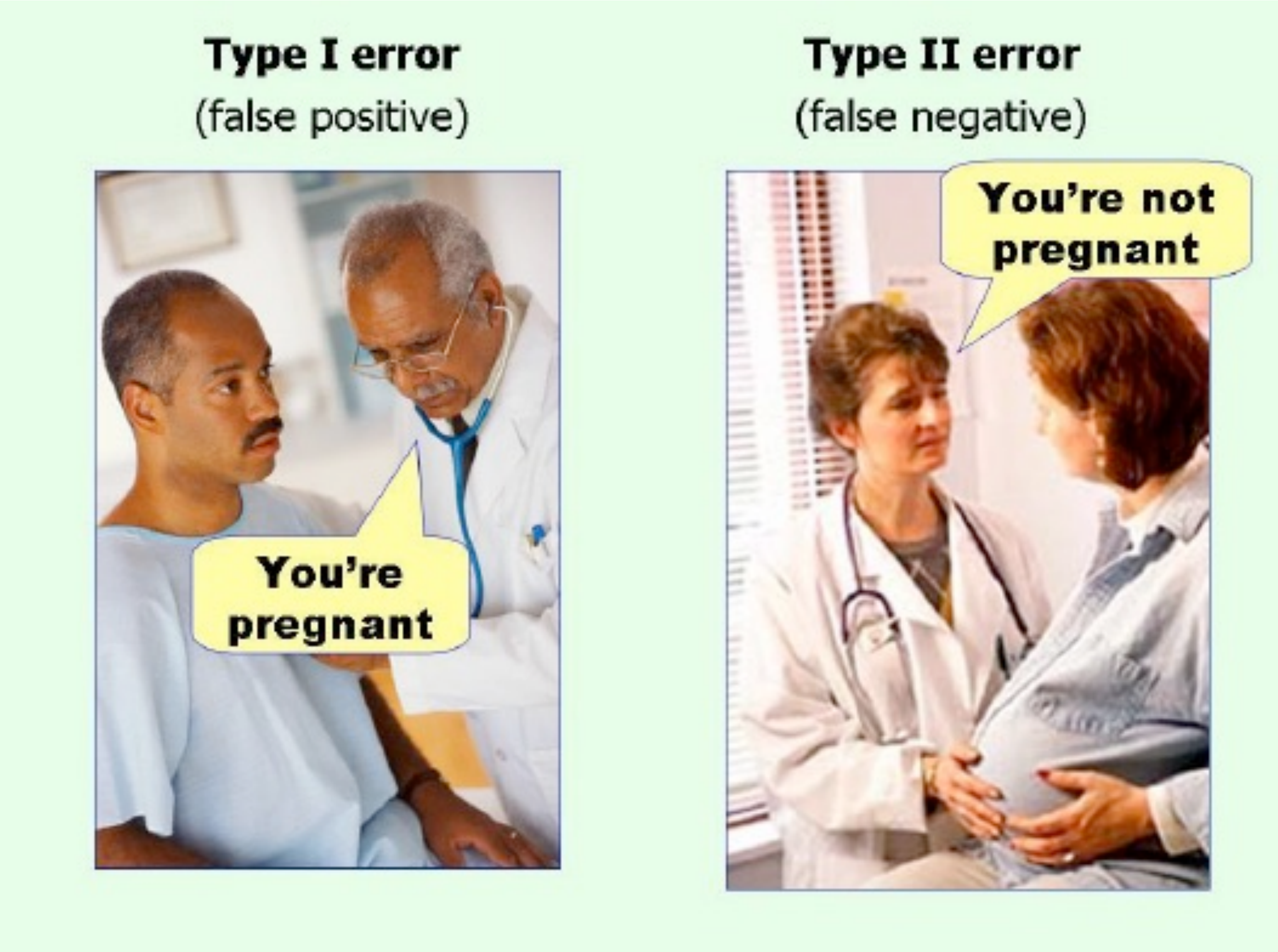
Reality

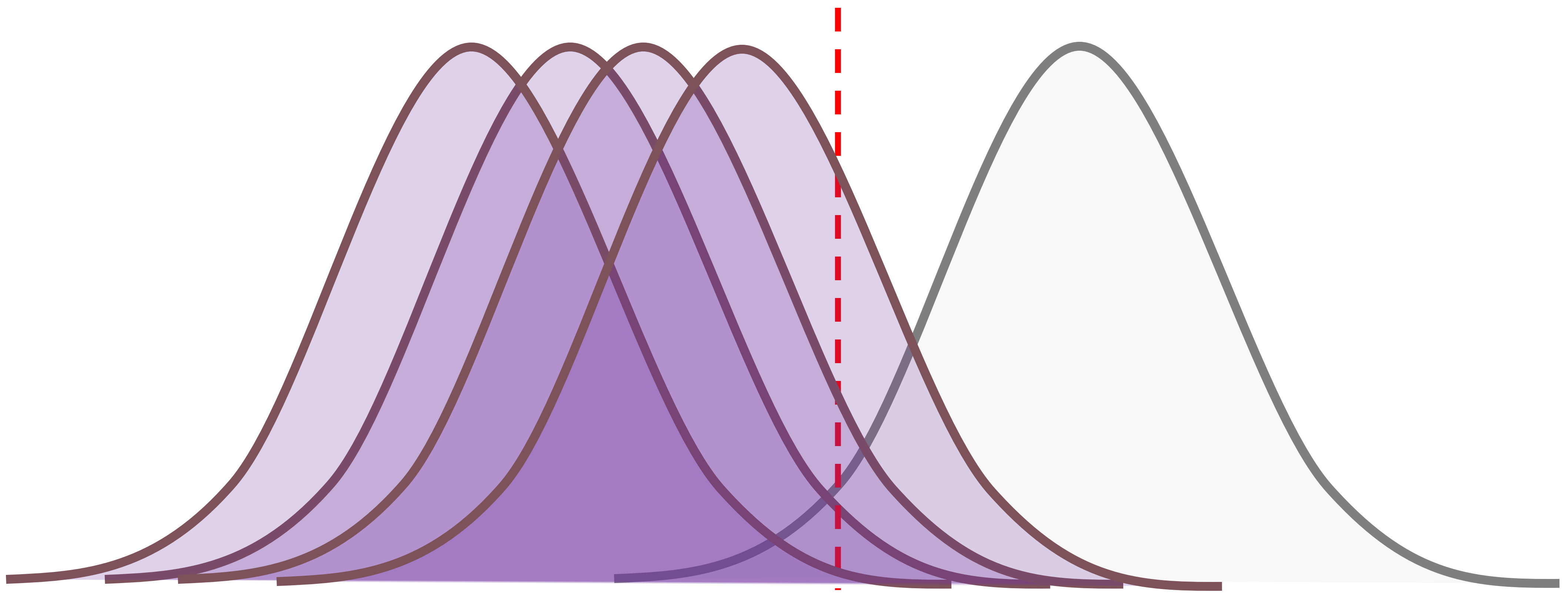
Decision		**H_0 is true**	**H_0 is false**
	Reject H_0		
	Fail to reject H_0		**Wrong** Type II Error

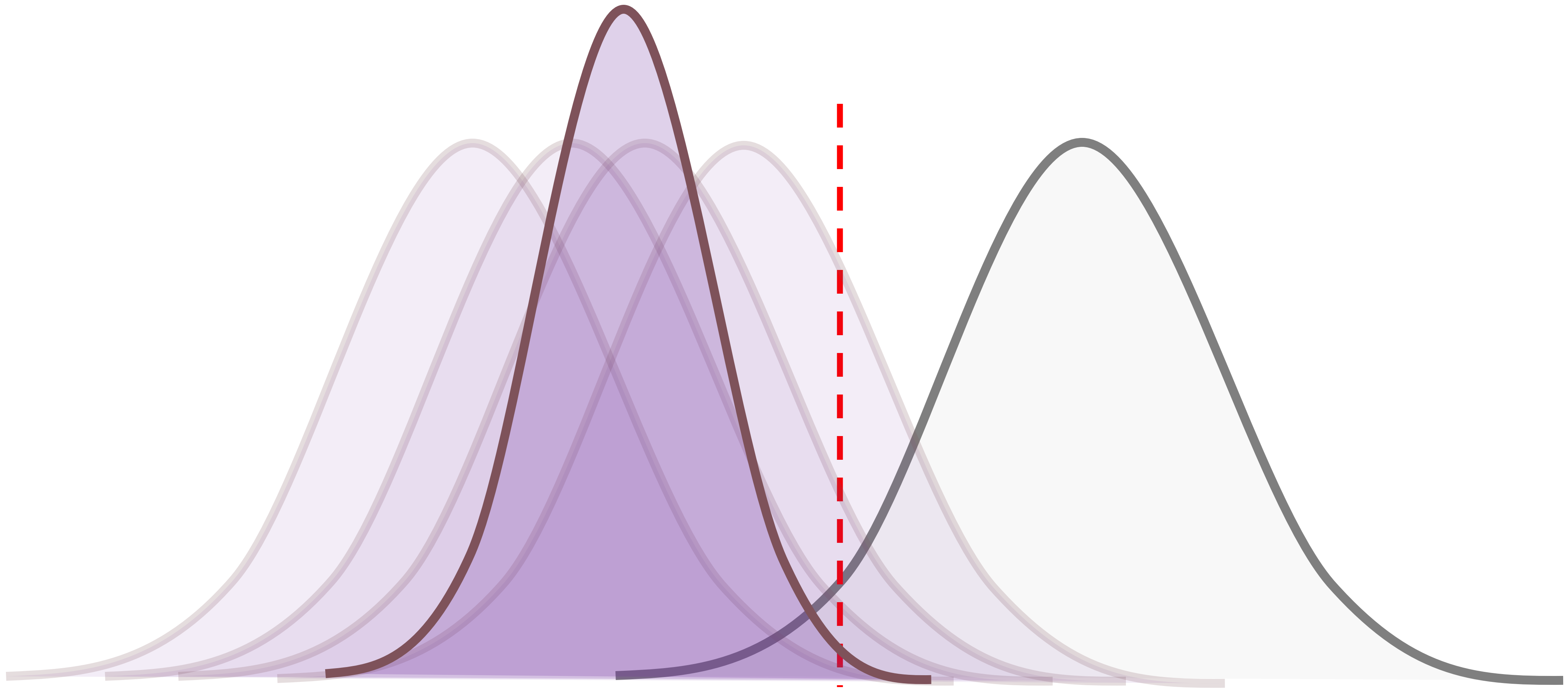




null hypothesis
theoretical
differences
α = 0.05
$\mu_1 - \mu_2$
$\mu_1 - \mu_2 = 0?$
"False positive"
"False negative"
Type I Error
Type II Error
-3
-2
-1
0
1
2
3


Type I error
(false positive)
You're pregnant
Type II error
(false negative)
You're not pregnant



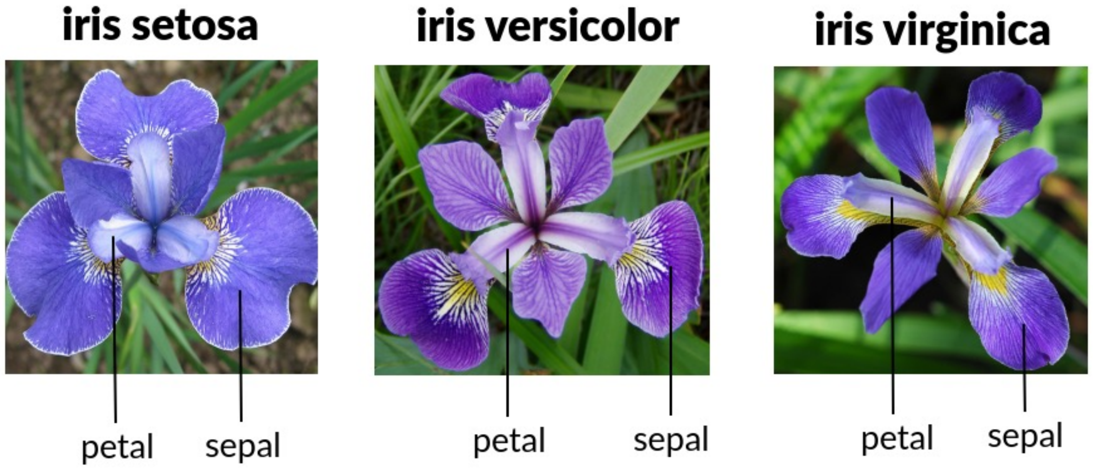


Summary I

- Mean & Standard deviation
- Null Hypothesis & Alternative Hypothesis
- α (alpha) & p-value
- Type I & Type II Error
- t-test

Introduction to R:
Part I

- R & RStudio environment
- Basic math operations:
  - addition, subtraction, multiplication, division, power, square-root
- Mean
- Standard deviation
- T-test



R:

https://cran.r-project.org/bin/windows/base/

RStudio

https://posit.co/products/open-source/rstudio/

```
packages <-
c("tidyverse","emmeans","ggeffects","lme4","lmerTest")

install.packages(packages)
```

t-test : comparison between two means -> $\boldsymbol{\mu_1 - \mu_2}$

t-test : comparison between two means -> $\boldsymbol{\mu_1 - \mu_2}$

$$\boldsymbol{\mu_1 = \mu_2 = \mu_3}?$$

t-test : comparison between two means -> $\boldsymbol{\mu_1 - \mu_2}$

Species: Setosa = Versicolor = Virginica ?

$$\boldsymbol{\mu_1 = \mu_2 = \mu_3}?$$

t-test : comparison between two means -> $\boldsymbol{\mu_1 - \mu_2}$

$$\boldsymbol{\mu_{A1} = \mu_{A2} = \mu_{A3} = \mu_{B1} = \mu_{B2} = \mu_{B3}}\ ??$$

t-test : comparison between two means -> $\boldsymbol{\mu_1 - \mu_2}$

<u>A: Species</u>

A1: Setosa

A2: Versicolor

A3: Virginica

$$\boldsymbol{\mu_{A1} = \mu_{A2} = \mu_{A3} = \mu_{B1} = \mu_{B2} = \mu_{B3}}\ ??$$

t-test : comparison between two means -> $\boldsymbol{\mu_1 - \mu_2}$

A: Species

A1: Setosa

A2: Versicolor

A3: Virginica

B: Place

B1: Your home

B2: Your neighbor's home

B3: Your grandma's home

$$\boldsymbol{\mu_{A1} = \mu_{A2} = \mu_{A3} = \mu_{B1} = \mu_{B2} = \mu_{B3}}\ ??$$

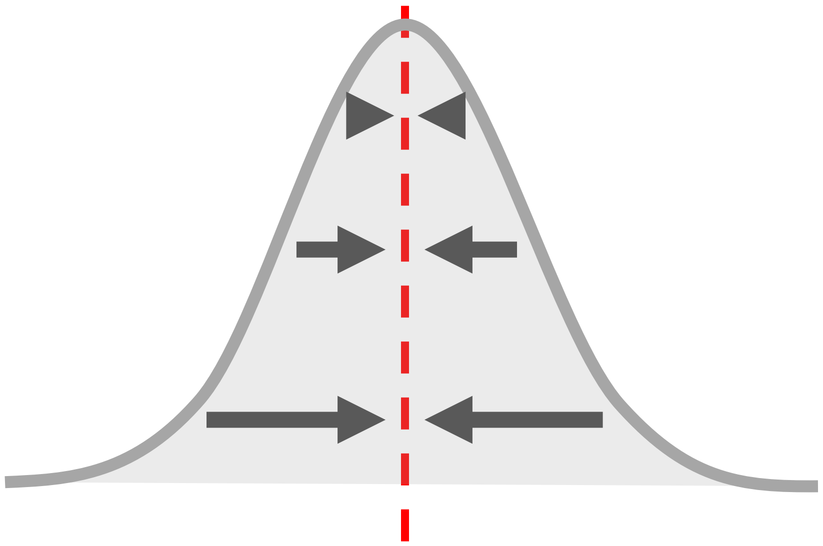
ANOVA:
Analysis of variance

ANOVA:
Analysis of variance

variance: “spread” of the data

ANOVA:
Analysis of variance

variance: “spread” of the data



ANOVA:
Analysis of variance

$$\textbf{Variance} = \frac{\textbf{Sum of squared deviations}, \boldsymbol{SS}}{\textbf{Degrees of freedom}, \boldsymbol{df}}$$

ANOVA:
Analysis of variance

$$\frac{\textbf{variance between groups}}{\textbf{variance within group}}$$

ANOVA:
Analysis of variance

$$\frac{\textbf{variance between groups}}{\textbf{variance within group}} =$$

$$\frac{MS_A}{MS_R} =$$

F-ratio

$$\text{Variance} = \frac{\text{Sum of squared deviations}, SS}{\text{Degrees of freedom}, df}$$

$$\textbf{Variance} = \frac{\textbf{Sum of squared deviations}, \boldsymbol{SS}}{\textbf{Degrees of freedom}, \boldsymbol{df}}$$

Squared deviation values are typically referred to as **sums of squares**, and abbreviated **SS**

$$\text{Variance} = \frac{\text{Sum of squared deviations, } SS}{f}$$

$$\frac{\textbf{variance between groups}}{\textbf{variance within group}}$$

Squared deviation values are typically referred to as **sums of squares**, and abbreviated **SS**

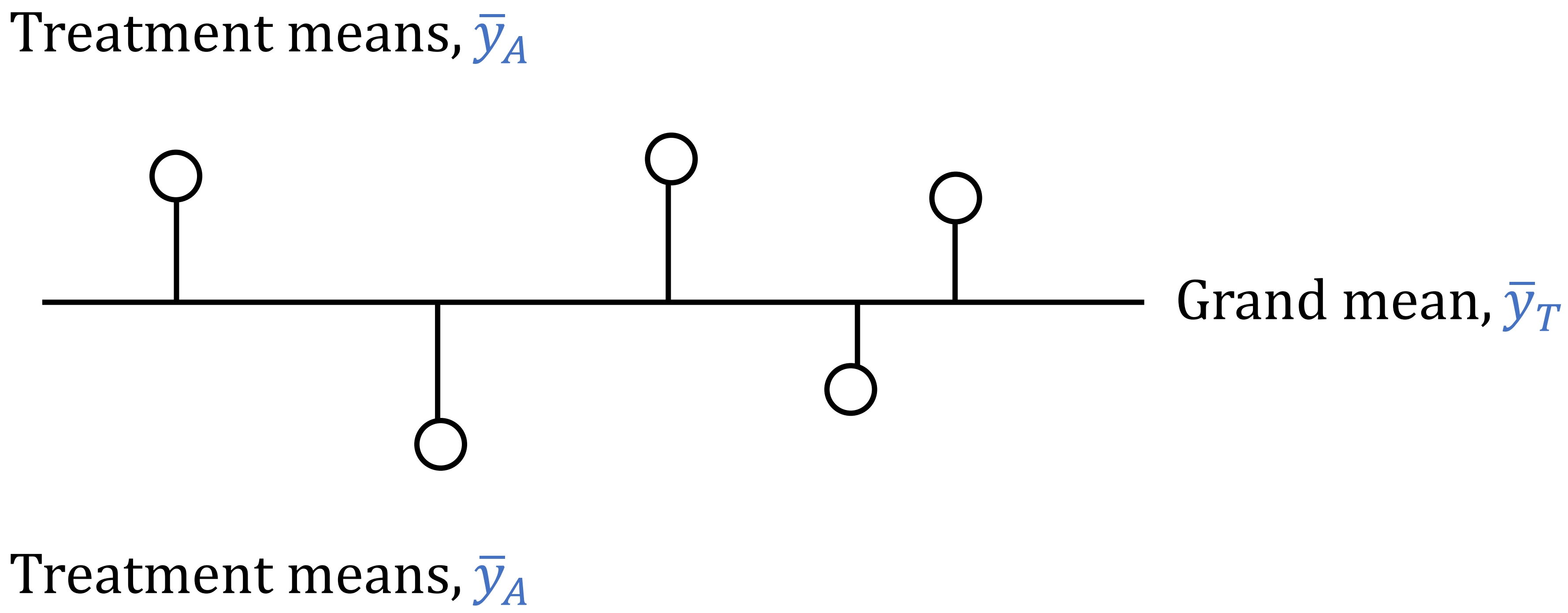
$$\frac{\textbf{variance between groups}}{\textbf{variance within group}}$$

- Between-groups: $SS_{between} = SS_A = \sum_A s \cdot (\bar{y}_A - \bar{y}_T)^2$
- Within-groups: $SS_{within} = SS_R = \sum_{A,i} \left(y_{A,i} - \bar{y}_A\right)^2$
- Total: $SS_{total} = SS_T = \sum_{A,i} \left(y_{A,i} - \bar{y}_T\right)^2$

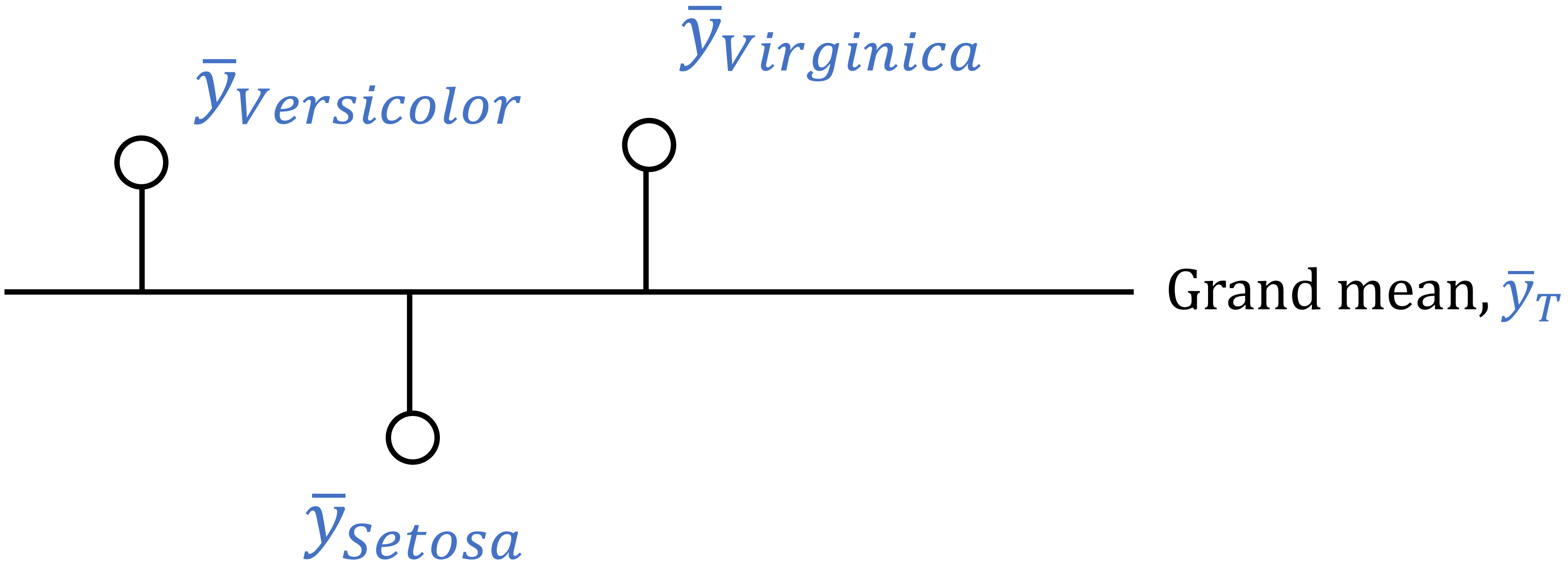
$$\frac{\textbf{variance between groups}}{\textbf{variance within group}}$$

- Between-groups: $SS_{between} = SS_A = \sum_A s \cdot (\bar{y}_A - \bar{y}_T)^2$
- Within-groups: $SS_{within} = SS_R = \sum_{A,i} \left(y_{A,i} - \bar{y}_A\right)^2$
- Total: $SS_{total} = SS_T = \sum_{A,i} \left(y_{A,i} - \bar{y}_T\right)^2$

Between-groups: $SS_{between} = SS_A = \sum_A s \cdot (\bar{y}_A - \bar{y}_T)^2$

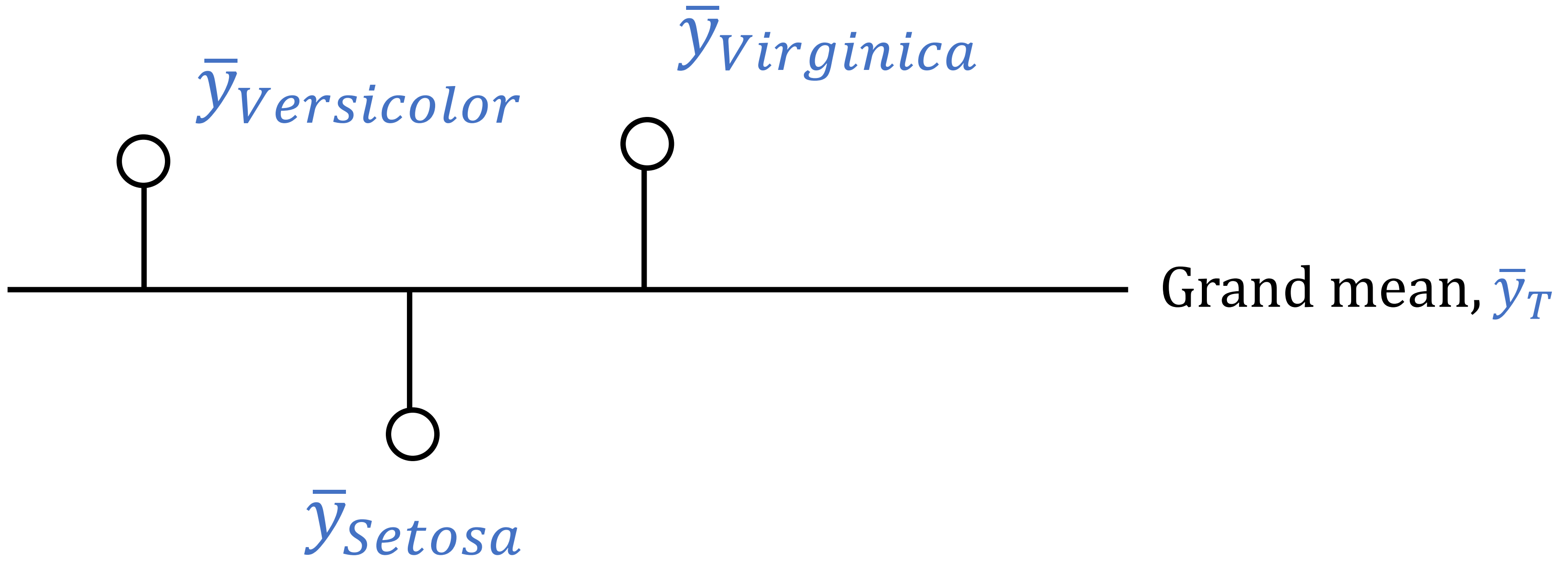


Between-groups: $SS_{between} = SS_A = \sum_A s \cdot (\bar{y}_A - \bar{y}_T)^2$



Between-groups: $SS_{between} = SS_A = \sum_A \boldsymbol{s} \cdot (\bar{y}_A - \bar{y}_T)^2$

For sample size ***s*** in each group

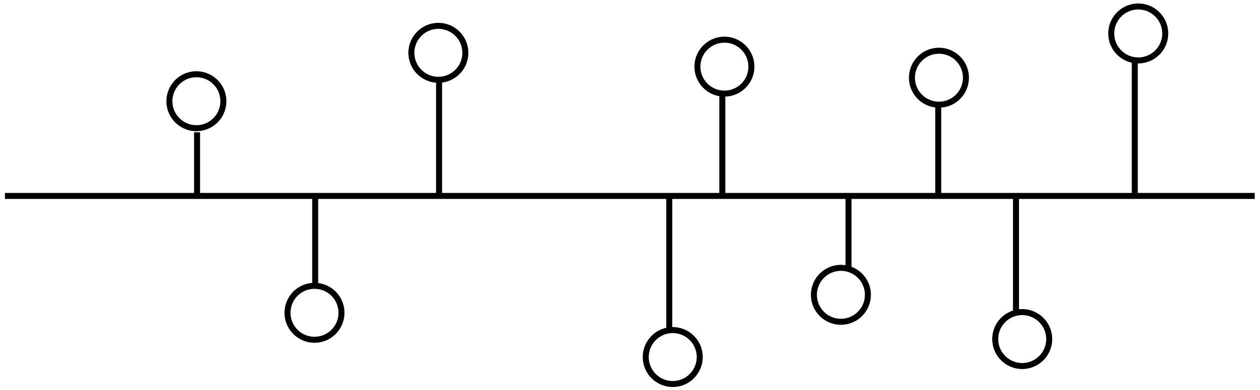


$$\frac{\textbf{variance between groups}}{\textbf{variance within group}}$$

- Between-groups: $SS_{between} = SS_A = \sum_A s \cdot (\bar{y}_A - \bar{y}_T)^2$
- Within-groups: $\boxed{SS_{within} = SS_R = \sum_{A,i} \left(y_{A,i} - \bar{y}_A\right)^2}$
- Total: $SS_{total} = SS_T = \sum_{A,i} \left(y_{A,i} - \bar{y}_T\right)^2$

Within-groups: $SS_{within} = SS_R = \sum_{A,i} (y_{A,i} - \bar{y}_A)^2$

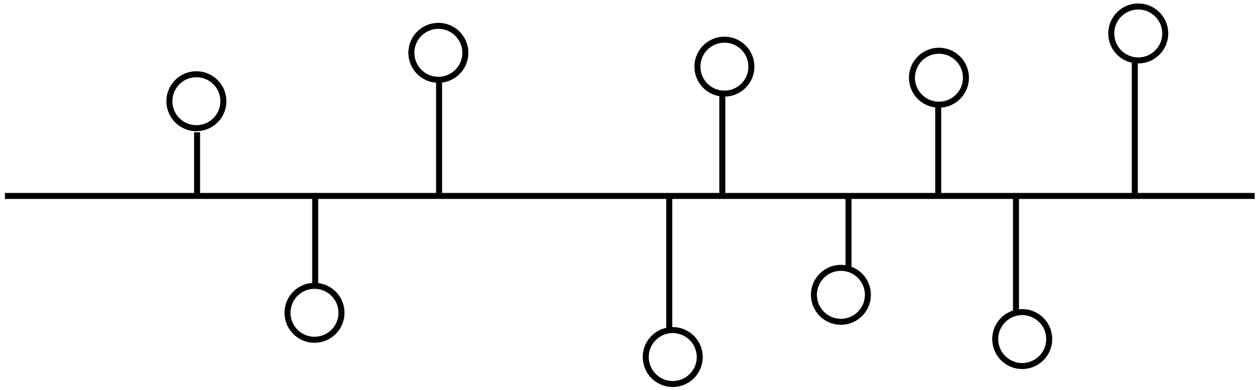
Individual data points, $y_{A,i}$



Treatment mean, $\bar{y}_A$

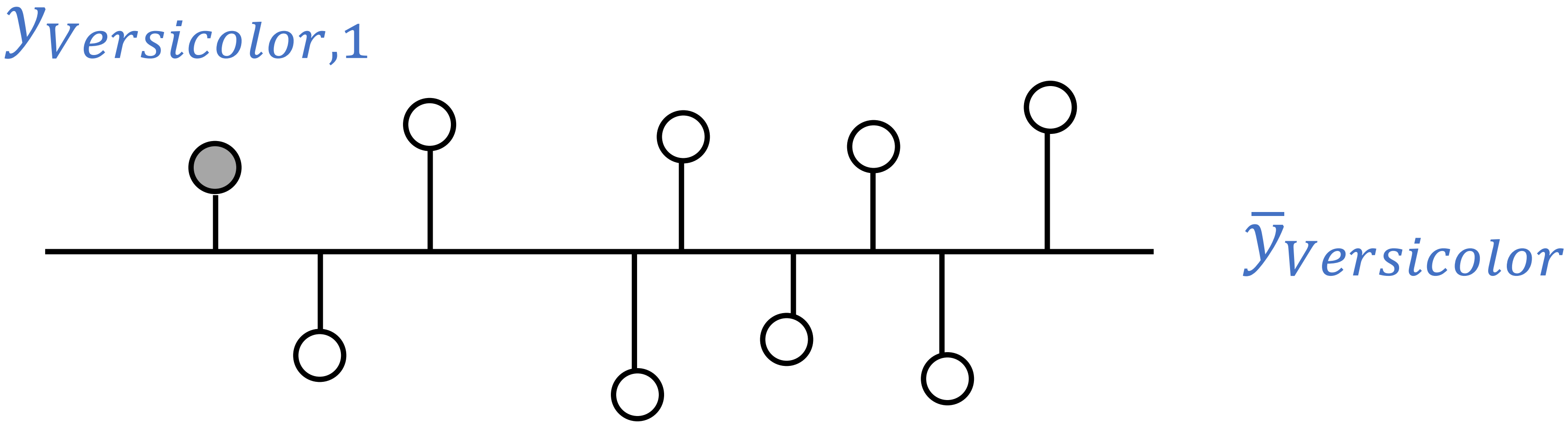
Individual data points, $y_{A,i}$

Within-groups: $SS_{within} = SS_R = \sum_{A,i}\left(y_{A,i} - \bar{y}_A\right)^2$

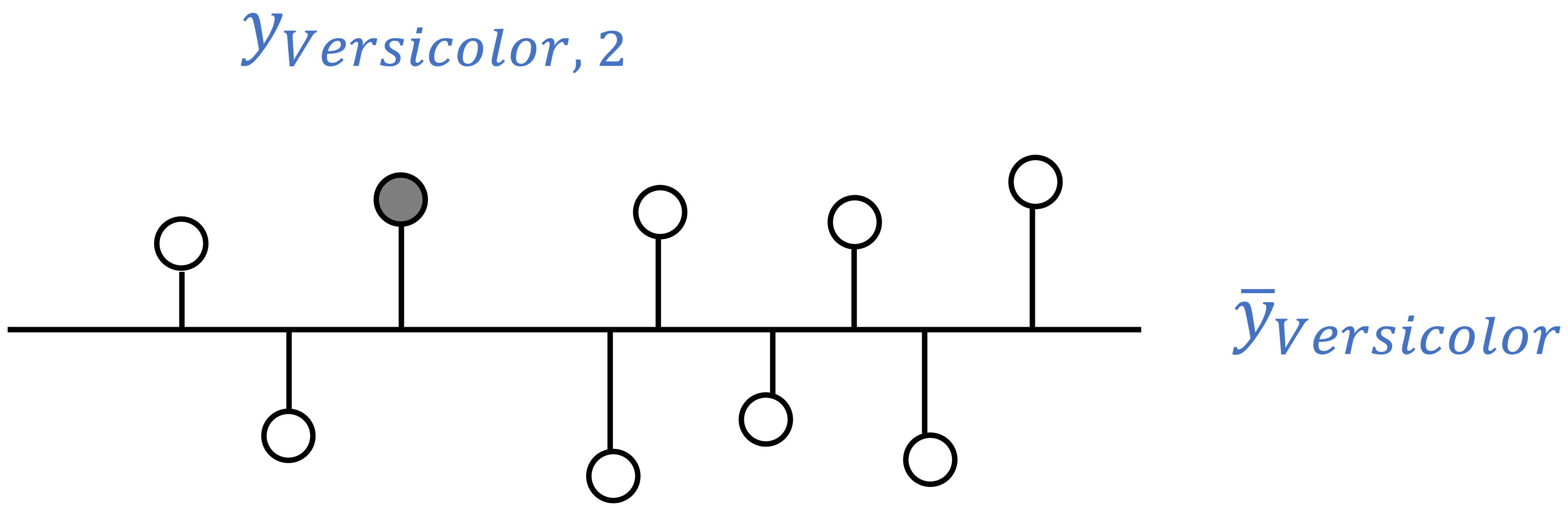


$\bar{y}_{Versicolor}$

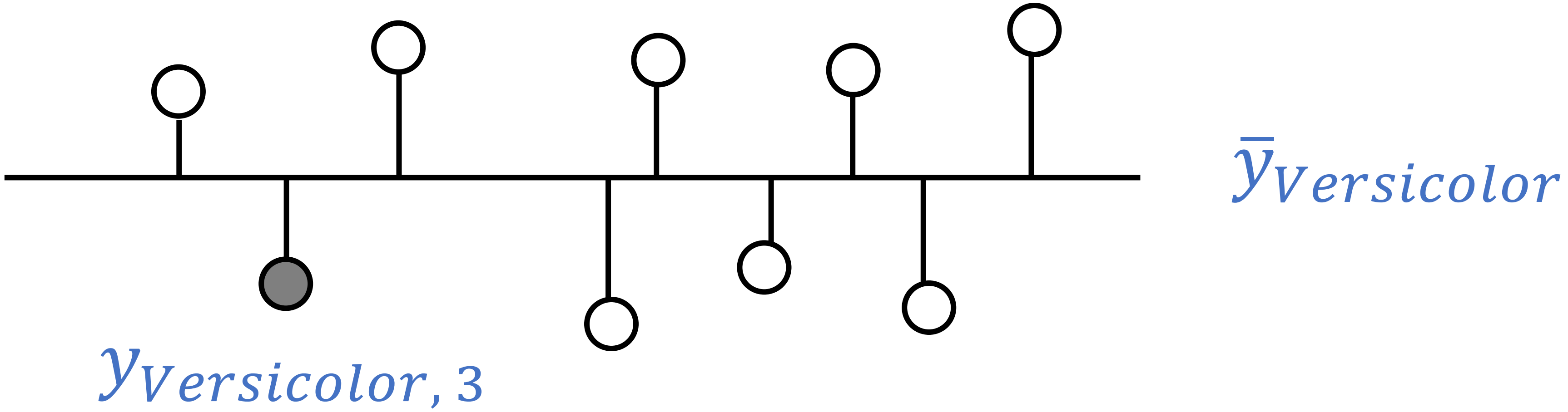
Within-groups: $SS_{within} = SS_R = \sum_{A,i}\left(y_{A,i} - \bar{y}_A\right)^2$



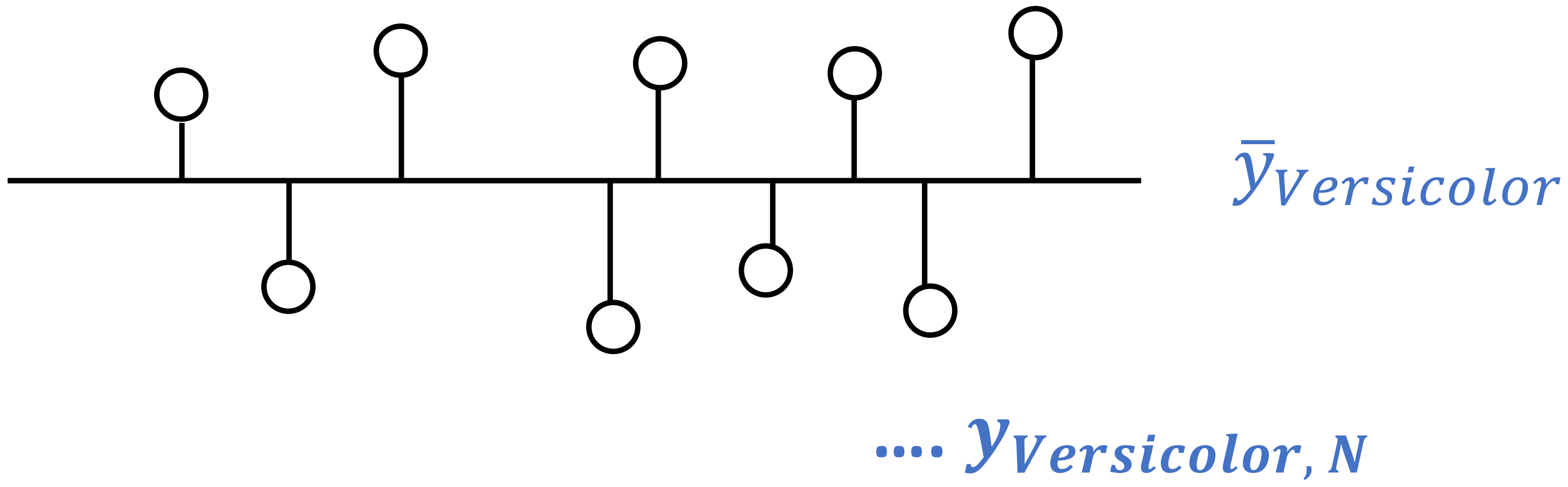
Within-groups: $SS_{within} = SS_R = \sum_{A,i}\left(y_{A,i} - \bar{y}_A\right)^2$



Within-groups: $SS_{within} = SS_R = \sum_{A,i}(y_{A,i} - \bar{y}_A)^2$



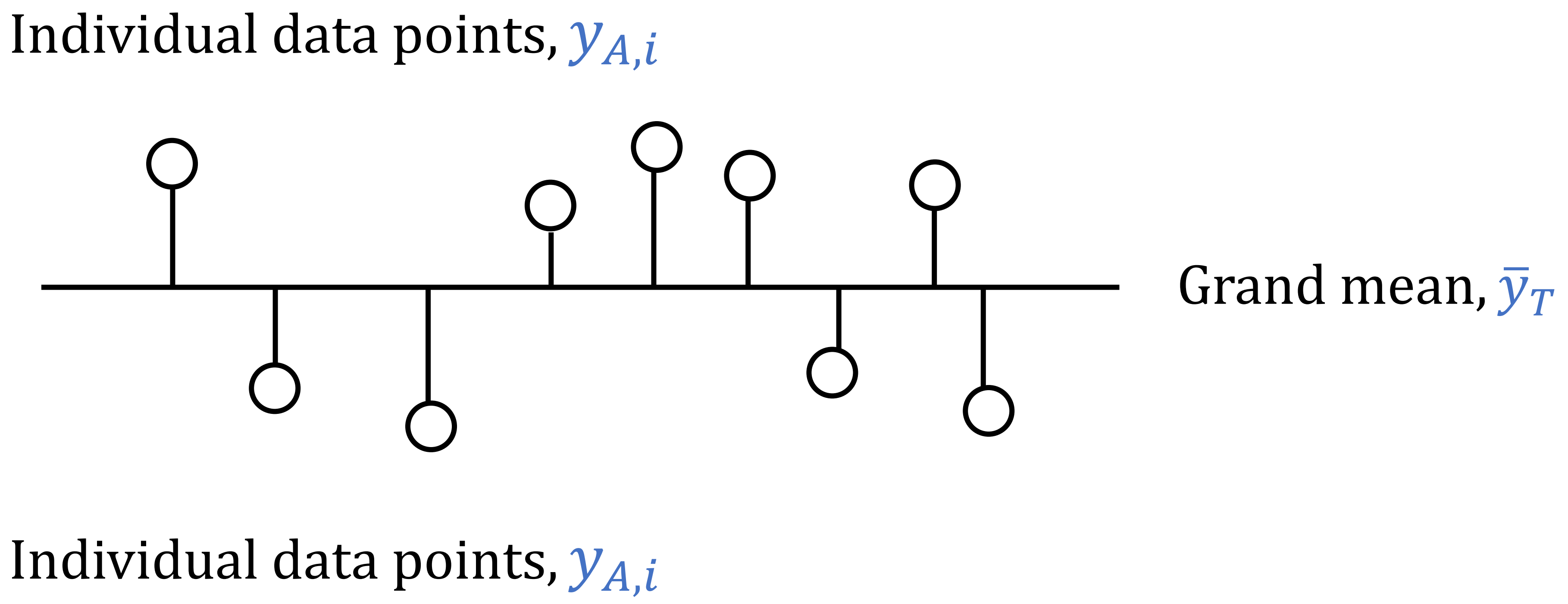
Within-groups: $SS_{within} = SS_R = \sum_{A,i} (y_{A,i} - \bar{y}_A)^2$



$$\frac{\textbf{variance between groups}}{\textbf{variance within group}}$$

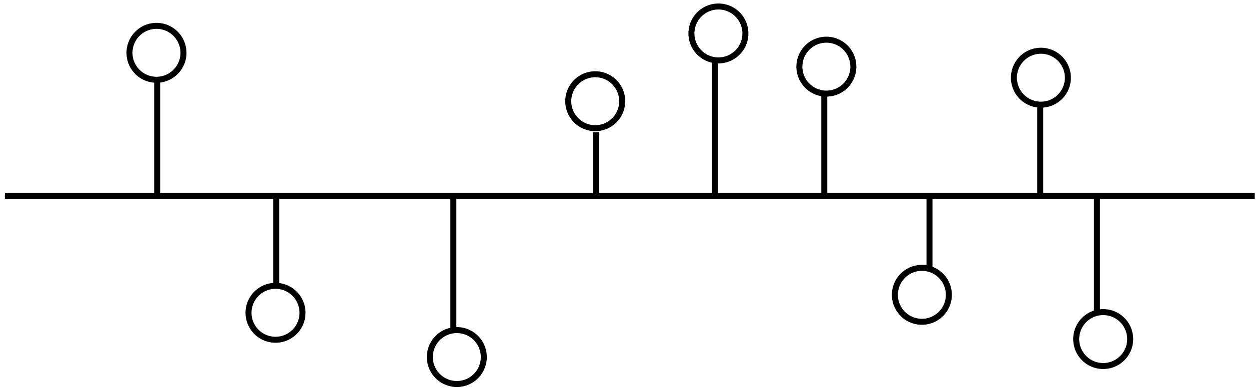
- Between-groups: $SS_{between} = SS_A = \sum_A s \cdot (\bar{y}_A - \bar{y}_T)^2$
- Within-groups: $SS_{within} = SS_R = \sum_{A,i} \left(y_{A,i} - \bar{y}_A\right)^2$
- Total: $SS_{total} = SS_T = \sum_{A,i} \left(y_{A,i} - \bar{y}_T\right)^2$

Total: $SS_{total} = SS_T = \sum_{A,i}(y_{A,i} - \bar{y}_T)^2$



Total: $SS_{total} = SS_T = \sum_{A,i} \left(y_{A,i} - \bar{y}_T \right)^2$

Individual data points, $y_{A,i}$



Grand mean, $\bar{y}_T$

Individual data points, $y_{A,i}$

$SS_{between} + SS_{within}$

Total: $SS_{total} = SS_T = \sum_{A,i} \left(y_{A,i} - \bar{y}_T\right)^2$

Useful to obtain ratio of variance explained by the treatment against noise (within-group variance)

$SS_{between} + SS_{within}$

Total: $$SS_{total} \quad = SS_T = \sum_{A,i}(y_{A,i} - \bar{y}_T)^2$$

Useful to obtain ratio of variance explained by the treatment against noise (within-group variance)

simplified e.g., $$\frac{SS_{between}}{SS_{between} + SS_{within}}$$

Mean Squared Error → F-ratio

Between-group Variance:

$$\boldsymbol{MS_A} = \frac{SS_A}{df_A}$$

Within-group Variance:

$$\boldsymbol{MS_R} = \frac{SS_R}{df_R}$$

→ $\dfrac{\boldsymbol{MS_A}}{\boldsymbol{MS_R}}$ **= F-ratio**

Summary II

- ANOVA
- Variance
- F-ratio
- Sum of squares deviation
- Mean square error

Introduction to R: Part II

- ANOVA
- emmeans()
- plot()

DAY 2

Winter Workshop:

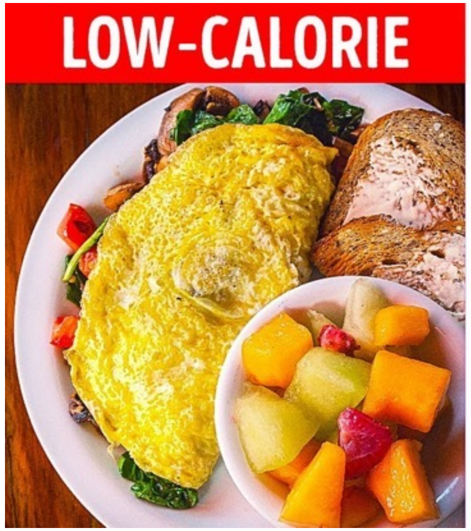
Basic and Intermediate Statistics with R

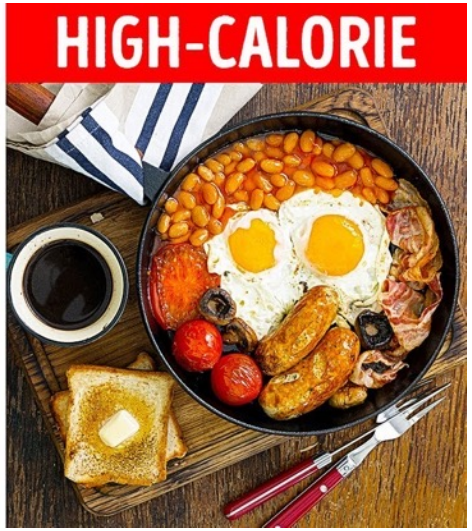
Chai Jun Ho, PhD

Schedule

~~Day 1~~	Day 2	Day 3
~~Introduction to Statistics~~	Linear Model I	Linear Model II
~~Introduction to R~~	Data Visualization	Mixed-Effect Model

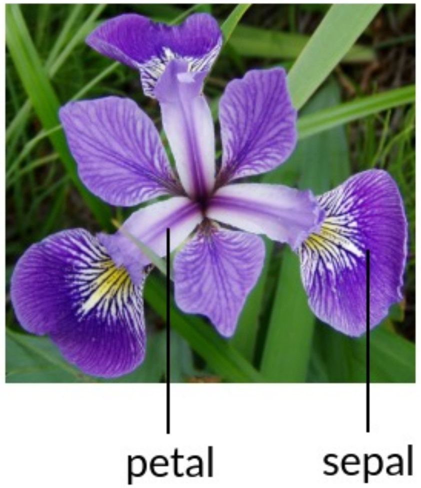
high calorie **vs** **low calorie**





<u>Categorical differences</u>

high calorie **vs** **low calorie**



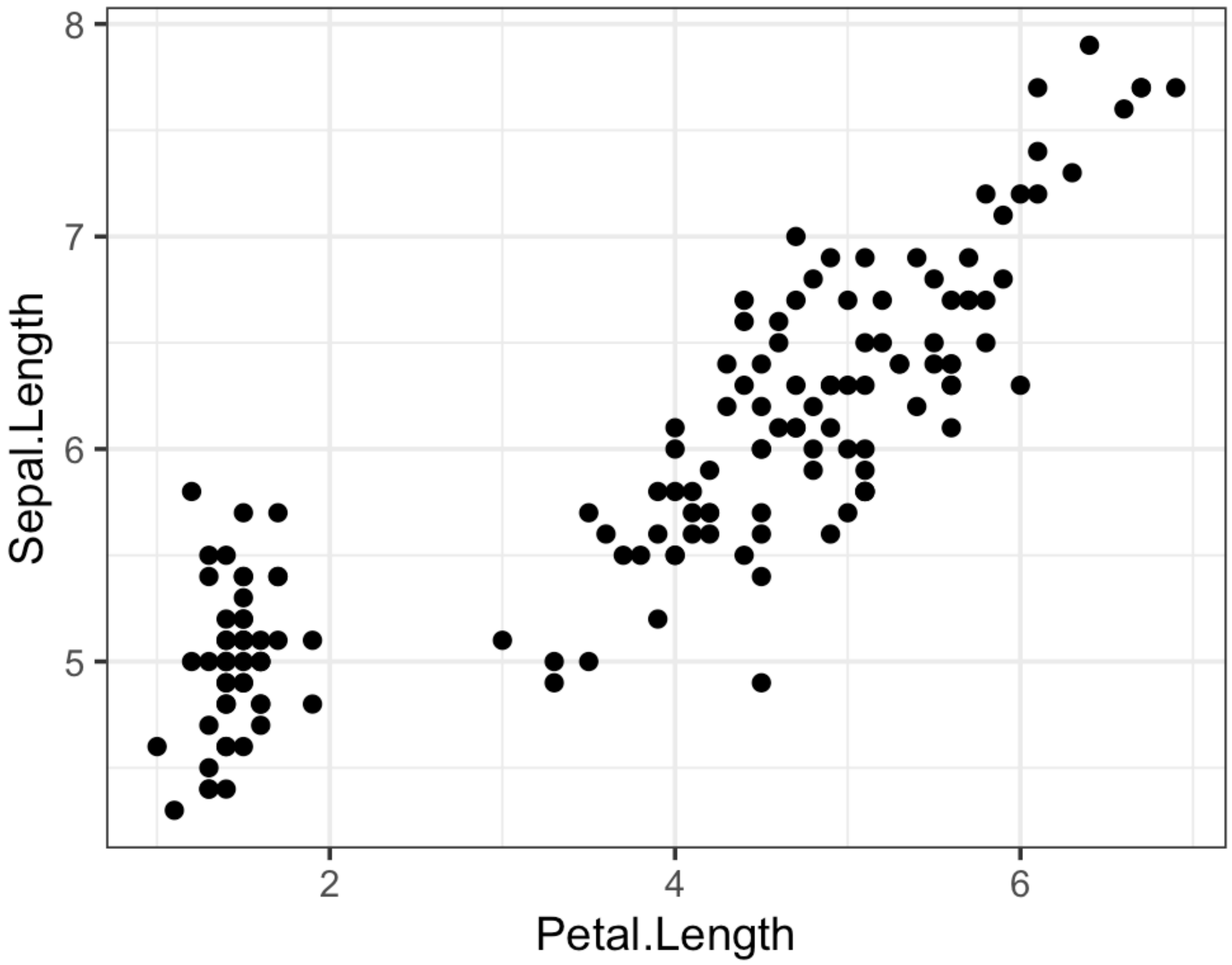
Relationships

Long petal, long sepal?

Long petal, short sepal?

Short petal, long sepal?

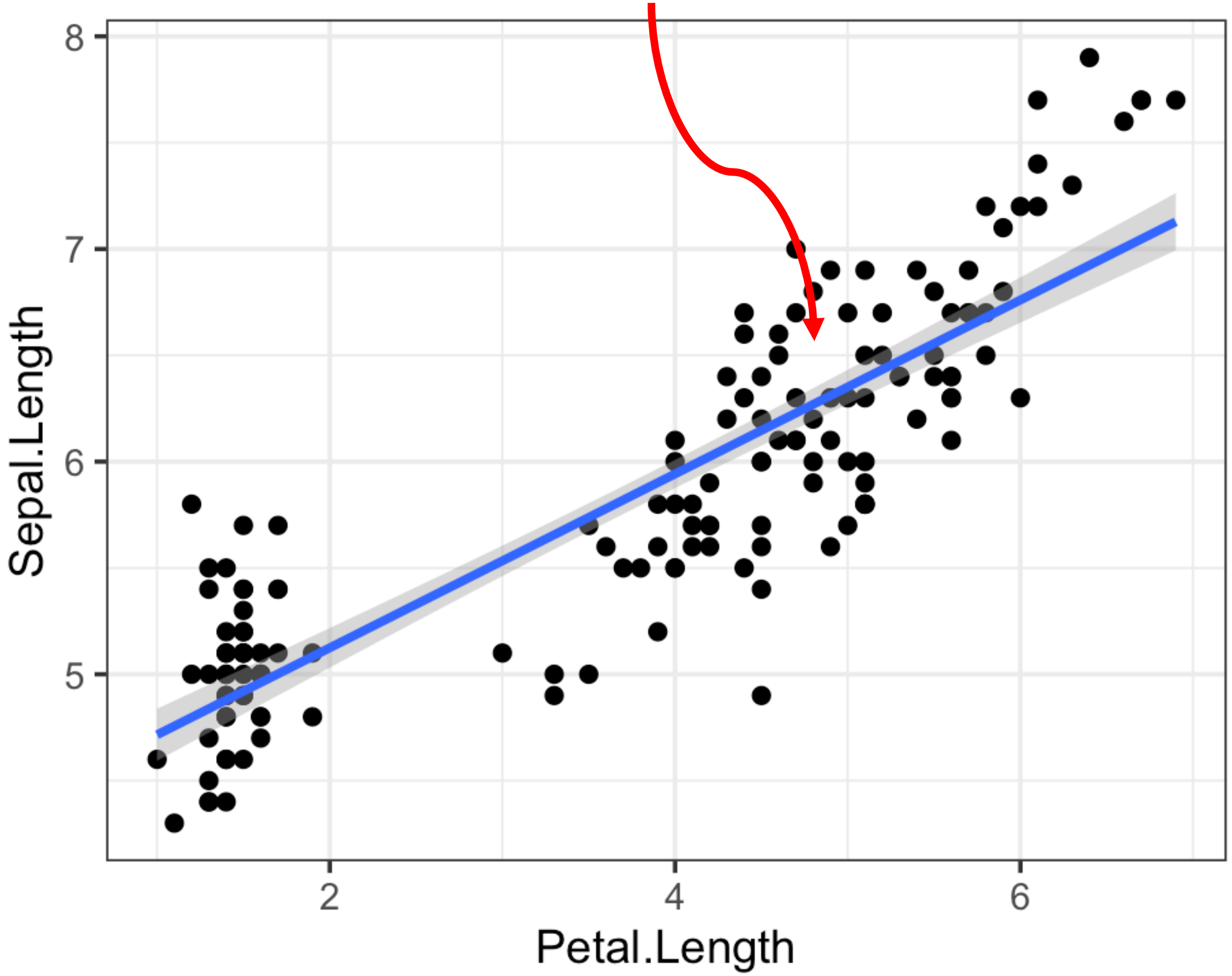
Short petal, short sepal?



8
7
6
5
Sepal.Length
2
4
6
Petal.Length

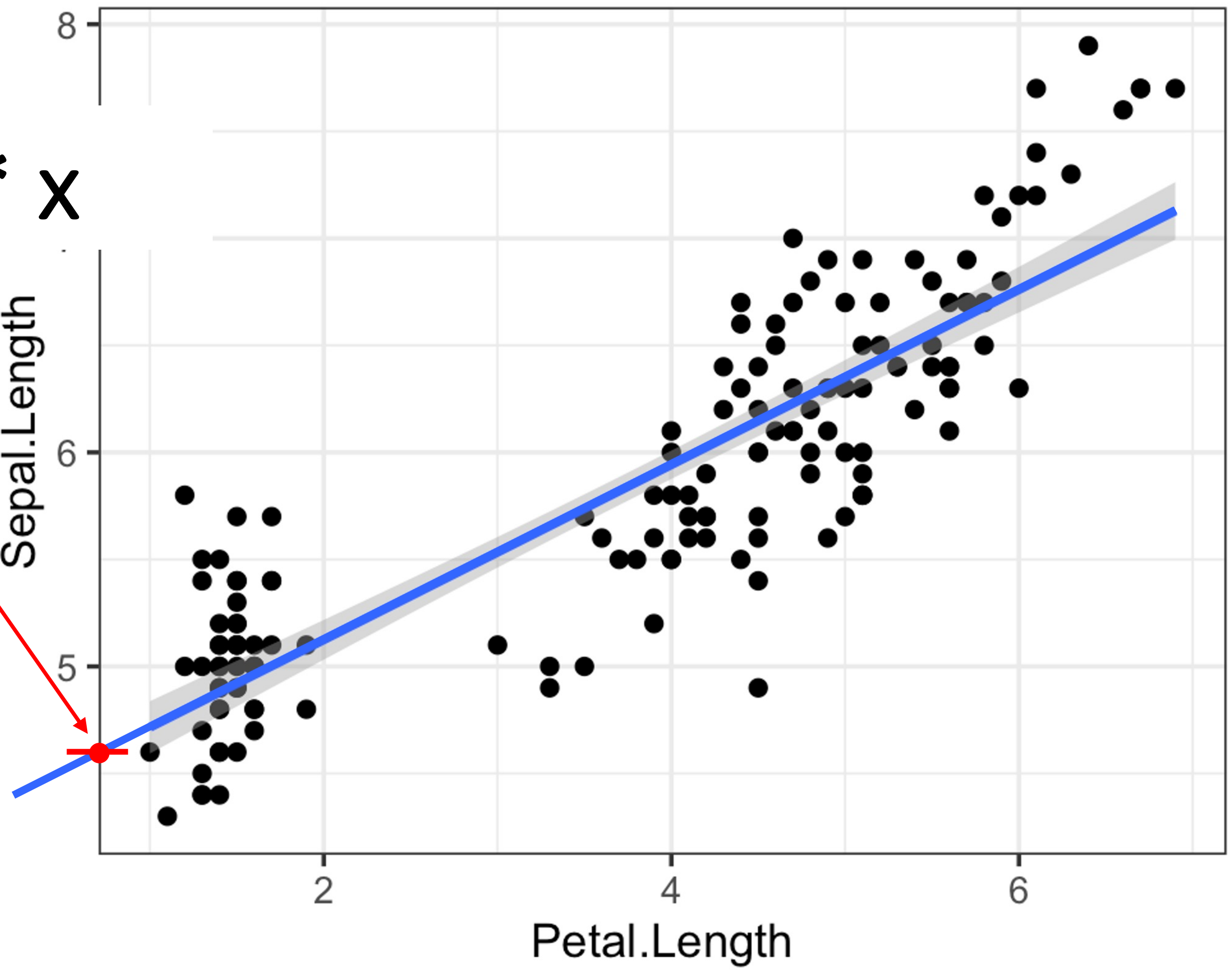
Linear model represented

by the regression line



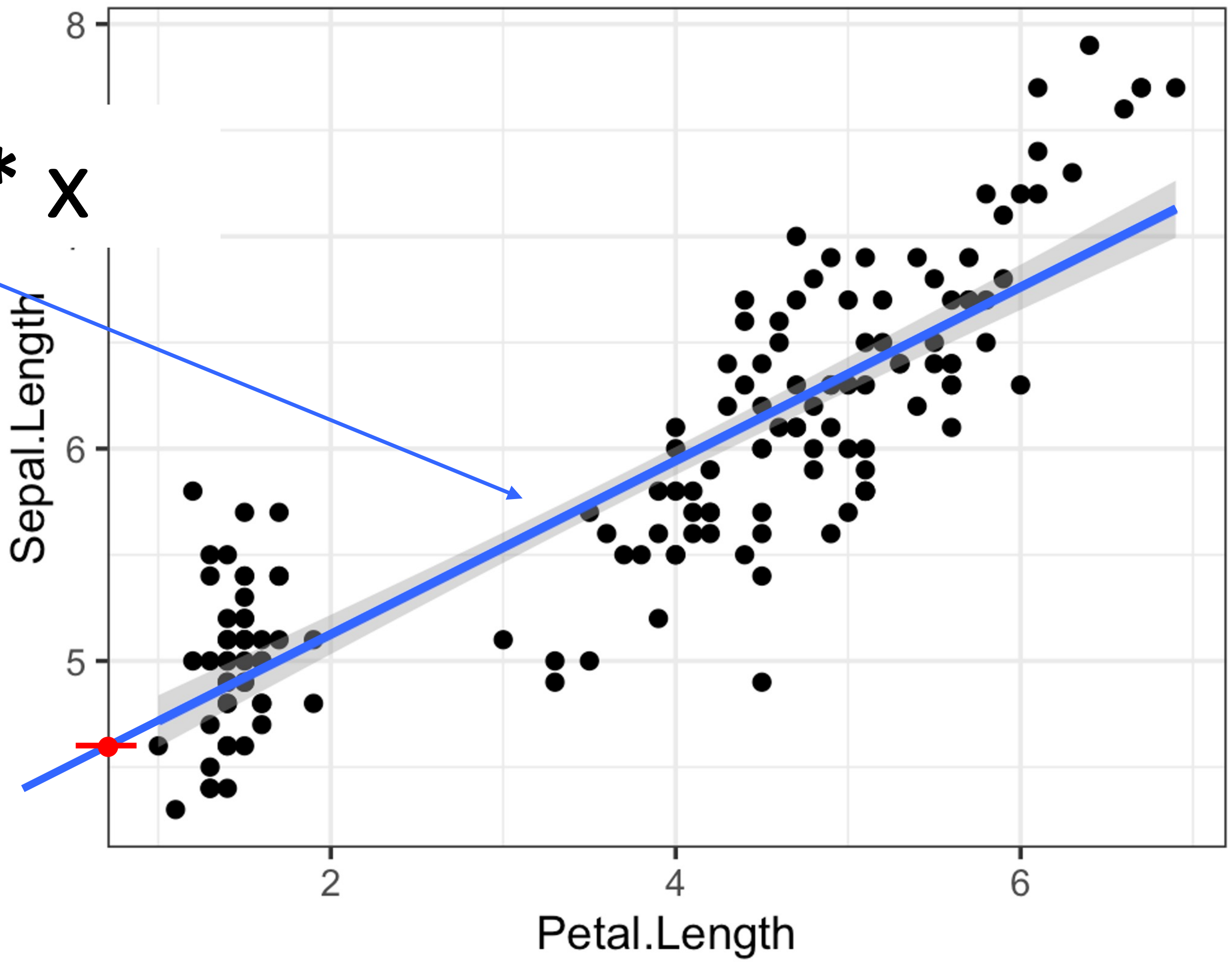
regression line

$$y = c + m * x$$



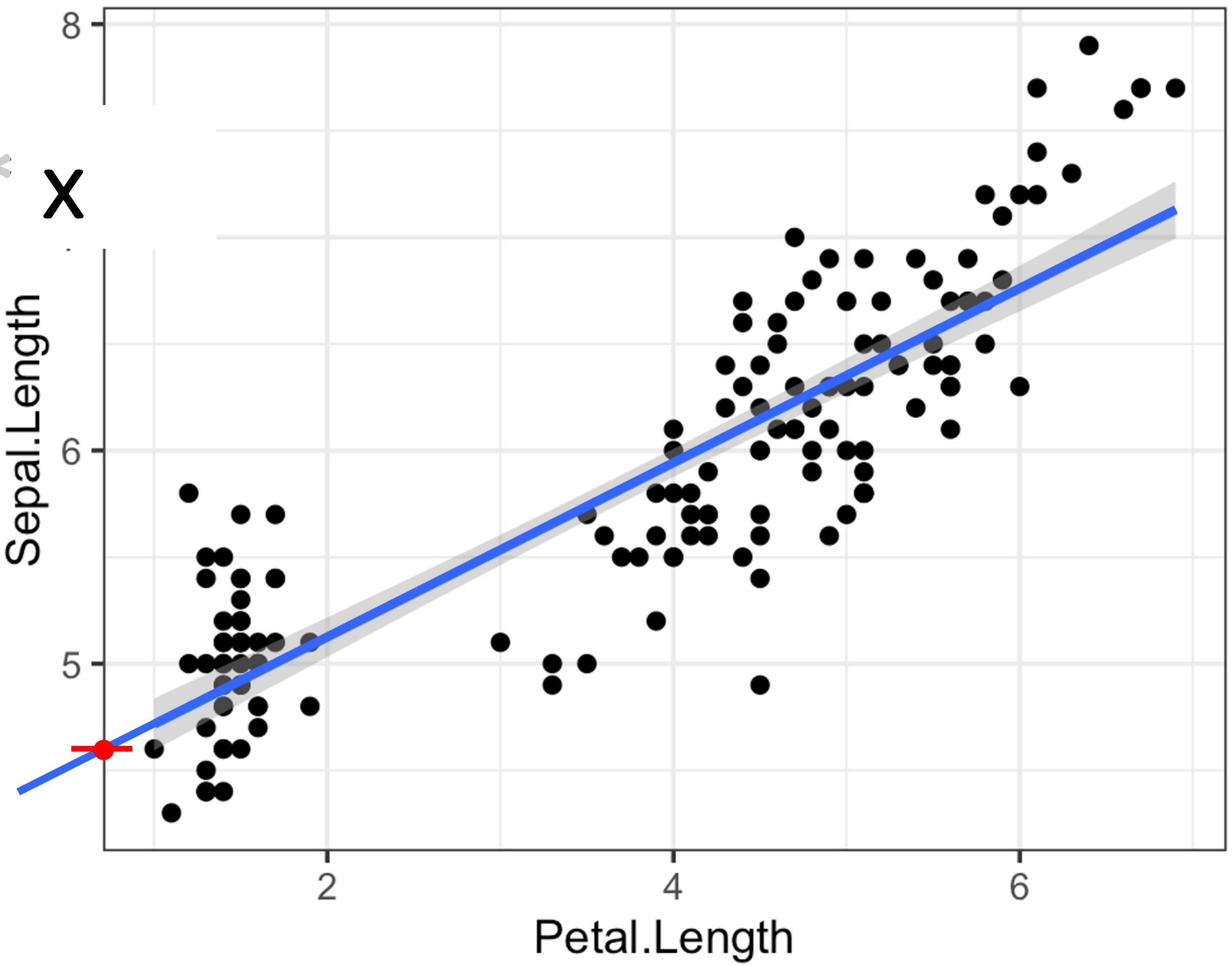
regression line

$$y = c + m * x$$

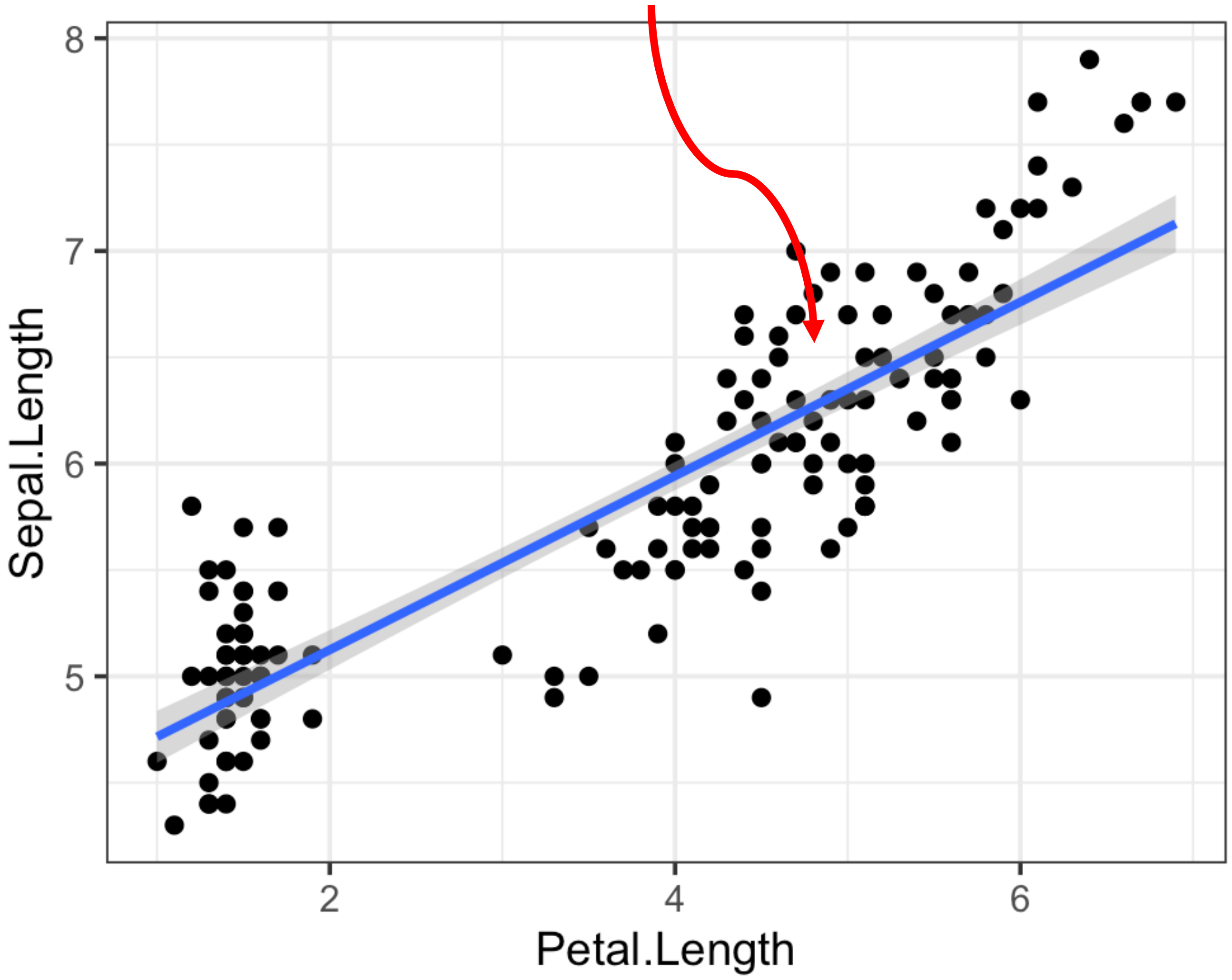


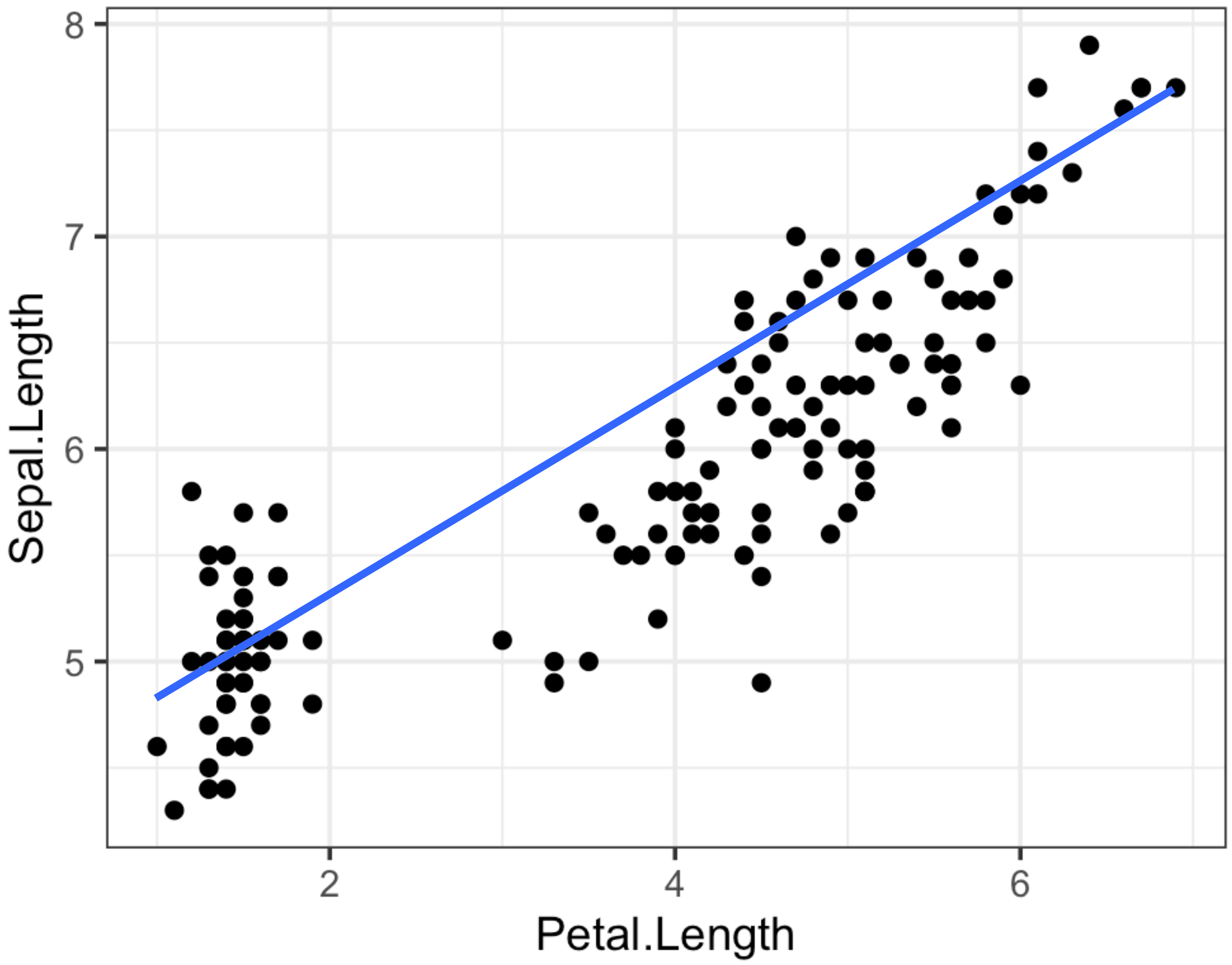
regression line

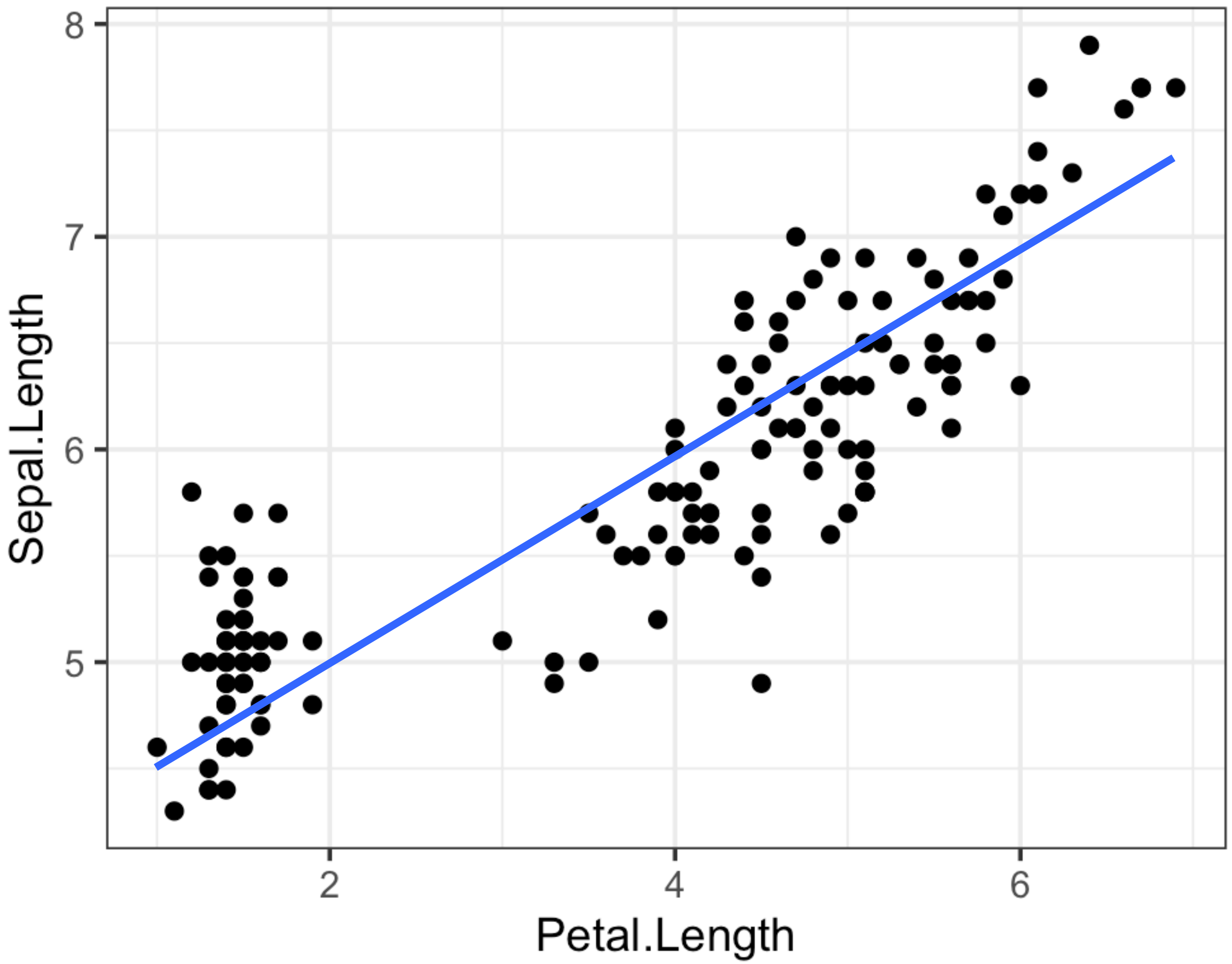
$$y = c + m * x$$

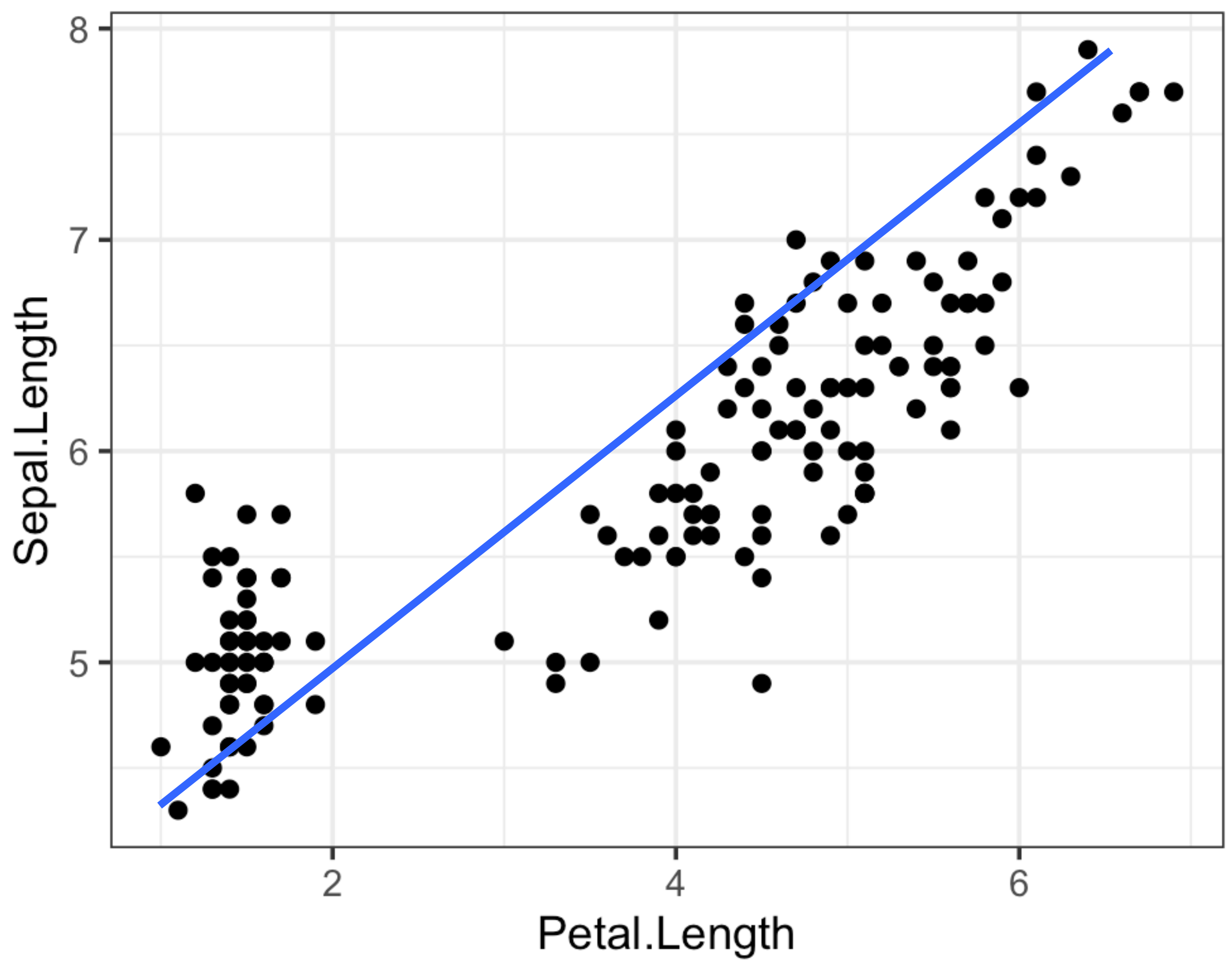


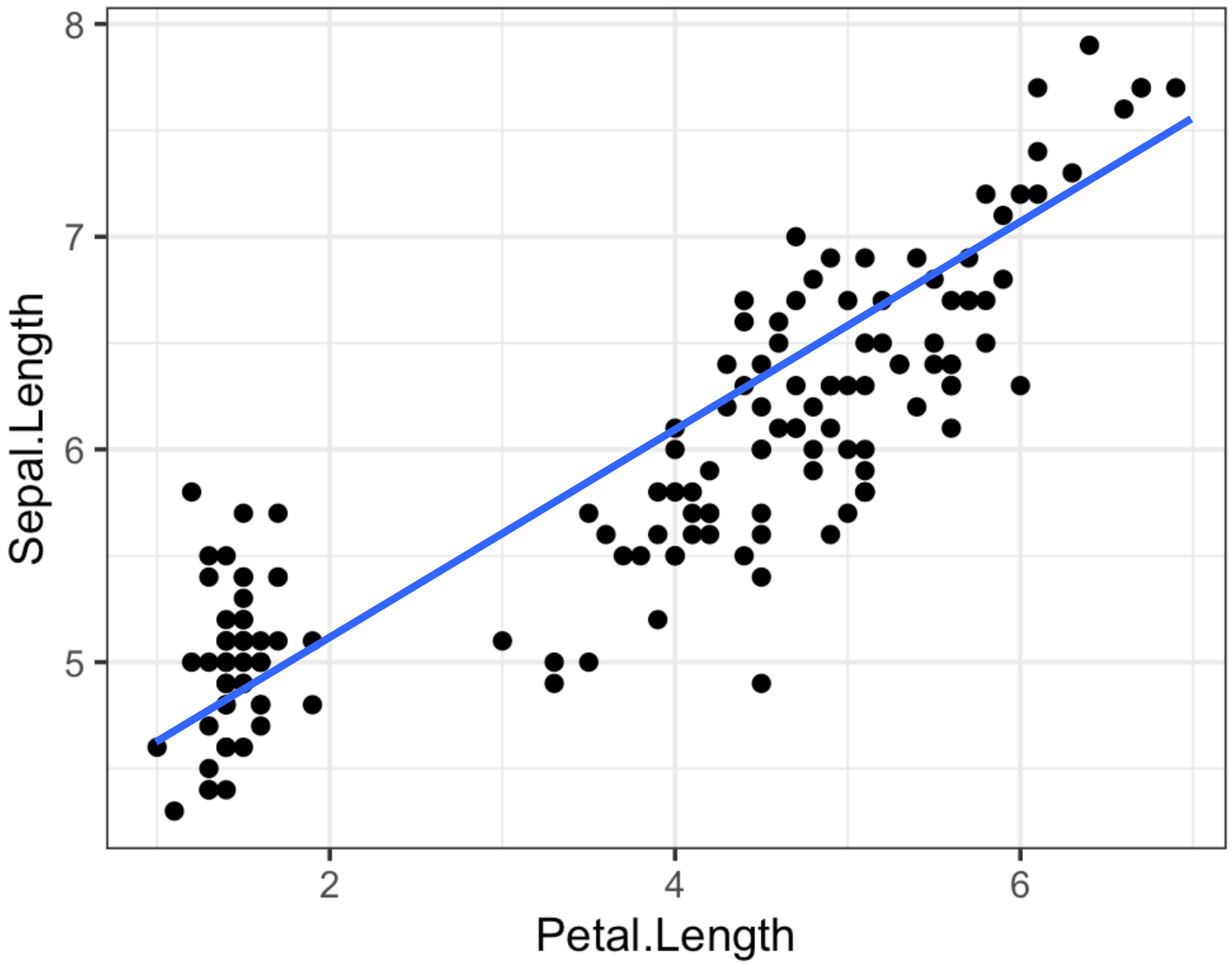
why this?



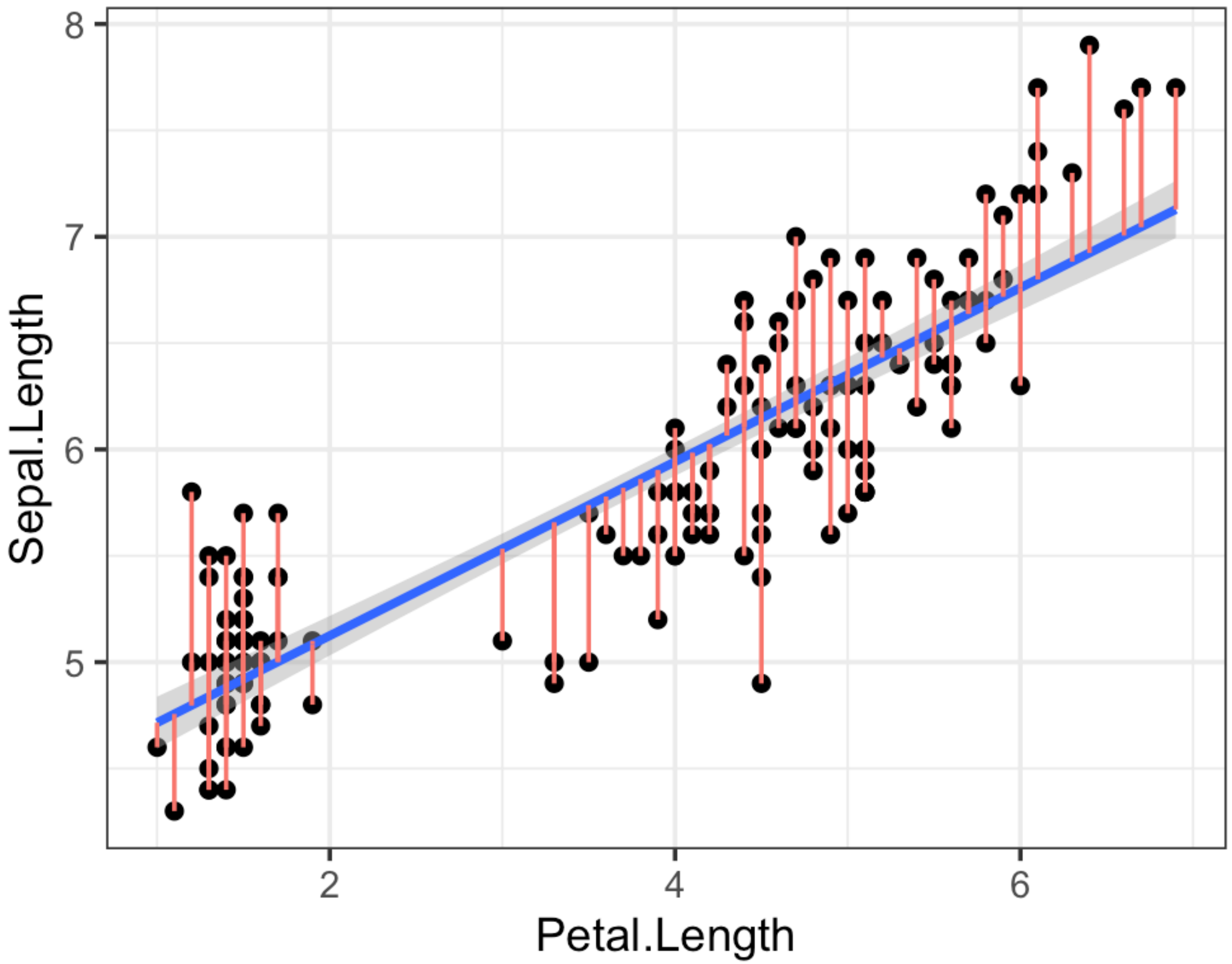

8
7
6
5
Sepal.Length
2
4
6
Petal.Length


8
7
6
5
2
4
6
Sepal.Length
Petal.Length


8
7
6
5
2
4
6
Sepal.Length
Petal.Length

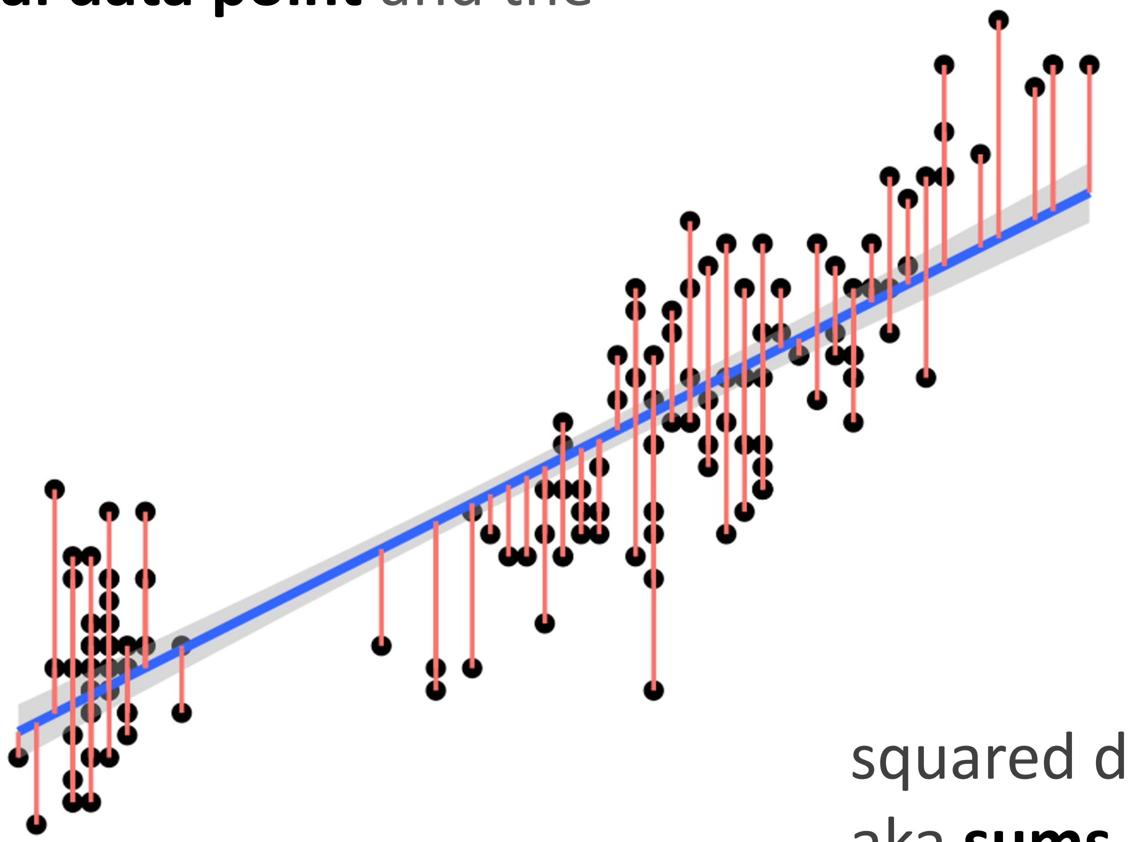


8
7
6
5
Sepal.Length
2
4
6
Petal.Length

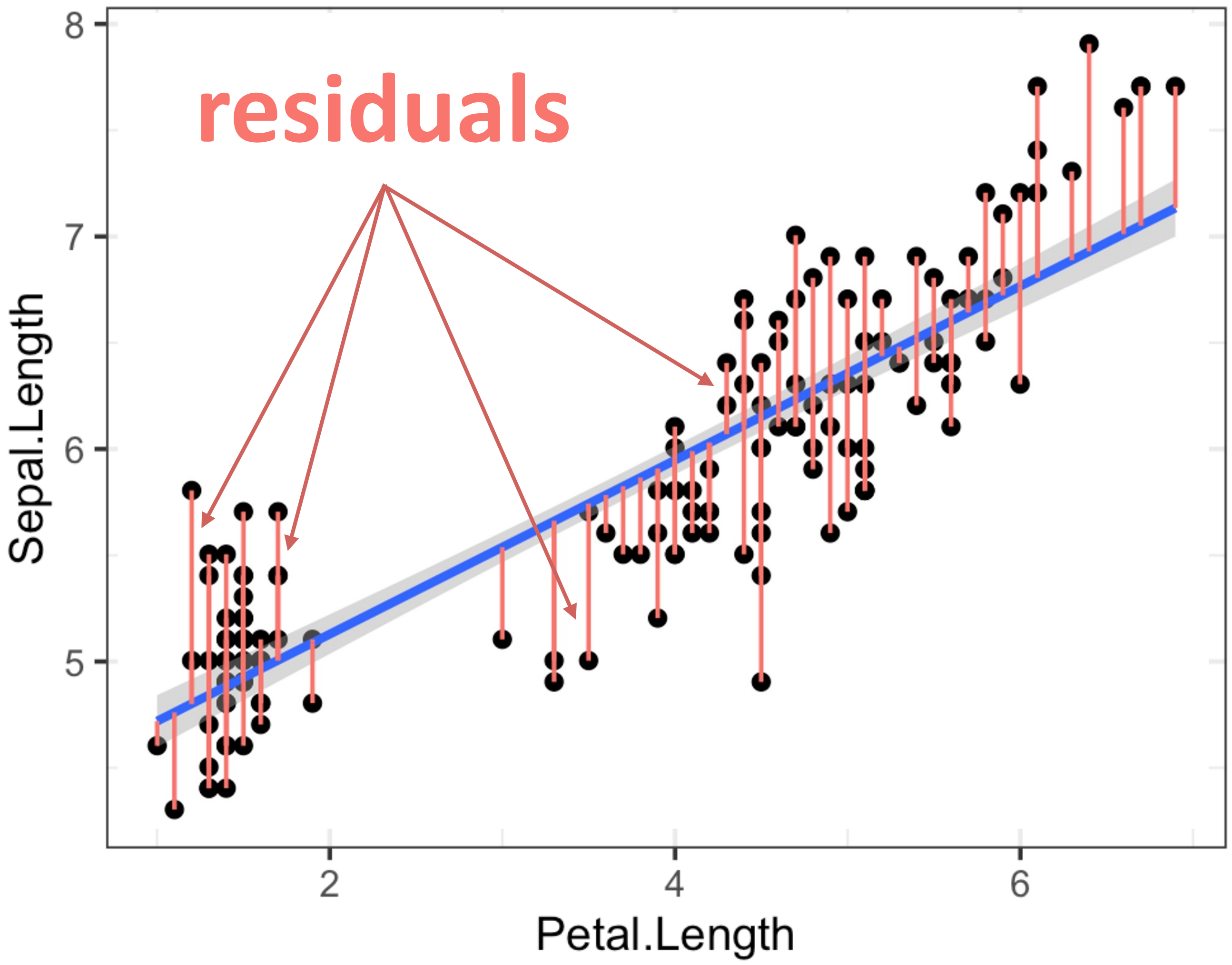


8
7
6
5
2
4
6
Sepal.Length
Petal.Length

Minimizing squared deviation/distance between **individual data point** and the **regression line**

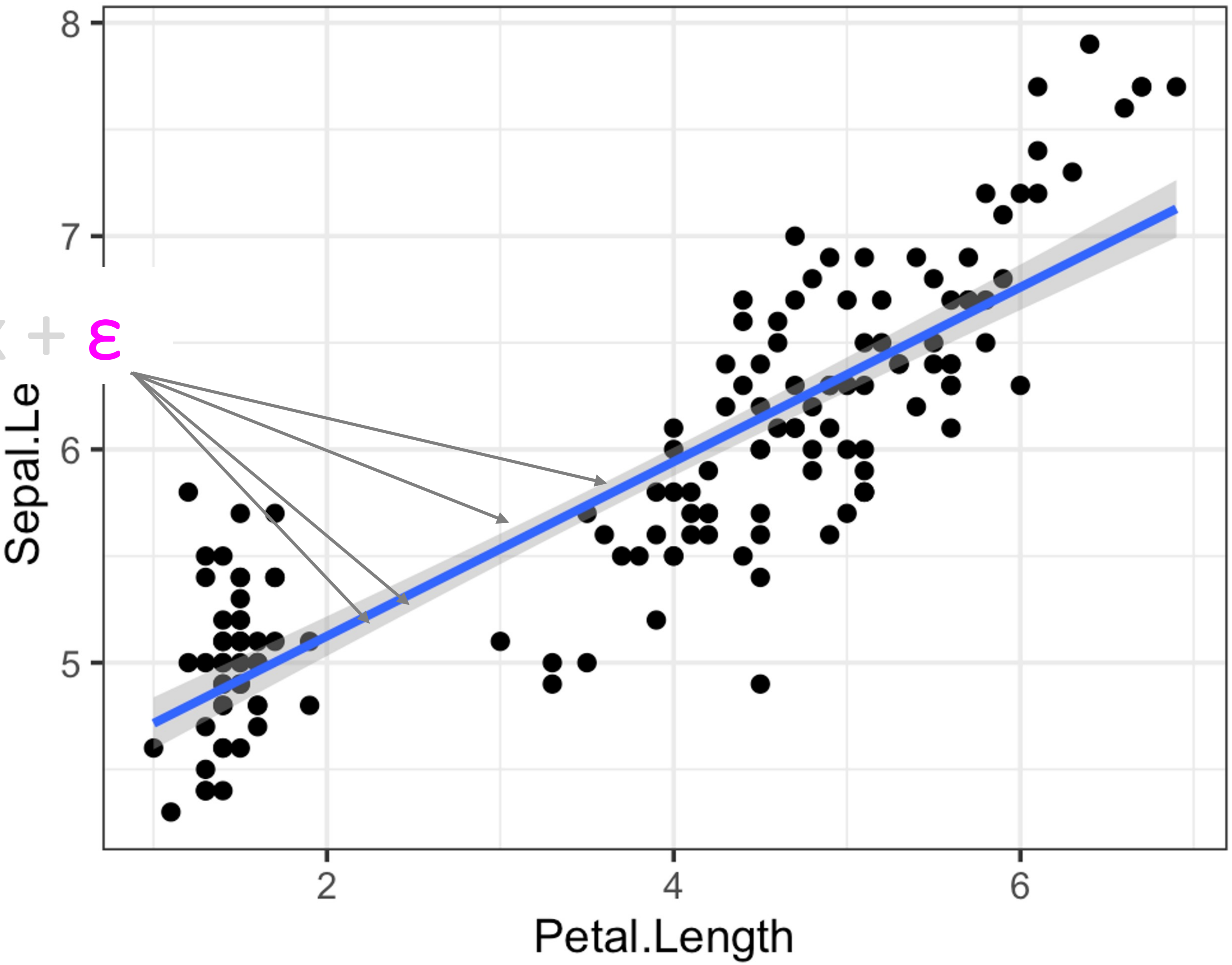


squared deviation/distance aka **sums of squares!**

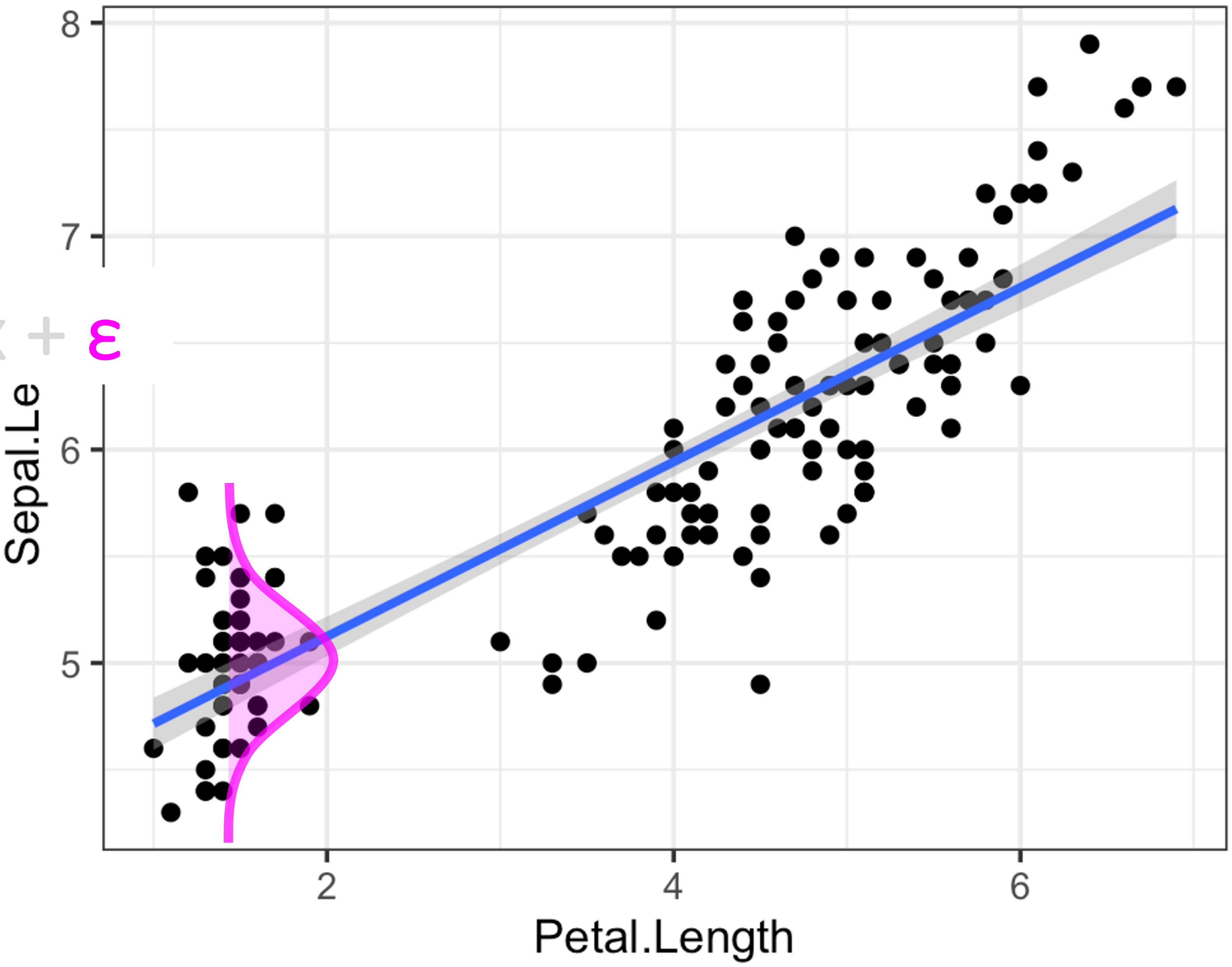


residuals
8
7
6
5
2
4
6
Sepal.Length
Petal.Length

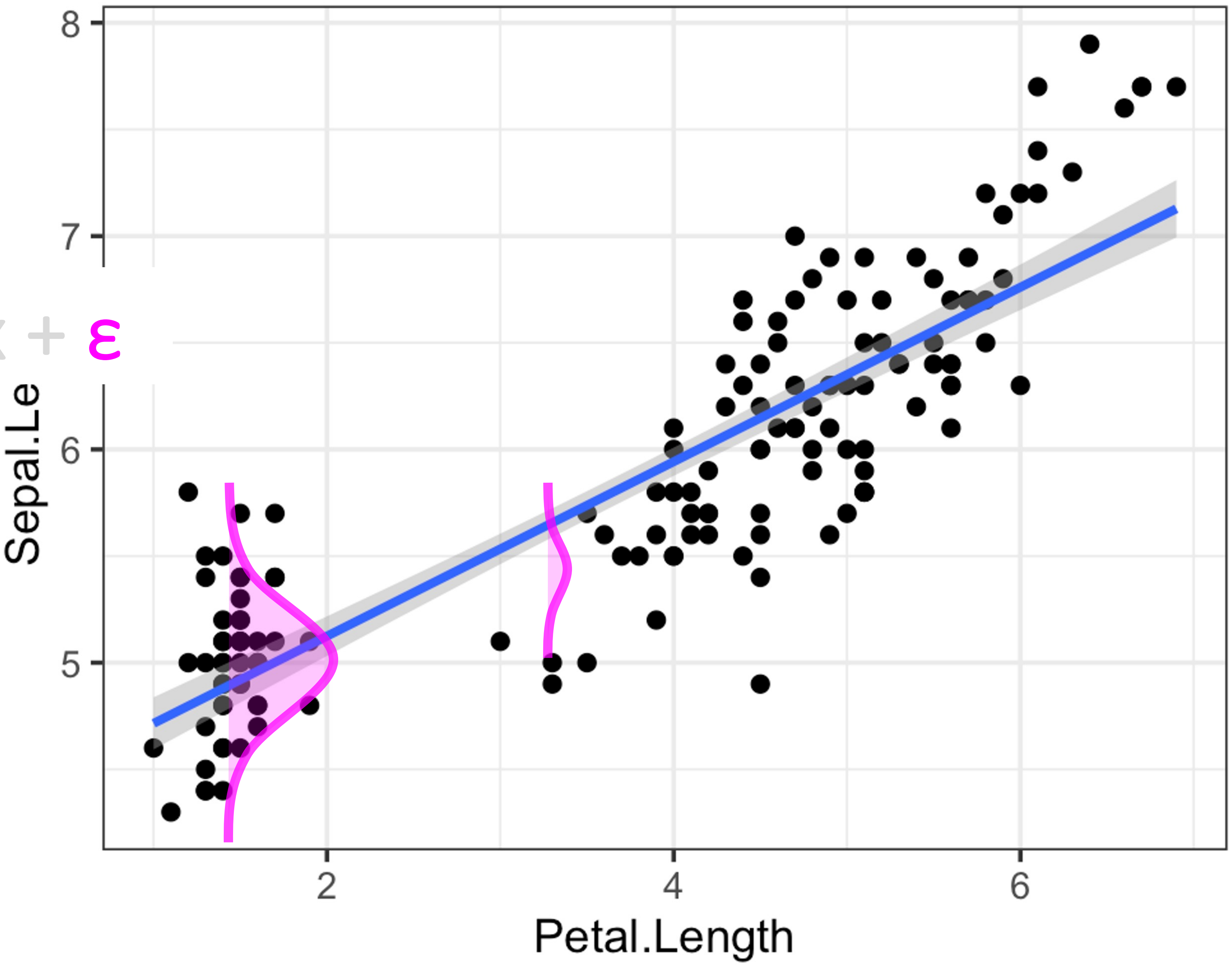
$$y = c + m * x + \varepsilon$$



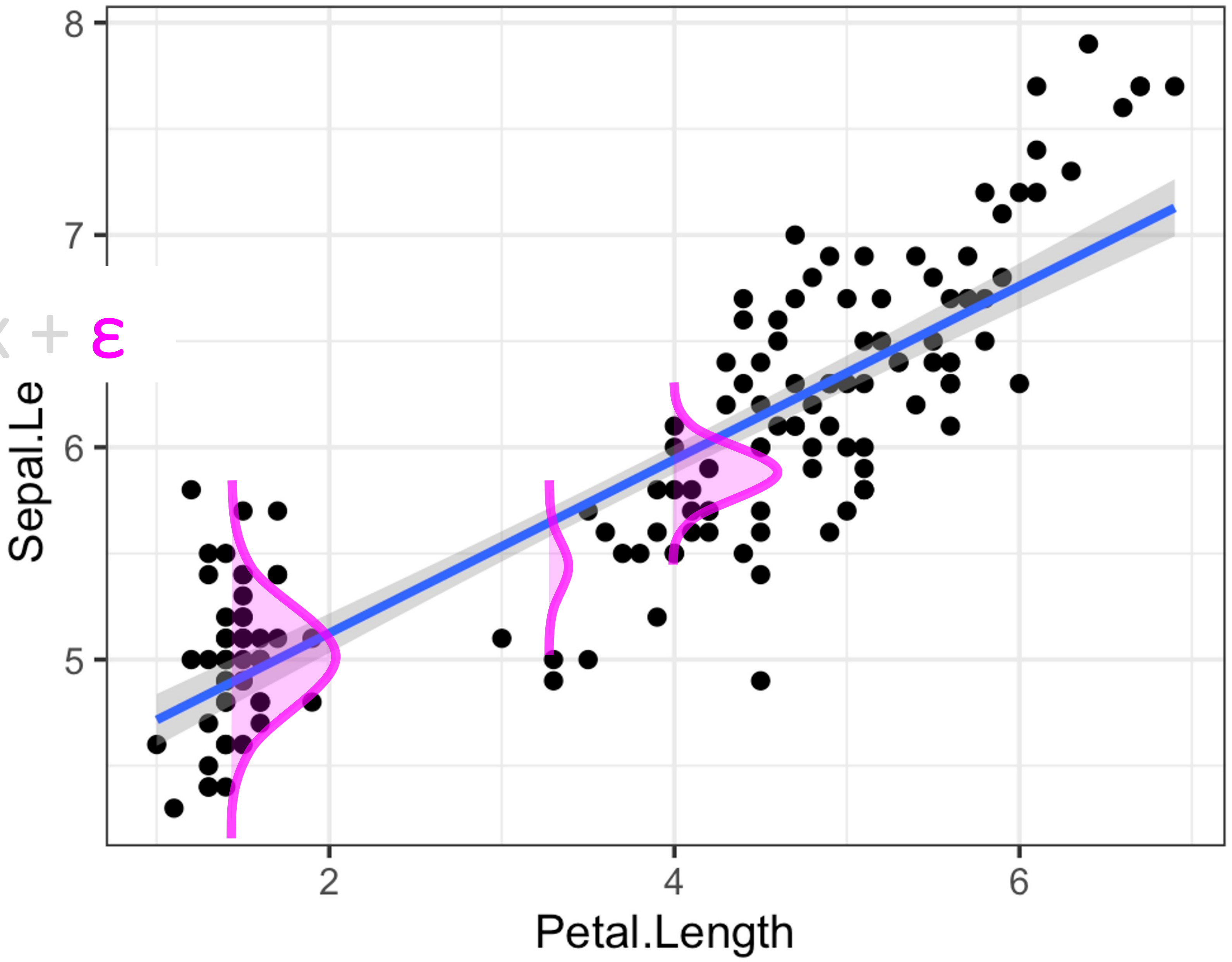
$$y = c + m * x + \varepsilon$$



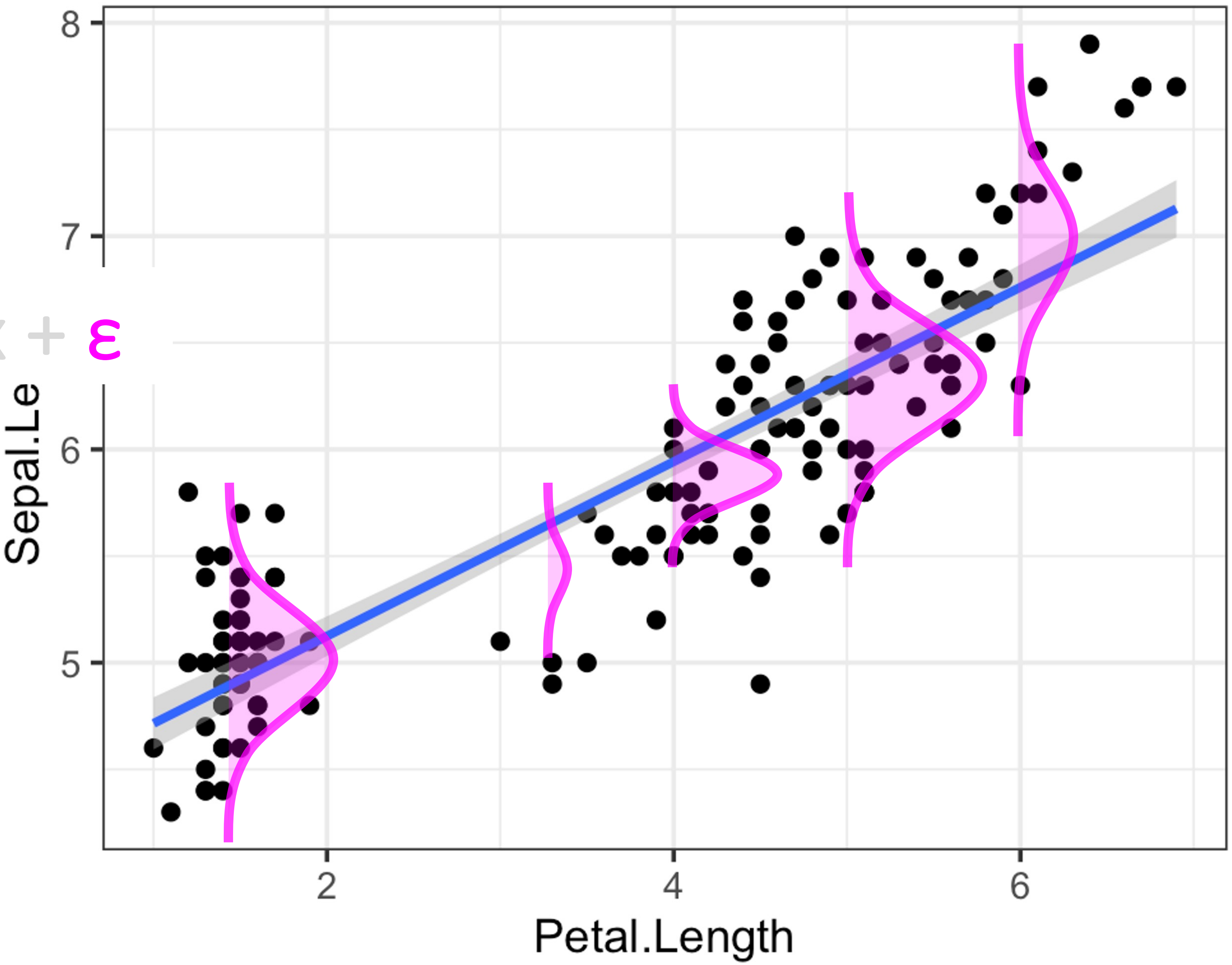
$$y = c + m * x + \varepsilon$$



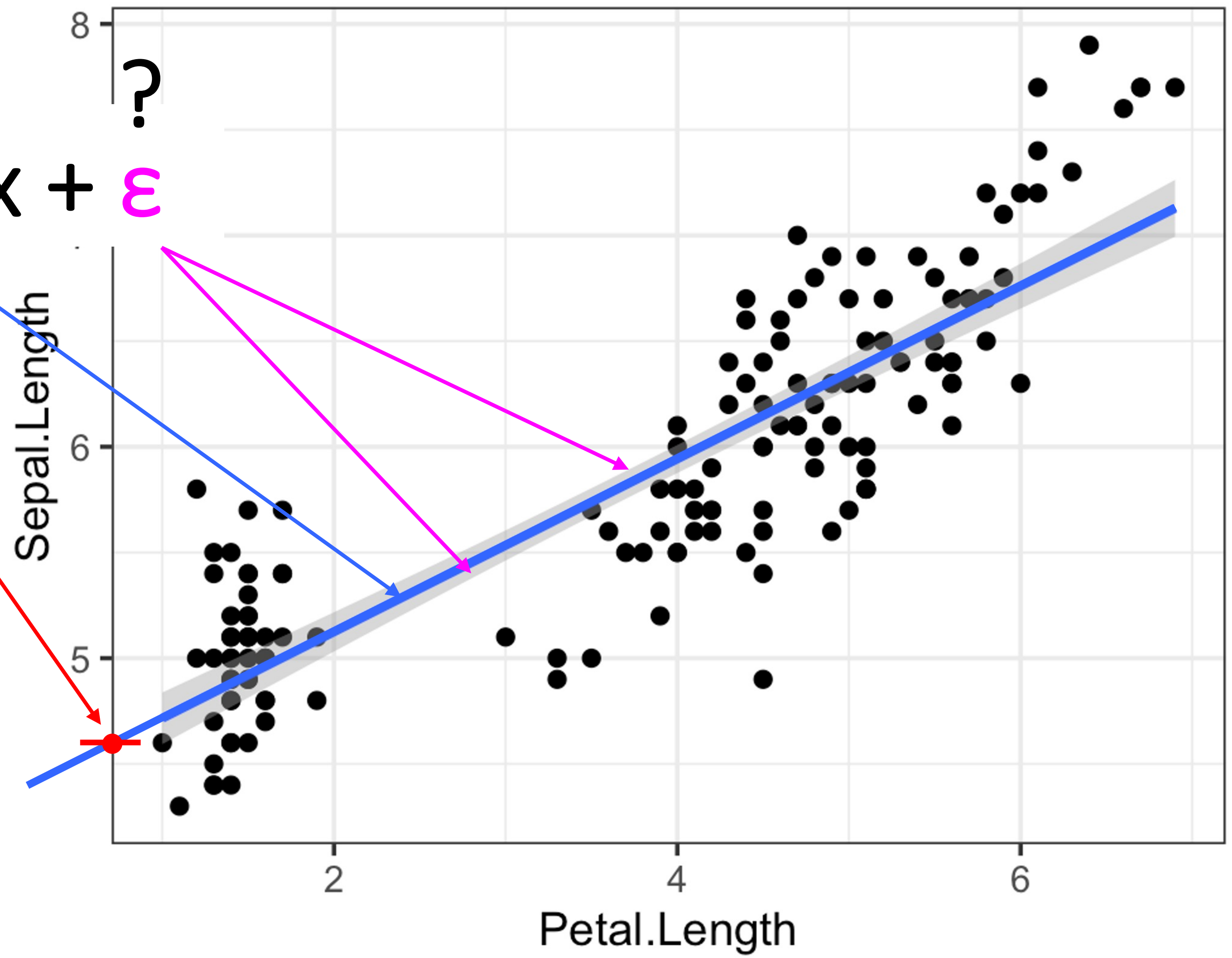
$$y = c + m * x + \varepsilon$$



$y = c + m * x + \varepsilon$



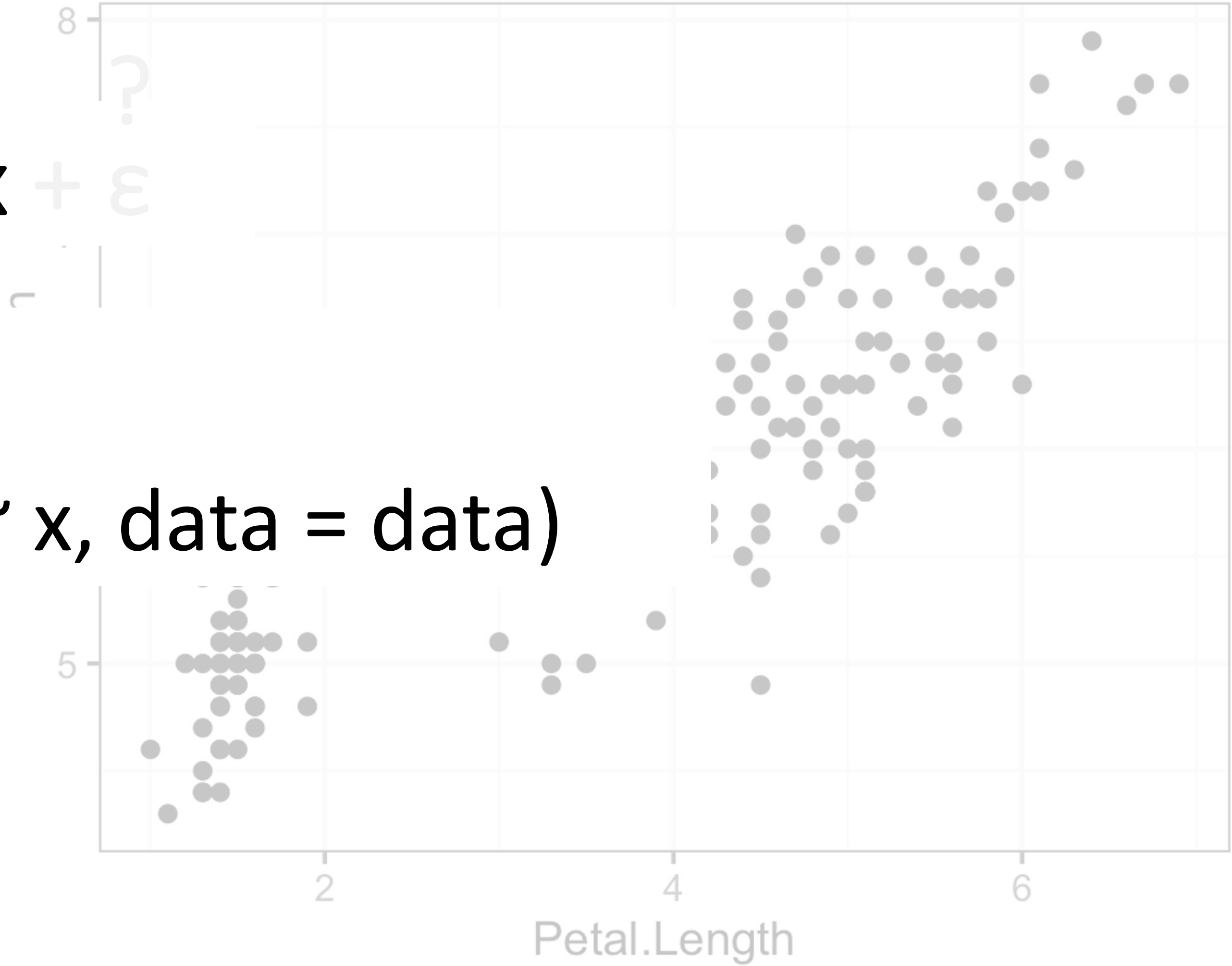
$$y = \overset{?}{c} + \overset{?}{m} * x + \overset{?}{\varepsilon}$$

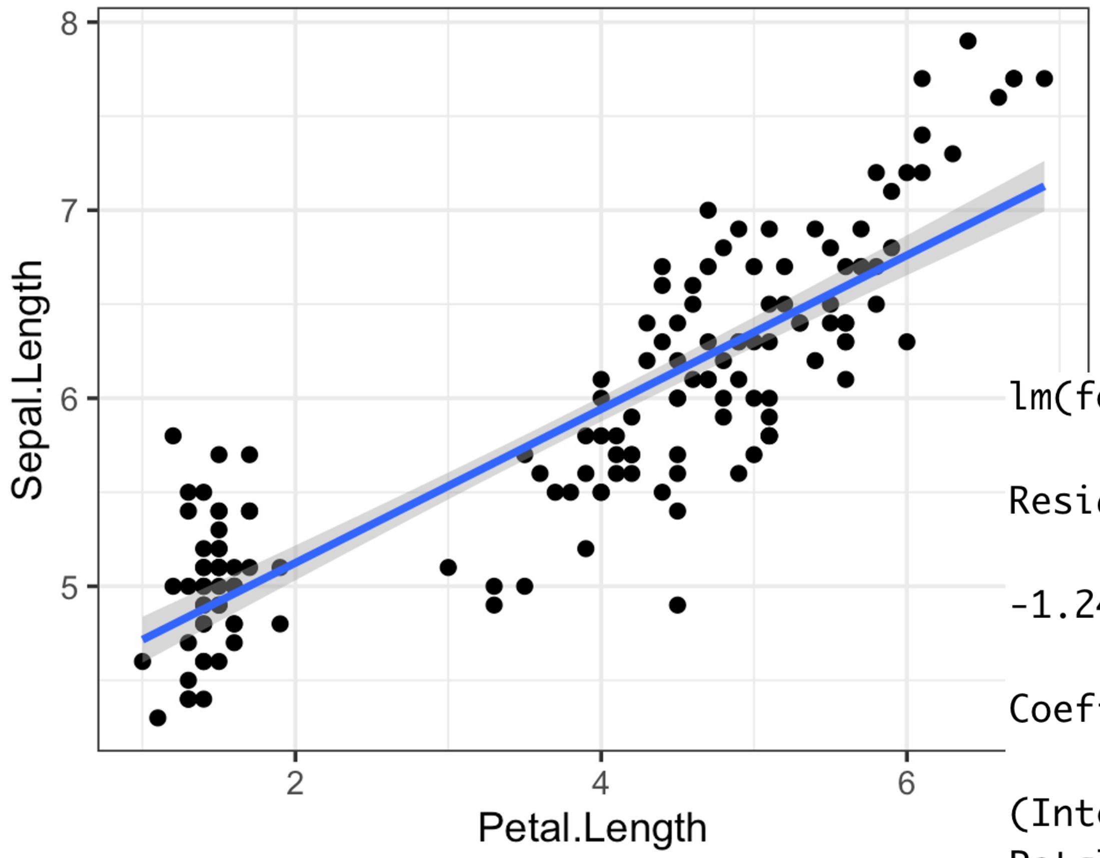


$$y \sim \overset{?}{c} + \overset{?}{m} * x + \overset{?}{\varepsilon}$$

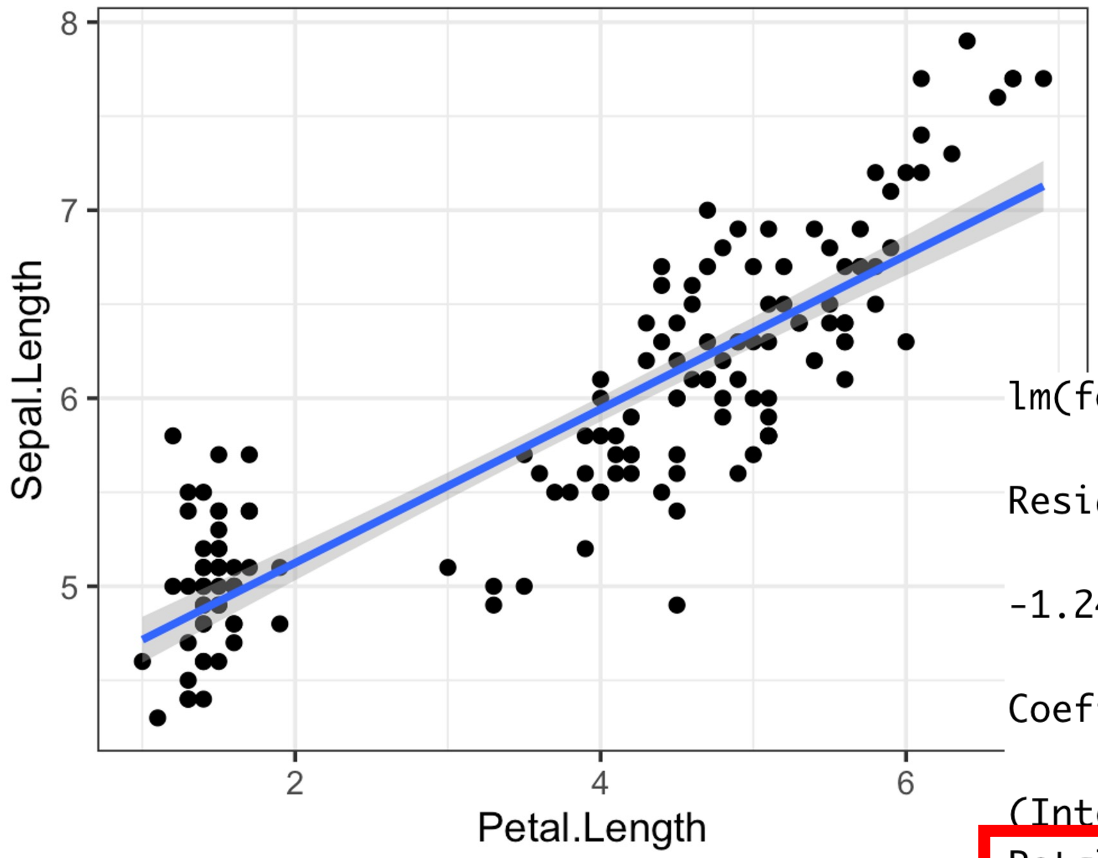
notation in R:

lmod = lm(y ~ x, data = data)

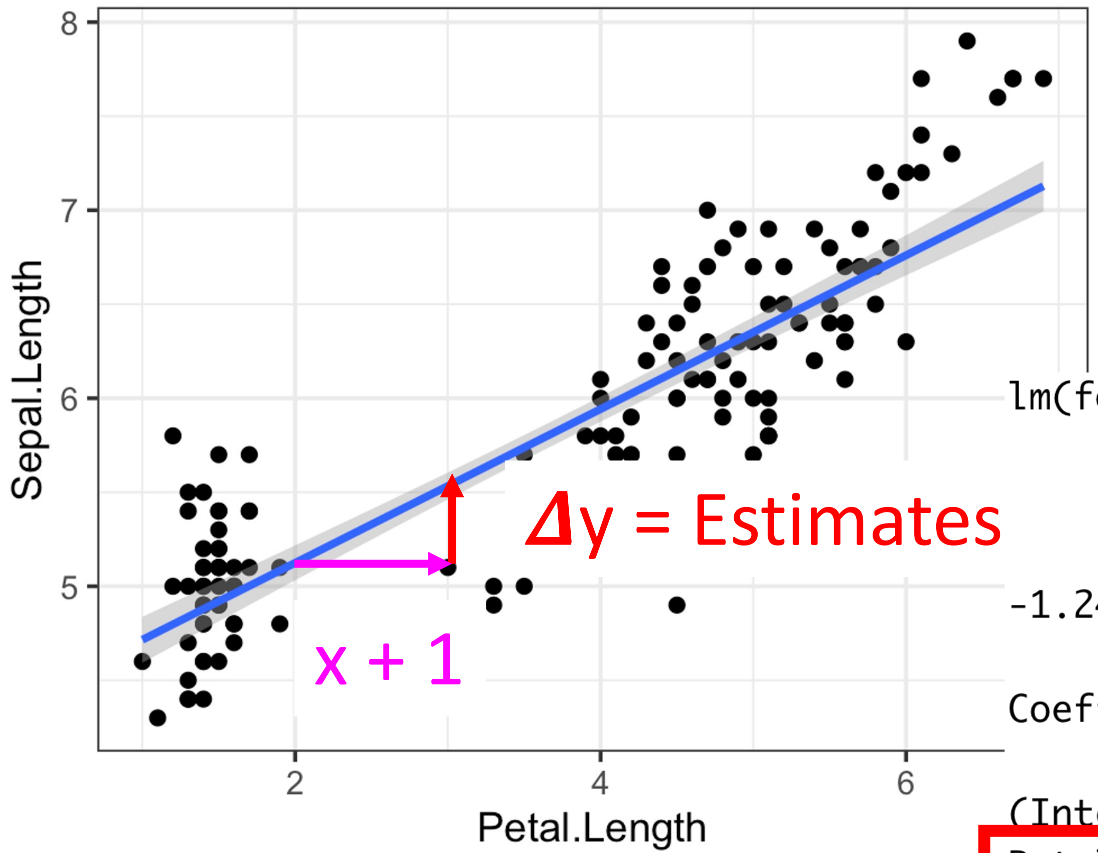




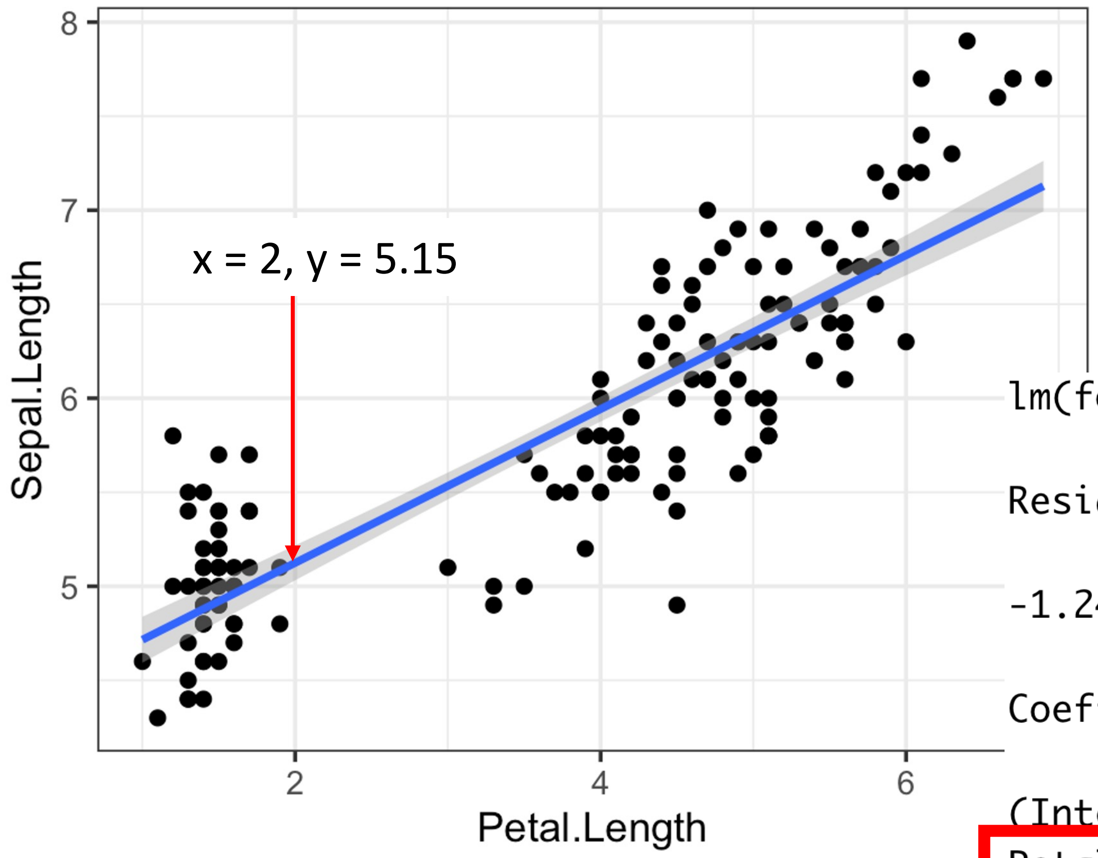

```
lm(formula = Sepal.Length ~ Petal.Length, data = iris)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.24675 -0.29657 -0.01515  0.27676  1.00269 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.30660    0.07839   54.94   <2e-16 ***
Petal.Length  0.40892    0.01889   21.65   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4071 on 148 degrees of freedom
Multiple R-squared:   0.76,	Adjusted R-squared:  0.7583 
F-statistic: 468.6 on 1 and 148 DF,  p-value: < 2.2e-16
```



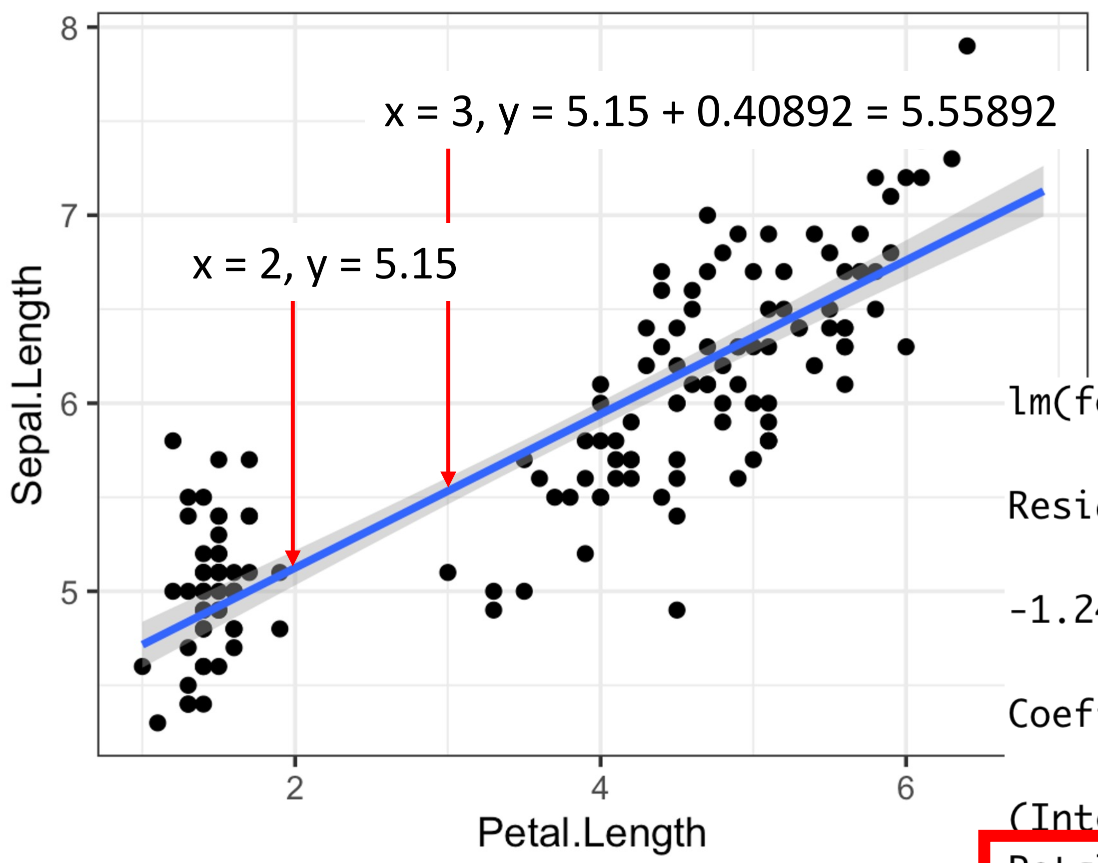

```
lm(formula = Sepal.Length ~ Petal.Length, data = iris)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.24675 -0.29657 -0.01515  0.27676  1.00269 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.30660    0.07839   54.94   <2e-16 ***
Petal.Length  0.40892    0.01889   21.65   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4071 on 148 degrees of freedom
Multiple R-squared:   0.76,     Adjusted R-squared:  0.7583 
F-statistic: 468.6 on 1 and 148 DF,  p-value: < 2.2e-16
```




```
lm(formula = Sepal.Length ~ Petal.Length, data = iris)


                     Median       3Q      Max
-1.24675 -0.29657 -0.01515  0.27676  1.00269

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.30660    0.07839   54.94   <2e-16 ***
Petal.Length  0.40892    0.01889   21.65   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4071 on 148 degrees of freedom
Multiple R-squared:   0.76,	Adjusted R-squared:  0.7583
F-statistic: 468.6 on 1 and 148 DF,  p-value: < 2.2e-16
```




```
lm(formula = Sepal.Length ~ Petal.Length, data = iris)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.24675 -0.29657 -0.01515  0.27676  1.00269 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)   4.30660    0.07839   54.94   <2e-16 ***
Petal.Length  0.40892    0.01889   21.65   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4071 on 148 degrees of freedom
Multiple R-squared:   0.76,	Adjusted R-squared:  0.7583 
F-statistic: 468.6 on 1 and 148 DF,  p-value: < 2.2e-16
```




```
lm(formula = Sepal.Length ~ Petal.Length, data = iris)

Residuals:
     Min       1Q   Median       3Q      Max
-1.24675 -0.29657 -0.01515  0.27676  1.00269

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.30660    0.07839   54.94   <2e-16 ***
Petal.Length  0.40892    0.01889   21.65   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4071 on 148 degrees of freedom
Multiple R-squared:   0.76,     Adjusted R-squared:  0.7583
F-statistic: 468.6 on 1 and 148 DF,  p-value: < 2.2e-16
```

Summary I

- Linear regression
- Regression line
- Slope & intercept
- Residual
- Standard error
- $y = c + m * x + \varepsilon$

Linear modeling using R: Part I

- linear modeling, lm()
- summary()
- car::Anova()
- ggeffects::ggemmeans()

https://osf.io/35htc/

Data Visualization

- ggplot2 useful for creating graphics
- based on The Grammar of Graphics

1. Provide data,
2. Map variables to aesthetics,
3. Declare what graphical geometries to use
4. ggplot2 generate the graphics!

1. Provide data

```
ggplot(data = iris,
aes(y = Sepal.Length, x = Petal.Length)) +
geom_point()
```

1. Provide data

2. Map variables to aesthetics

```
ggplot(data = iris,
aes(y = Sepal.Length, x = Petal.Length)) +
geom_point()
```

1. Provide data

2. Map variables to aesthetics

```
ggplot(data = iris,
aes(y = Sepal.Length, x = Petal.Length)) +
geom_point()
```

3. Declare what graphical geometries to use

1. Provide data

2. Map variables to aesthetics

```
ggplot(data = iris,
aes(y = Sepal.Length, x = Petal.Length)) +
geom_point()
```

3. Declare what graphical geometries to use

aes(x, y, color, fill, shape, etc.)

geom_[point(), smooth(), bar(), line(), ribbon(), boxplot(), histogram(), etc]

Minimum to ggplot!

3. Declare what graphical geometries to use

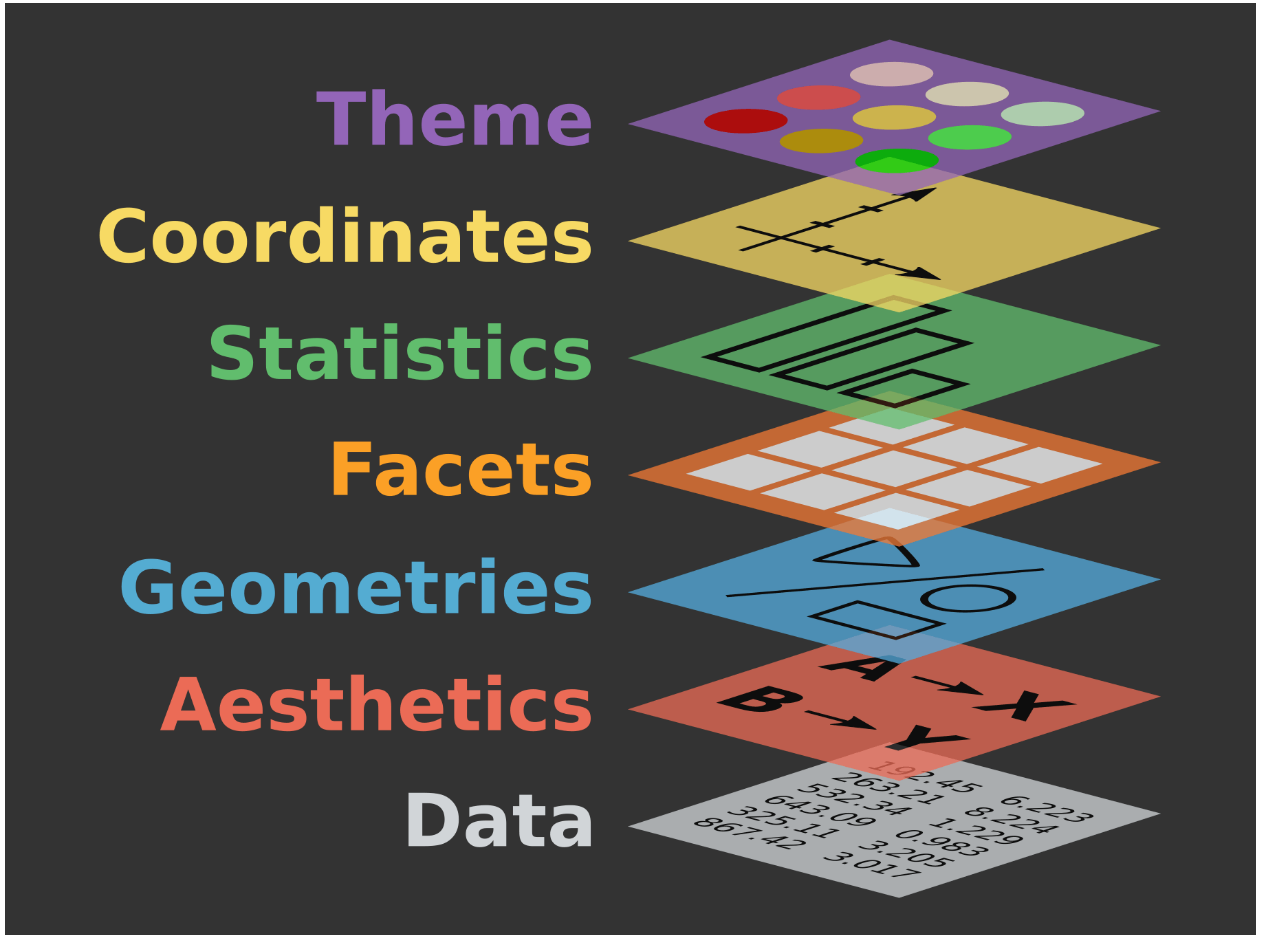
2. Map variables to aesthetics

1. Provide data

3. Declare what graphical geometries to use

2. Map variables to aesthetics

1. Provide data



Define overall visuals

Set limits for graph

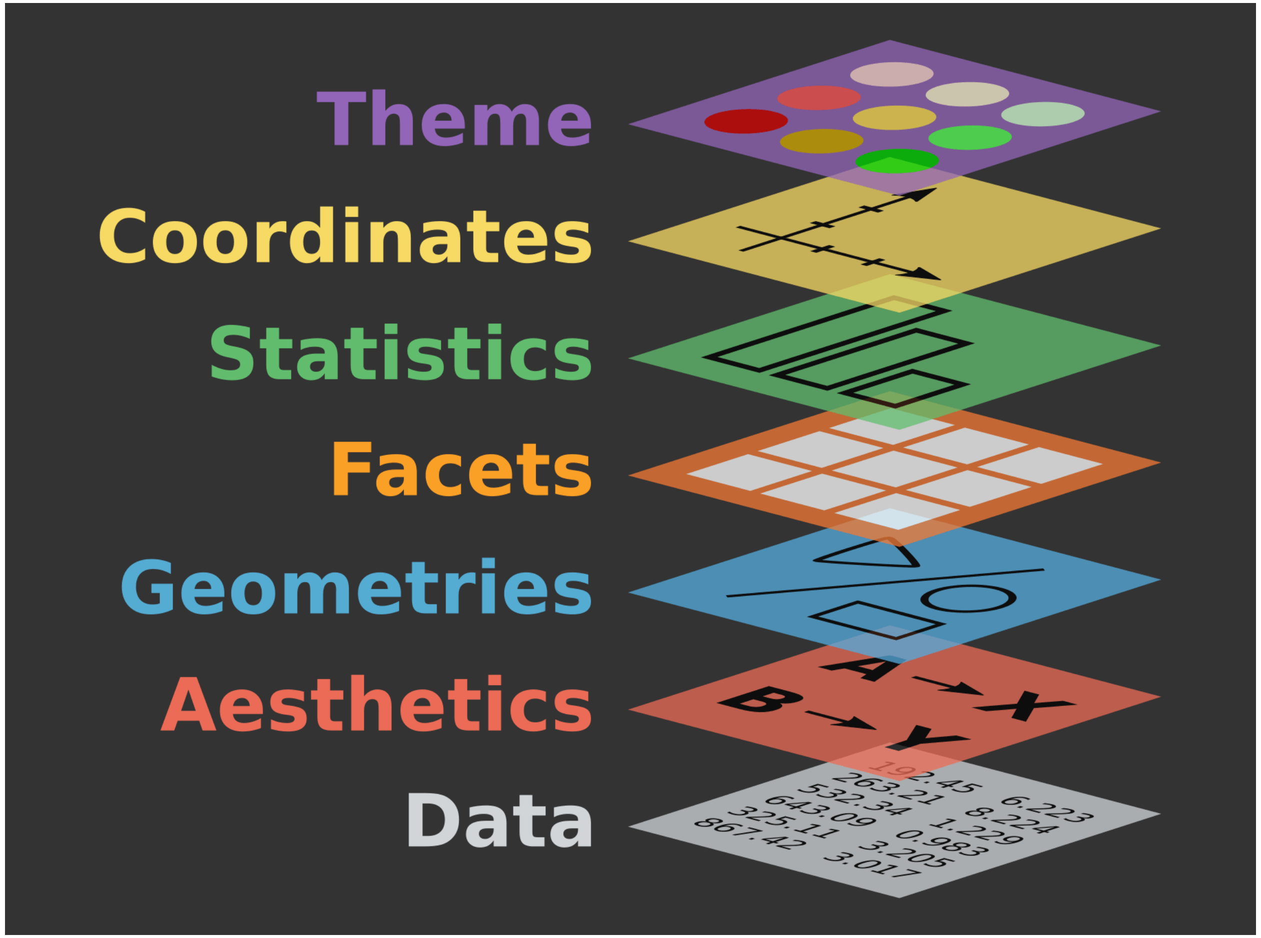
Summarise data

Plot data into facets

3. Declare what graphical geometries to use

2. Map variables to aesthetics

1. Provide data



Google!

adjust axis limit ggplot in r

All · Videos · Images · Books · News · More · Tools

About 548,000 results (0.48 seconds)

https://ggplot2.tidyverse.org › reference › lims

Set scale limits — lims • ggplot2

This is a shortcut for supplying the **limits** argument to the individual scales. By default, any values outside the **limits** specified are replaced with NA.

Scale_x_continuous · Coord_cartesian · Scale_x_discrete · Expand_limits

People also ask

How to set limits for axes in ggplot2 R plots?

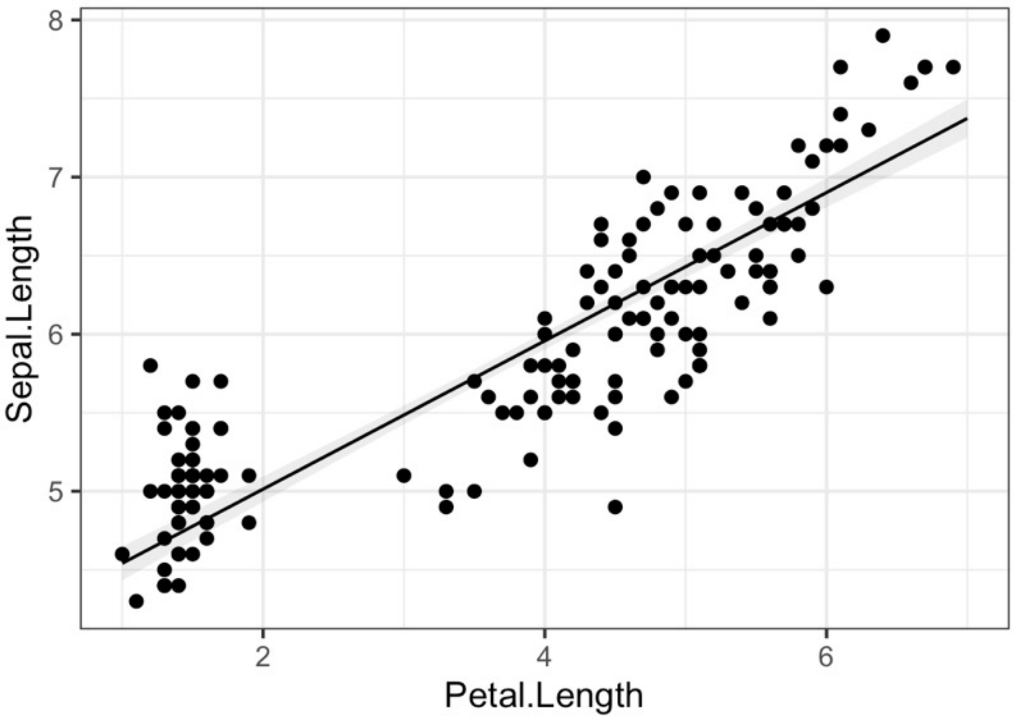
How do I change the axis limit in R?

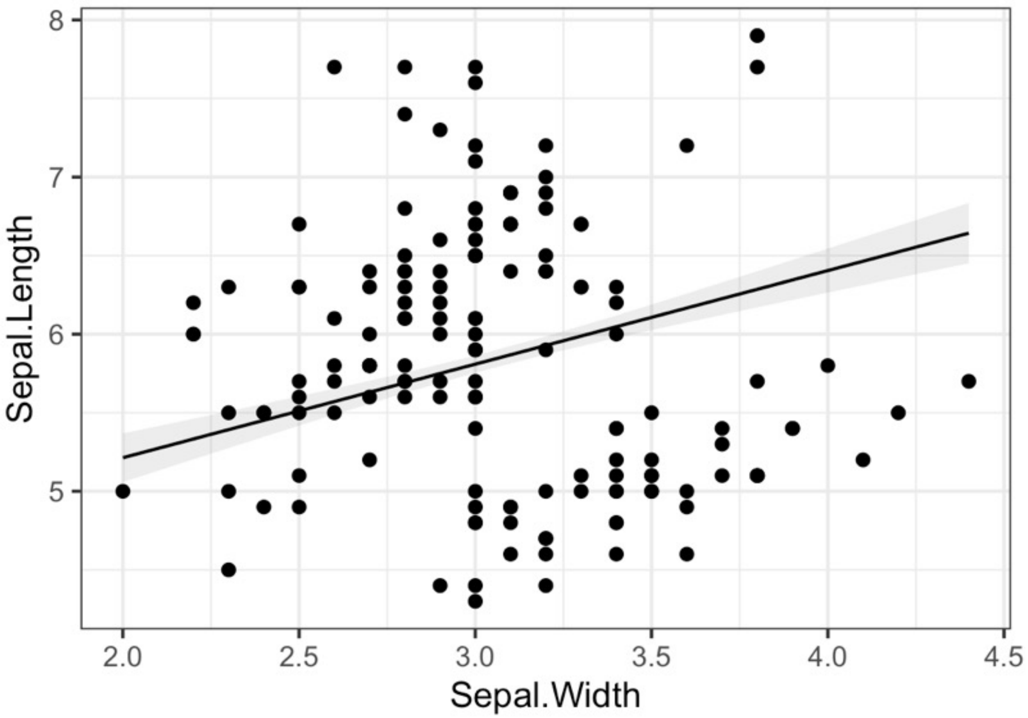
How to specify y axis range in ggplot2?

Which function used for setting axis limits in any graph in R?

Feedback

```
mlmod <- lm(Sepal.Length ~ Petal.Length + Sepal.Width, data = iris)
```





multiple linear regression

DAY 3

Winter Workshop:

Basic and Intermediate Statistics with R

Chai Jun Ho, PhD

CHILD. LANGUAGE. LAB.

CHOSUN UNIVERSITY 1946

조선대학교
CHOSUN UNIVERSITY

Schedule

~~Day 1~~	~~Day 2~~	Day 3
~~Introduction to Statistics~~	~~Linear Model I~~	Linear Model II
~~Introduction to R~~	~~Data Visualization~~	Mixed-Effect Model

summary(lmod)

```
Call:
Sepal.Length ~ Sepal.Width, data = iris)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.5561 -0.6333 -0.1120  0.5579  2.2226 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)   6.5262     0.4789   13.63   <2e-16 ***
Sepal.Width  -0.2234     0.1551   -1.44    0.152    
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8251 on 148 degrees of freedom
Multiple R-squared:  0.01382,	Adjusted R-squared:  0.007159 
F-statistic: 2.074 on 1 and 148 DF,  p-value: 0.1519
```

slope

p-value

```
Call:
lm(formula = Sepal.Length ~ Sepal.Width, data = iris)

Resi
    Min      1Q  Median      3Q     Max
-1.5561 -0.6333 -0.1120  0.5579  2.2226

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   6.5262     0.4789   13.63   <2e-16 ***
Sepal.Width  -0.2234     0.1551   -1.44    0.152
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8251 on 148 degrees of freedom
Multiple R-squared:  0.01382,	Adjusted R-squared:  0.007159
F-statistic: 2.074 on 1 and 148 DF,  p-value: 0.1519
```

summary()

```
Call:
lm(formula = Sepal.Length ~ Species, data = iris)

duals:
    Min      1Q  Median      3Q     Max 
-1.6880 -0.3285 -0.0060  0.3120  1.3120 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5148 on 147 degrees of freedom
Multiple R-squared:  0.6187,	Adjusted R-squared:  0.6135 
F-statistic: 119.3 on 2 and 147 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Sepal.Length ~ Species, data = iris)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.6880 -0.3285 -0.0060  0.3120  1.3120 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
---
Signif. codes:  
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5148 on 147 degrees of freedom
Multiple R-squared:  0.6187,	Adjusted R-squared:  0.6135 
F-statistic: 119.3 on 2 and 147 DF,  p-value: < 2.2e-16
```

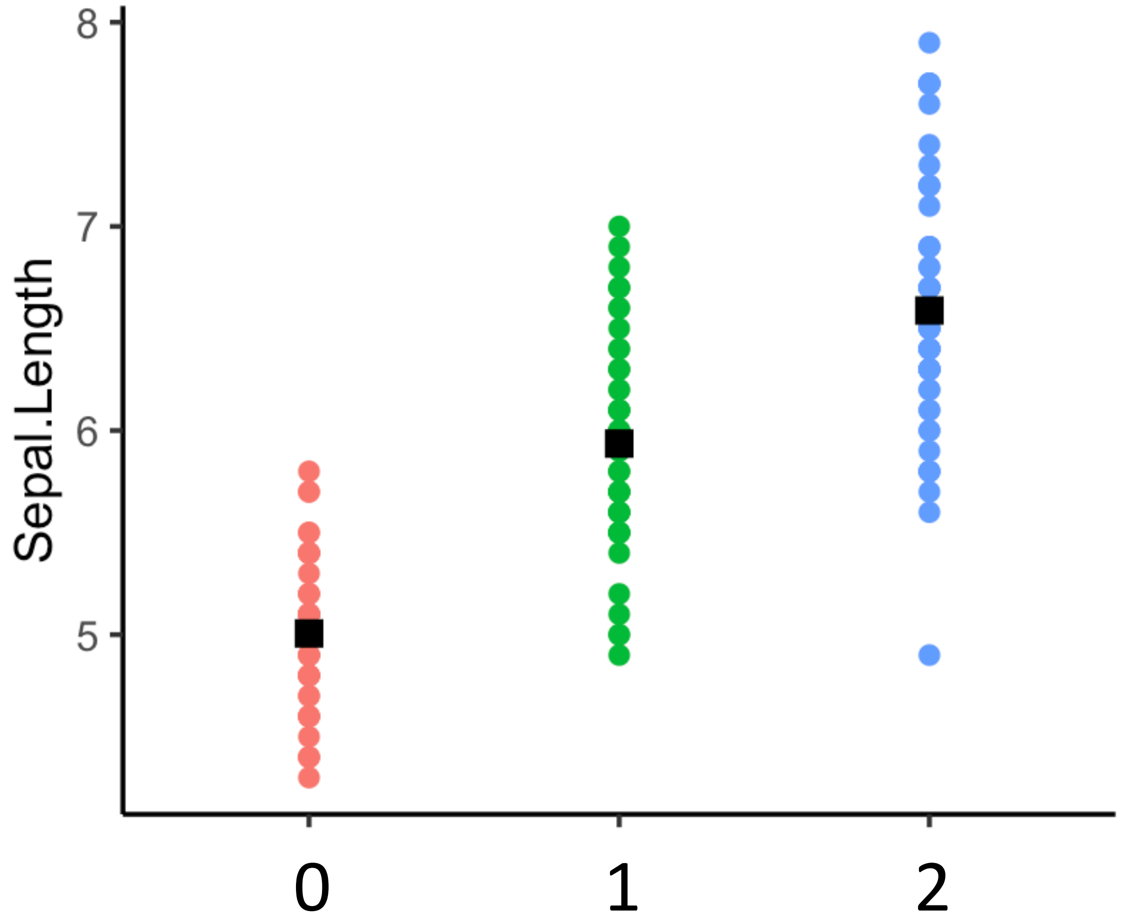
```
Call:
lm(formula = Sepal.Length ~ Species, data = iris)

Residuals:
    Min      1Q  Median      3Q     Max 
-1.6880 -0.3285 -0.0060  0.3120  1.3120 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
---
Signif. codes:  
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5148 on 147 degrees of freedom
Multiple R-squared:  0.6187,	Adjusted R-squared:  0.6135 
F-statistic: 119.3 on 2 and 147 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = Sepal.Length ~ Species, data = iris)

Residuals:      ???
    Min      1Q  Median      3Q     Max
-1.6880 -0.3285 -0.0060  0.3120  1.3120

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5148 on 147 degrees of freedom
Multiple R-squared:  0.6187,    Adjusted R-squared:  0.6135
F-statistic: 119.3 on 2 and 147 DF,  p-value: < 2.2e-16
```

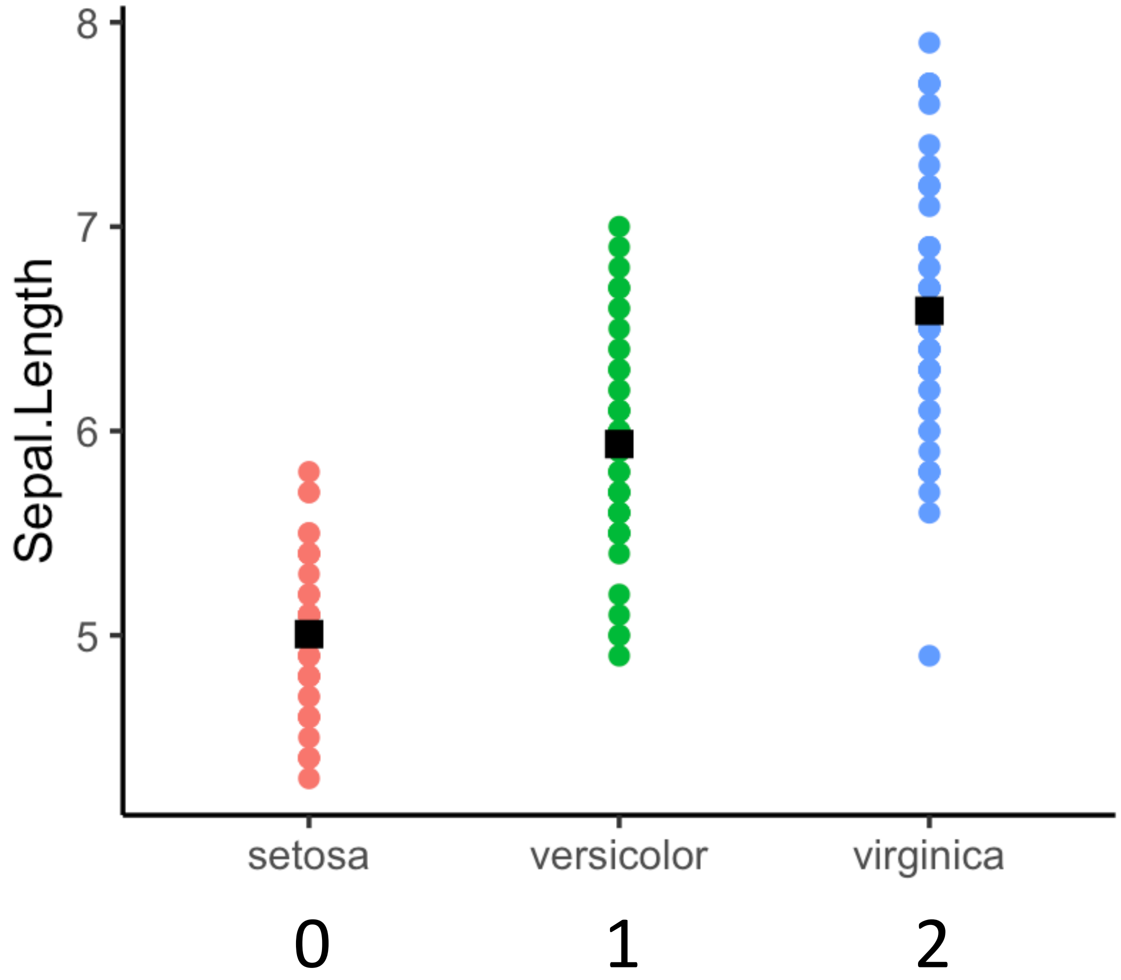
```
Call:
lm(formula = Sepal.Length ~ Species, data = iris)
```

Speciessetosa ????

```
    3Q     Max
..120  1.3120

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5148 on 147 degrees of freedom
Multiple R-squared:  0.6187,    Adjusted R-squared:  0.6135
F-statistic: 119.3 on 2 and 147 DF,  p-value: < 2.2e-16
```



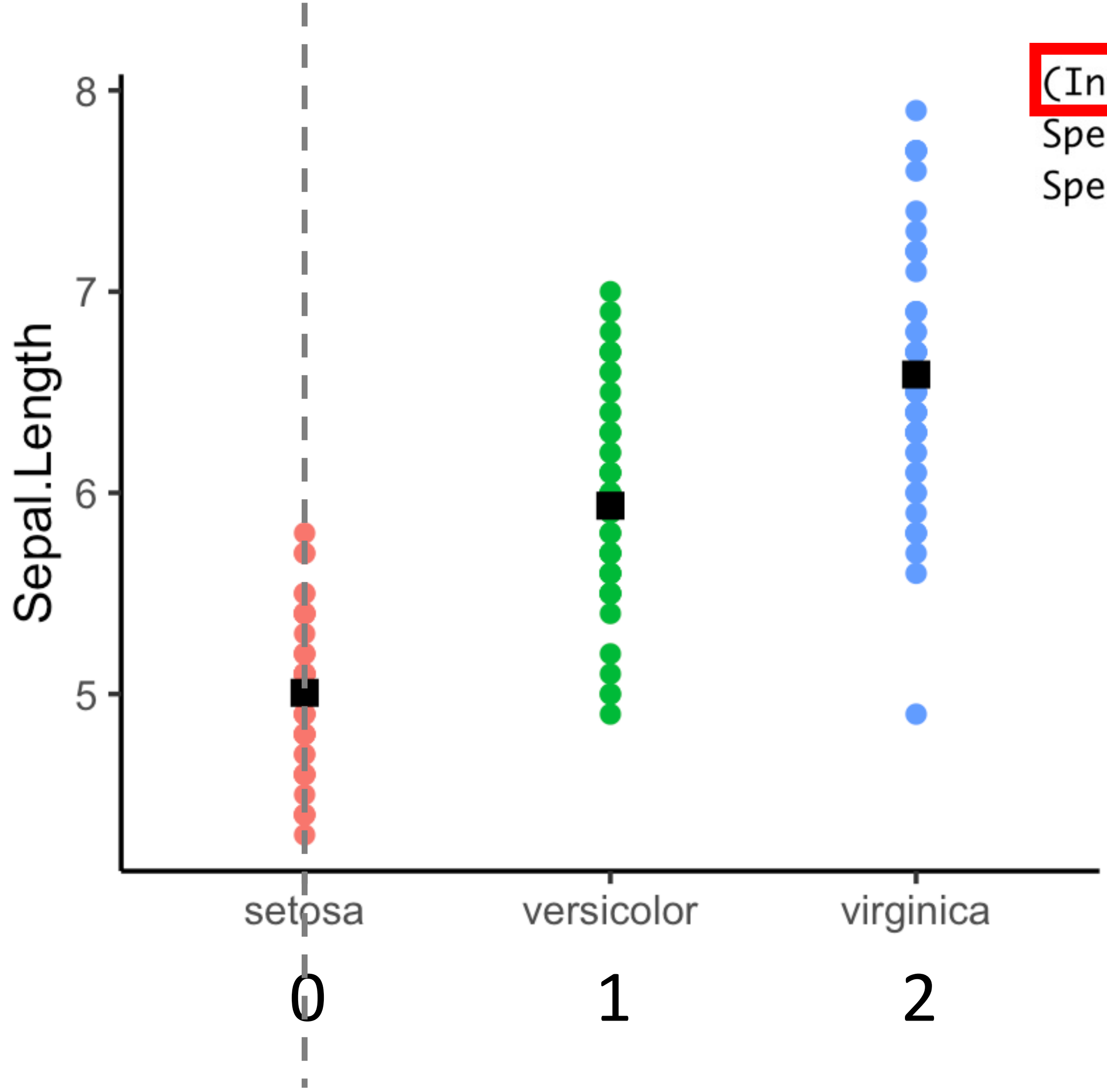
Sepal.Length
8
7
6
5
0
1
2




```
Coefficients:
                   Estimate Std. Error t
(Intercept)          5.0060     0.0728  6
Speciesversicolor    0.9300     0.1030
Speciesvirginica     1.5820     0.1030  1
```

“intercept”

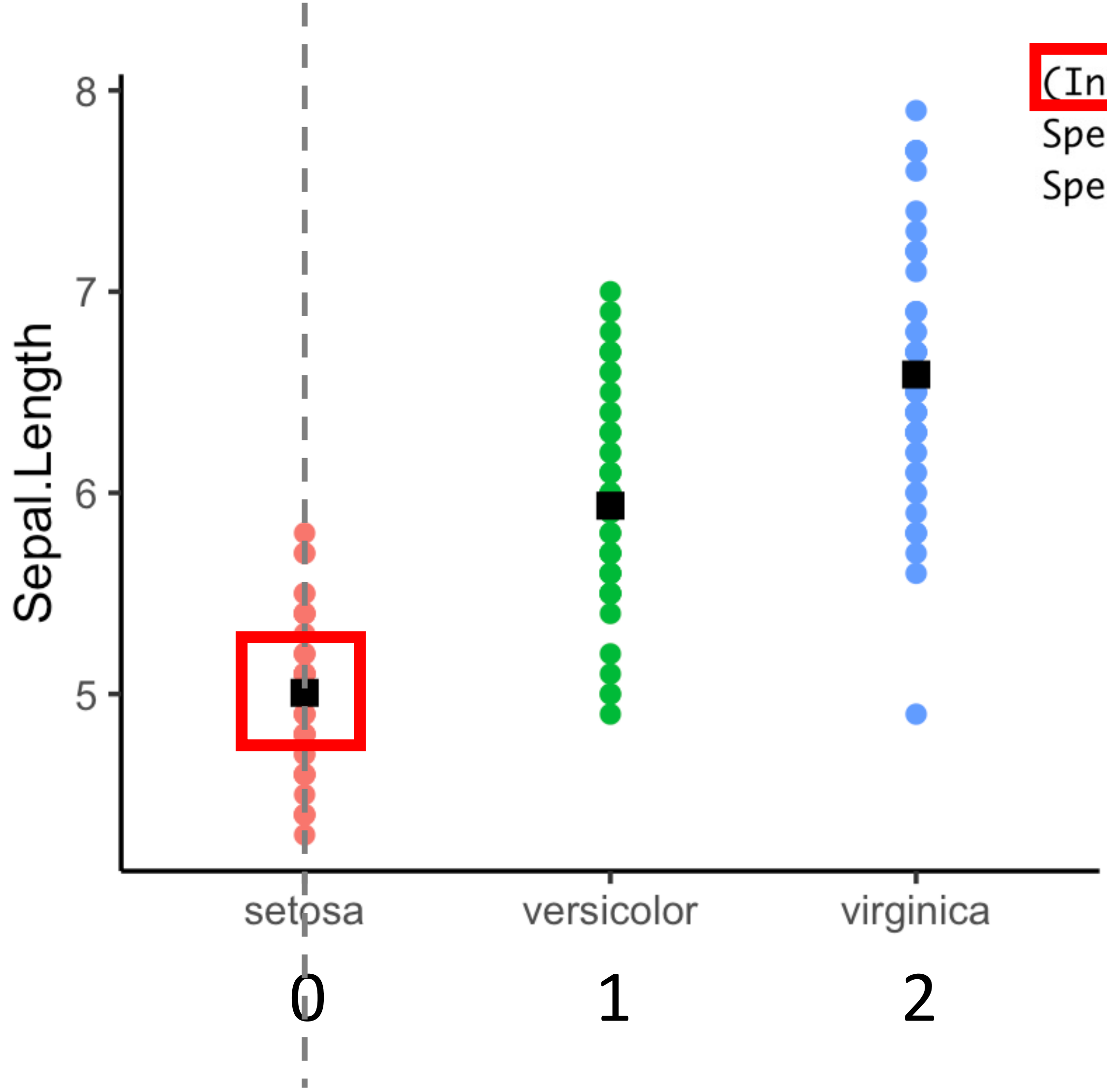
point of reference




```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
```

“intercept”

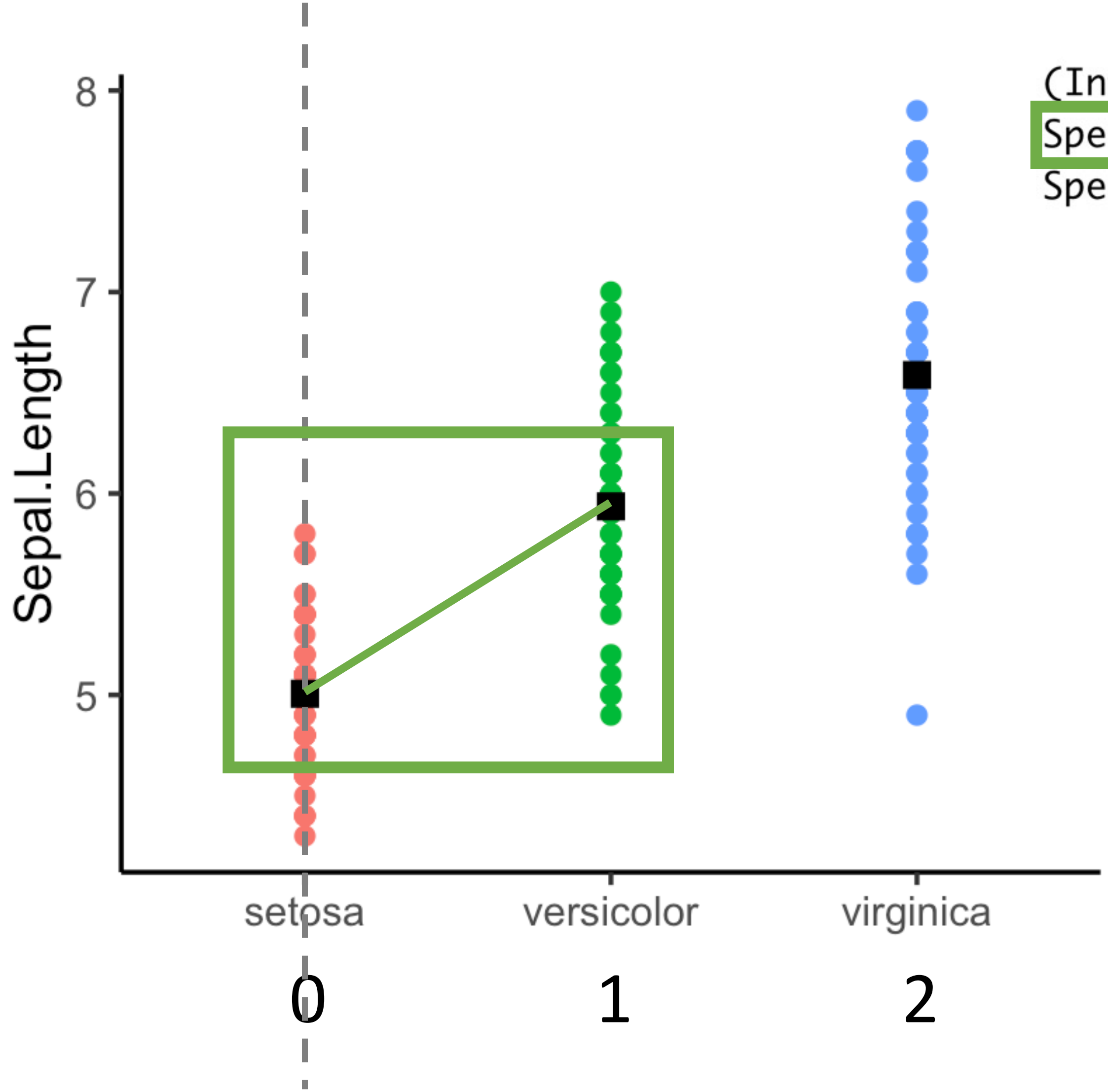
point of reference




```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
```

“intercept”

point of reference

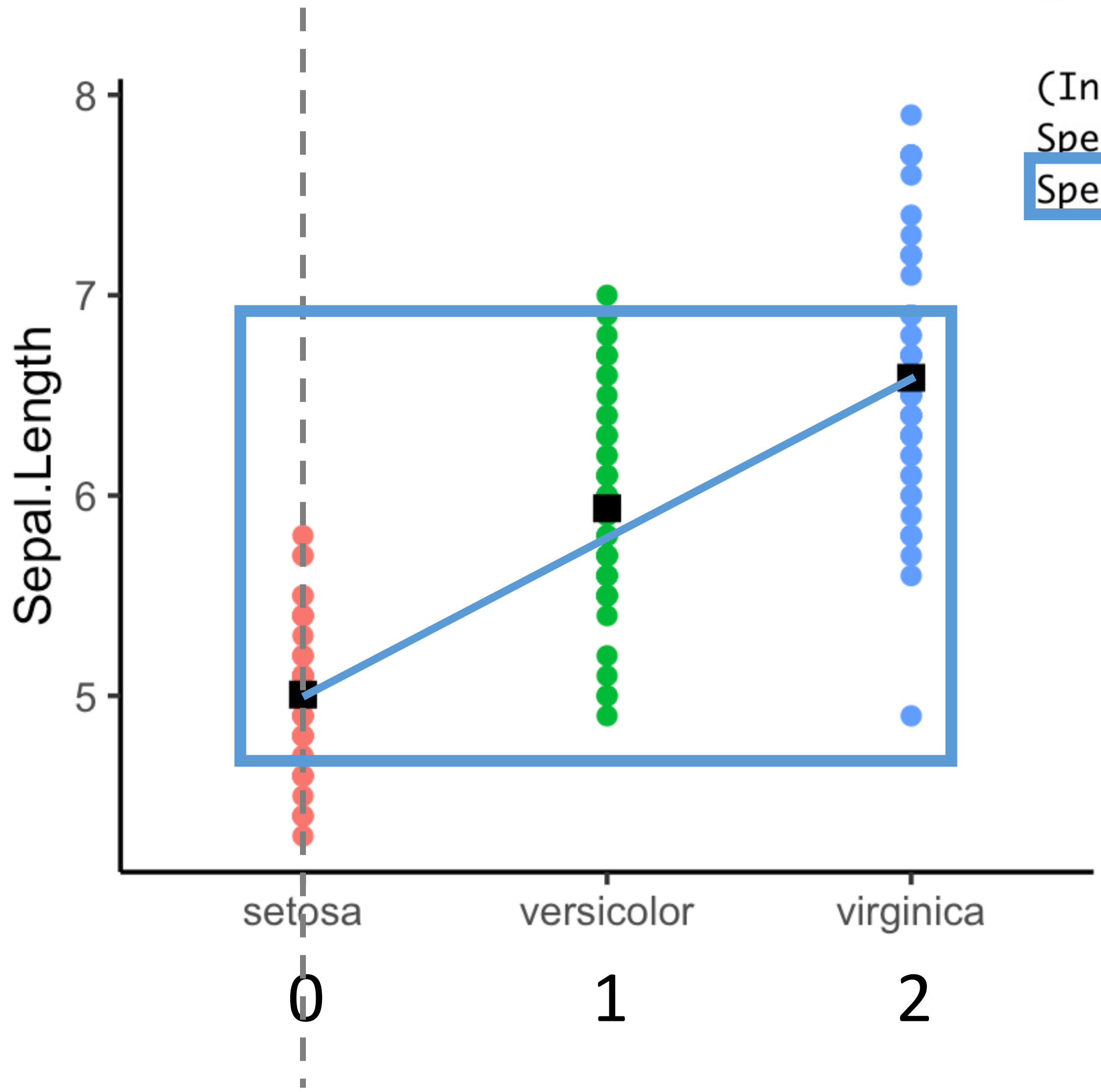



```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
```

5.0060 +
0.9300 =
5.9360

“intercept”

point of reference

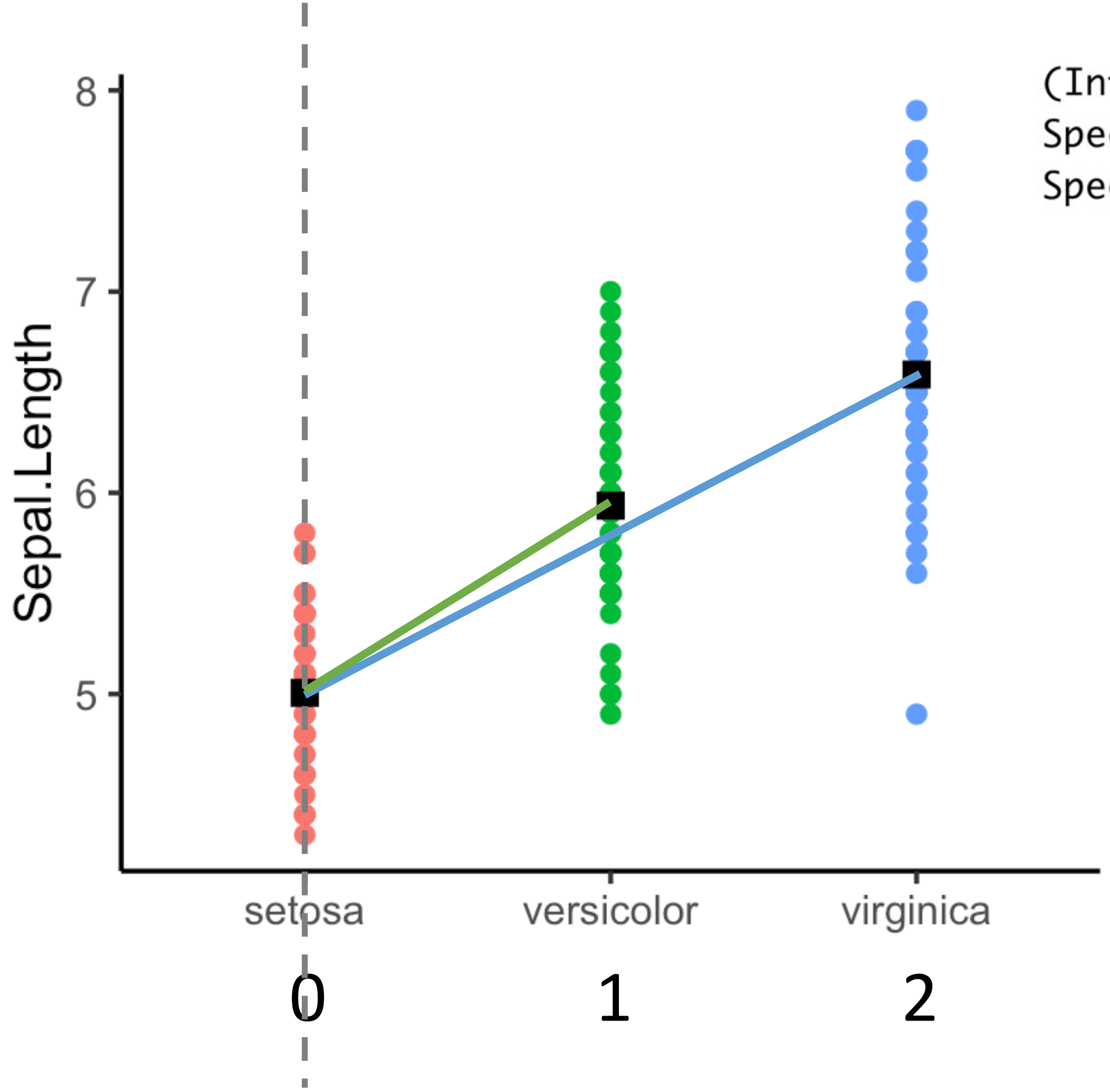



```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
```

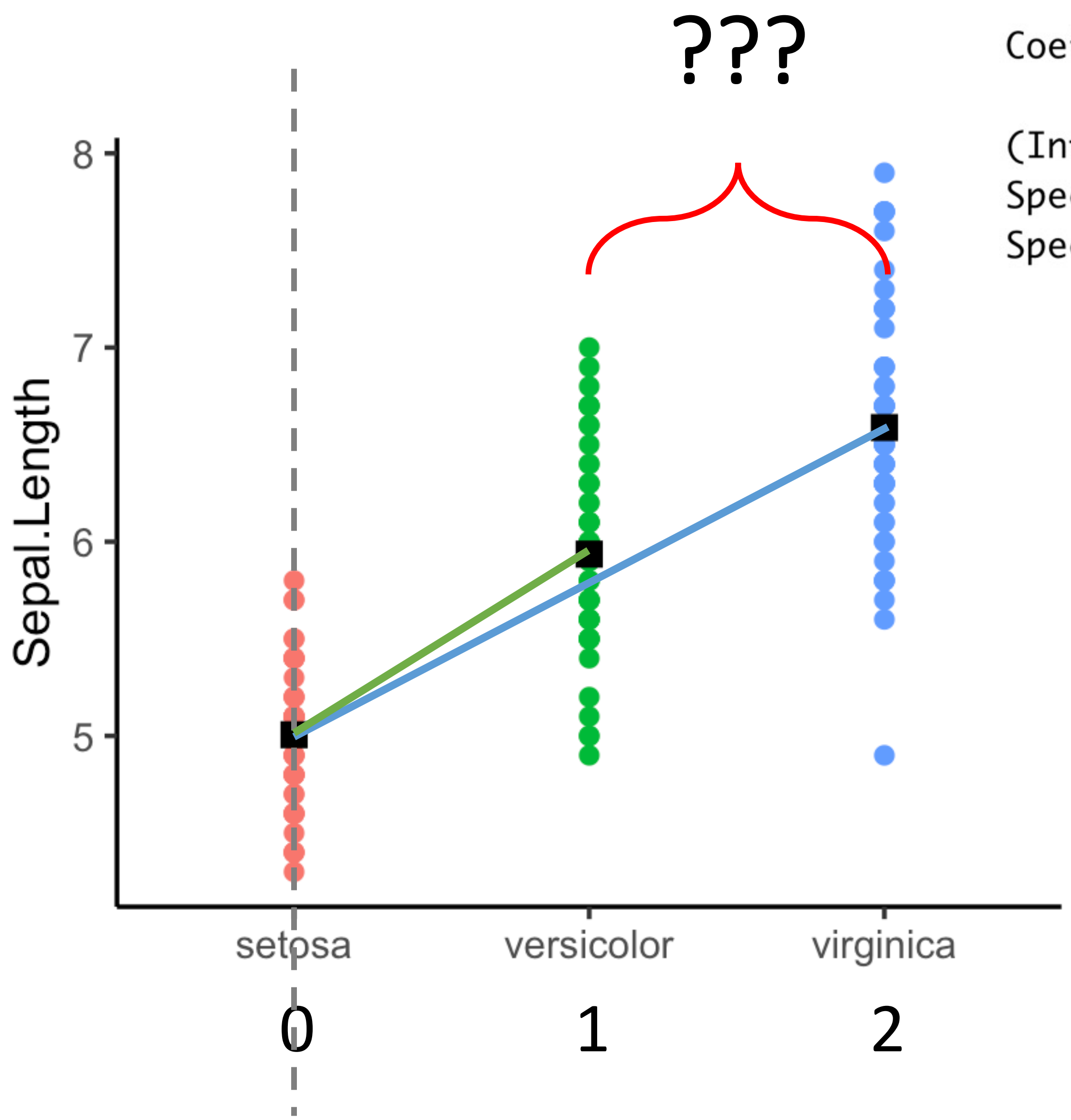
$$5.0060 + 1.5820 = 6.5880$$

“intercept”

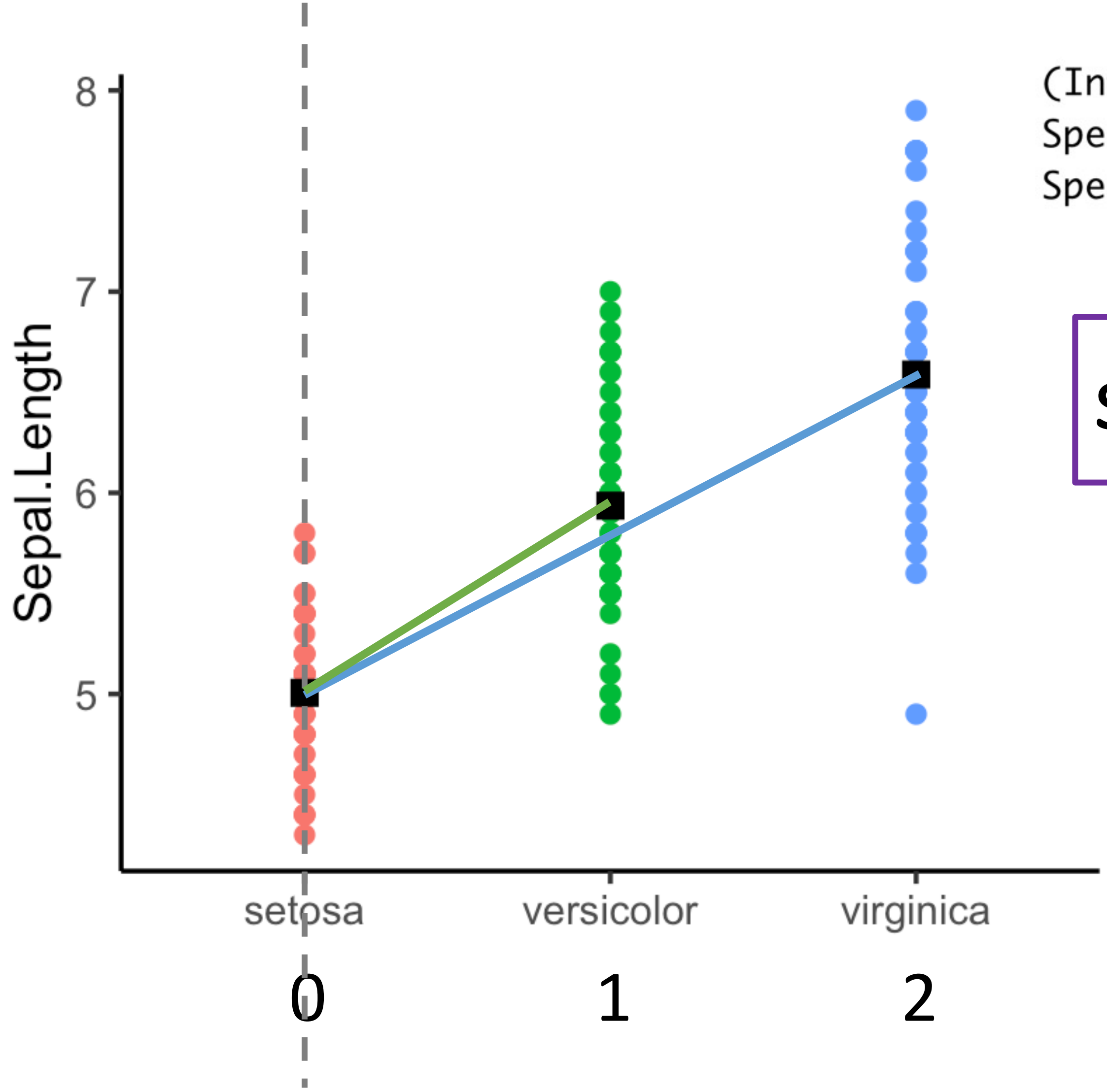
point of reference




```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
```




```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
```




```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
```

summary()

vs

car::Anova()

lmcat <- lm(Sepal.Length ~ Species, data = iris)

```
> car::Anova(lmcat, type = "III")
Anova Table (Type III tests)

Response: Sepal.Length
             Sum Sq  Df F value    Pr(>F)
(Intercept) 1253.00   1 4728.16 < 2.2e-16 ***
Species       63.21   2  119.26 < 2.2e-16 ***
Residuals     38.96 147
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

```
> car::Anova(lmcat, type = "III")
Anova Table (Type III tests)

Response: Sepal.Length
             Sum Sq  Df F valu
(Intercept) 1253.00   1 4728.1
Species       63.21   2  119.26 < 2.2e-16 ***
Residuals     38.96 147
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Significant main effect of Species,
$F(2, 147) = 119.26$, $p < .001$

Setosa – Versicolor = ???

Versicolor – Virginia = ???

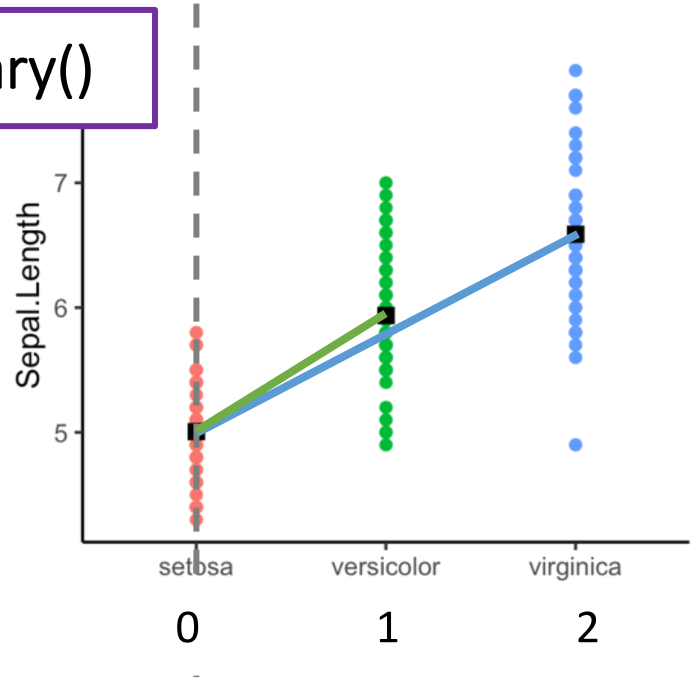
Setosa – Virginia = ???

```
> car::An
Anova Tab

Response:
                  Sum Sq  Df  F value    Pr(>F)
(Intercept) 1253.00    1 4728.16 < 2.2e-16 ***
Species       63.21    2  119.26 < 2.2e-16 ***
Residuals     38.96  147
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$\frac{\text{variance } \textbf{between groups}}{\text{variance } \textbf{within group}} = \textbf{F-ratio}$$

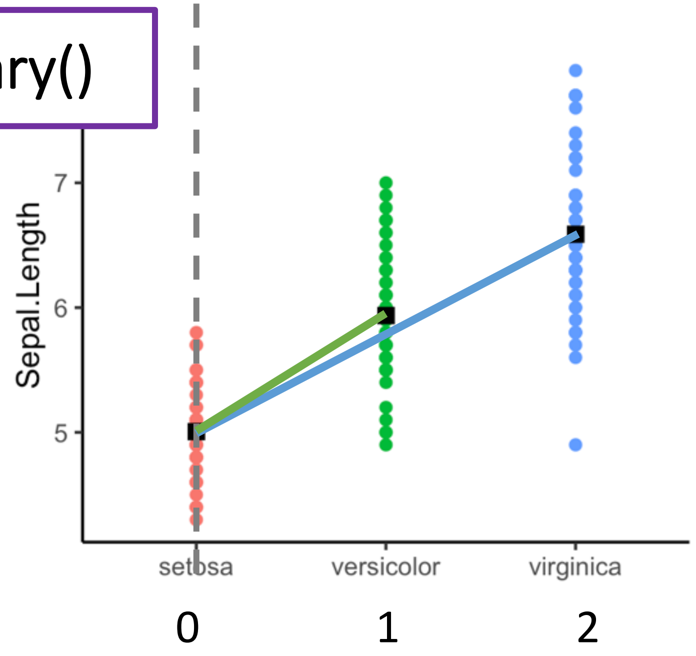
```
> co
Anov

Response: Sepal.Length
             Sum Sq  Df F value    Pr(>F)
(Intercept) 1253.00   1 4728.16 < 2.2e-16 ***
Species       63.21   2  119.26 < 2.2e-16 ***
Residuals     38.96 147
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

summary()



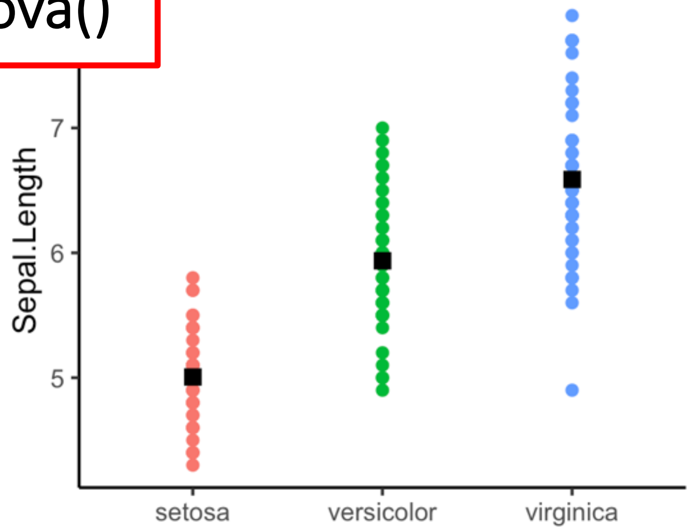

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
```

summary()



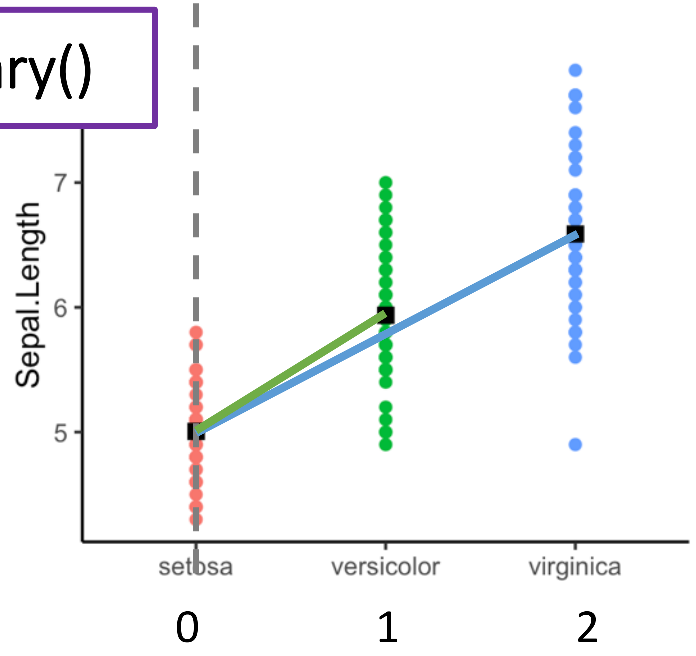

```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
```

car::Anova()



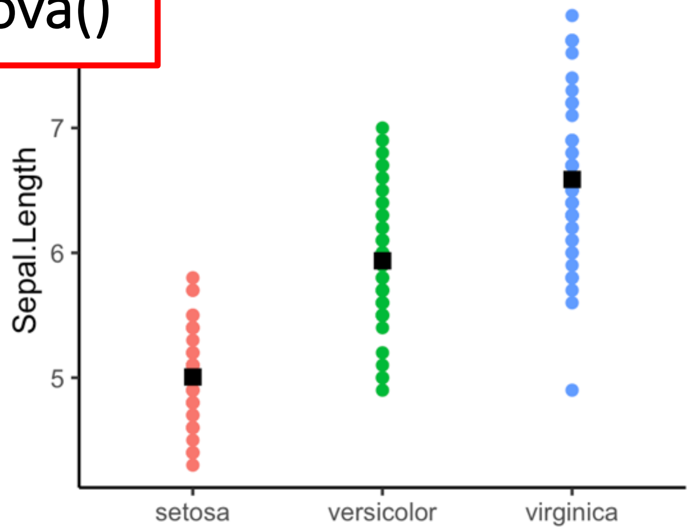

```
Response: Sepal.Length
             Sum Sq  Df F value    Pr(>F)
(Intercept) 1253.00   1 4728.16 < 2.2e-16 ***
Species       63.21   2  119.26 < 2.2e-16 ***
Residuals     38.96 147
```

summary()




```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
```

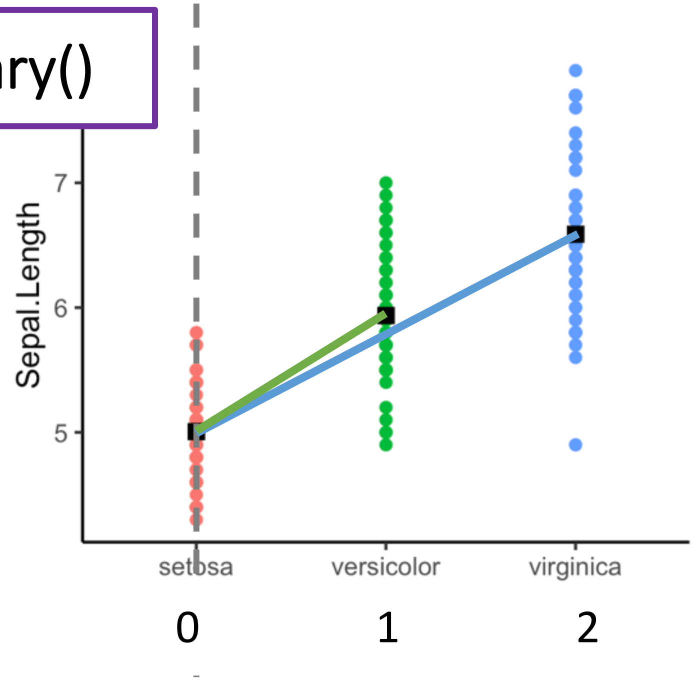
car::Anova()




```
Response: Sepal.Length
             Sum Sq  Df F value    Pr(>F)
(Intercept) 1253.00   1 4728.16 < 2.2e-16 ***
Species       63.21   2  119.26 < 2.2e-16 ***
Residuals     38.96 147
```

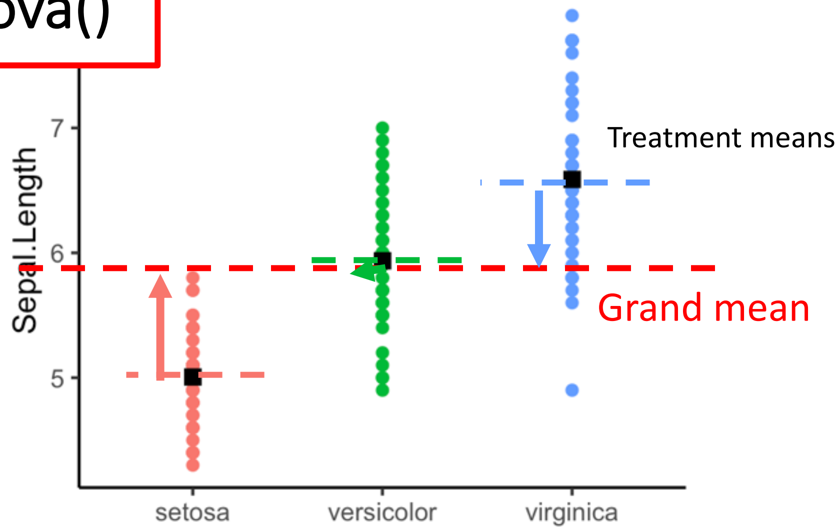
$$\frac{\text{variance between groups}}{\text{variance within group}} = \text{F-ratio}$$

summary()




```
Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         5.0060     0.0728  68.762  < 2e-16 ***
Speciesversicolor   0.9300     0.1030   9.033 8.77e-16 ***
Speciesvirginica    1.5820     0.1030  15.366  < 2e-16 ***
```

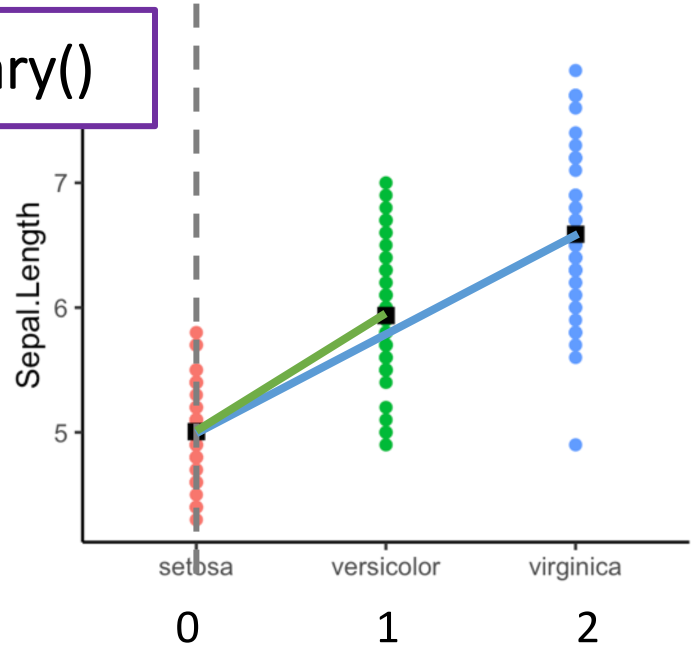
car::Anova()




```
Response: Sepal.Length
             Sum Sq  Df F value    Pr(>F)
(Intercept) 1253.00   1 4728.16 < 2.2e-16 ***
Species       63.21   2  119.26 < 2.2e-16 ***
Residuals     38.96 147
```

$$\frac{\text{variance between groups}}{\text{variance within group}} = \text{F-ratio}$$

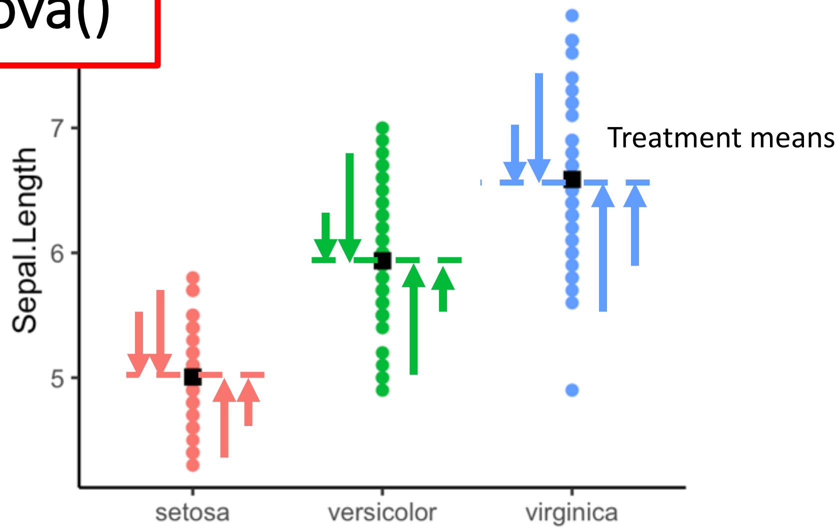
summary()



Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	5.0060	0.0728	68.762	< 2e-16	***
Speciesversicolor	0.9300	0.1030	9.033	8.77e-16	***
Speciesvirginica	1.5820	0.1030	15.366	< 2e-16	***

car::Anova()

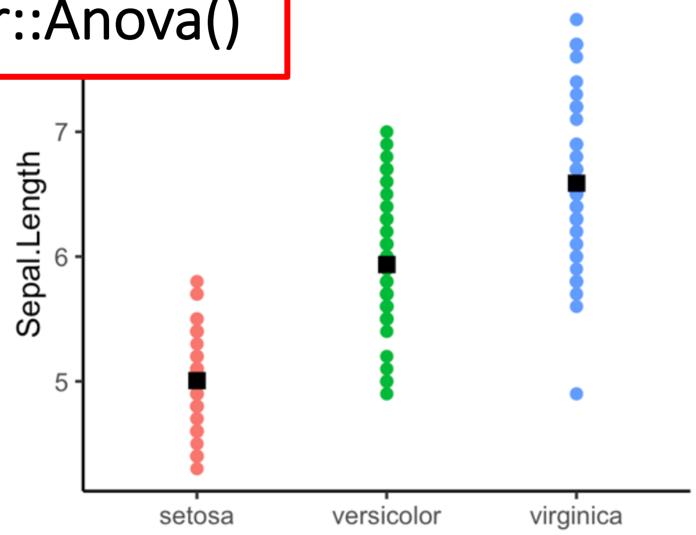


Response: Sepal.Length

	Sum Sq	Df	F value	Pr(>F)	
(Intercept)	1253.00	1	4728.16	< 2.2e-16	***
Species	63.21	2	119.26	< 2.2e-16	***
Residuals	38.96	147			

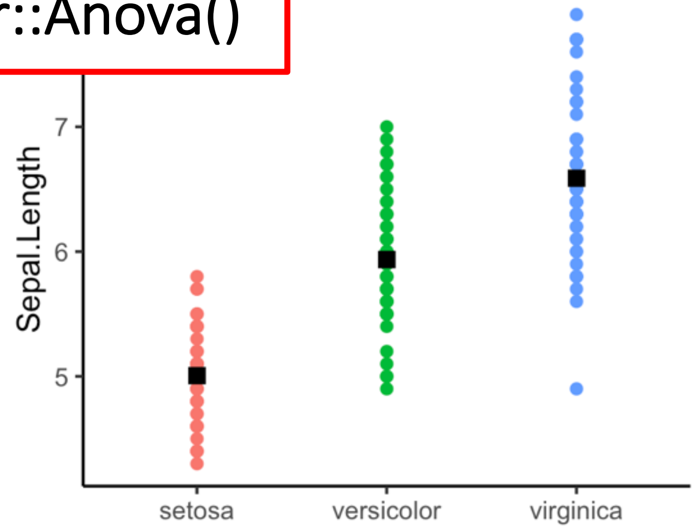
$$\frac{\text{variance between groups}}{\text{variance within group}} = \text{F-ratio}$$

car::Anova()




```
Response: Sepal.Length
            Sum Sq  Df F value    Pr(>F)    
(Intercept) 1253.00   1 4728.16 < 2.2e-16 ***
Species       63.21   2  119.26 < 2.2e-16 ***
Residuals     38.96 147
```

car::Anova()

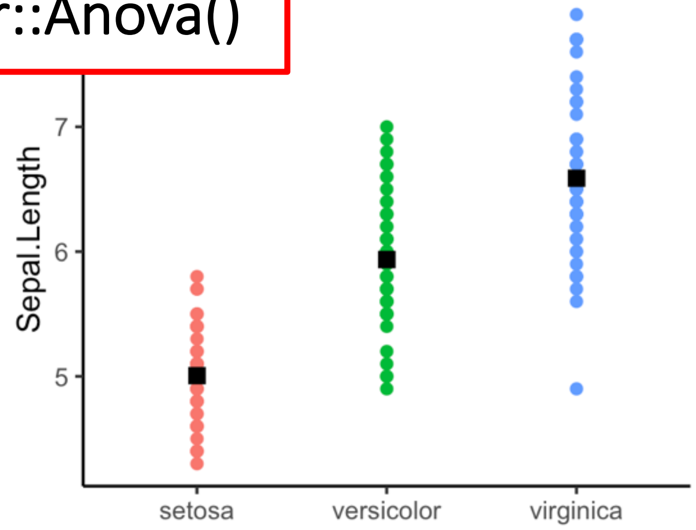


Response: Sepal.Length

	Sum Sq	Df	F value	Pr(>F)	
(Intercept)	1253.00	1	4728.16	< 2.2e-16	***
Species	63.21	2	119.26	< 2.2e-16	***
Residuals	38.96	147			

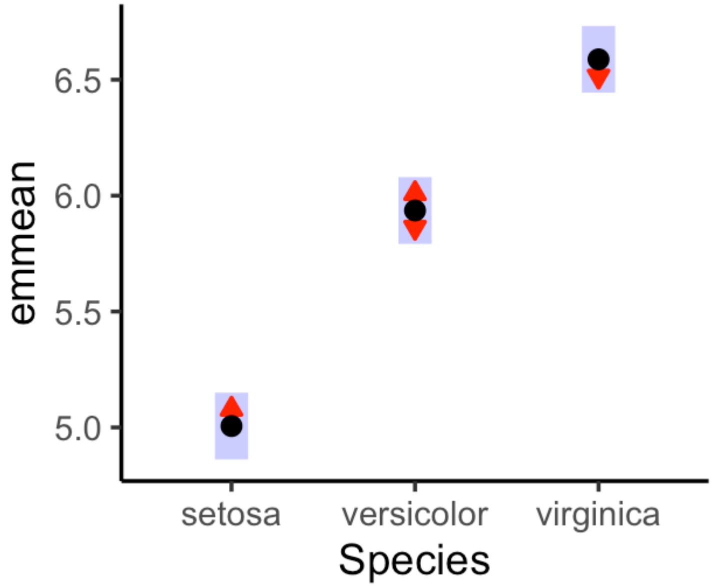
variance between groups was **sufficiently huge** while accounting for **individual variance within group**

car::Anova()



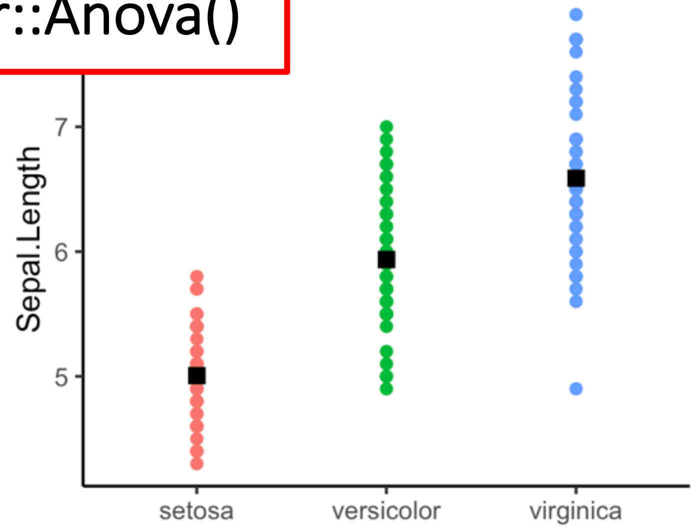

```
Response: Sepal.Length
             Sum Sq  Df F value    Pr(>F)
(Intercept) 1253.00   1 4728.16 < 2.2e-16 ***
Species       63.21   2  119.26 < 2.2e-16 ***
Residuals     38.96 147
```

emmeans()



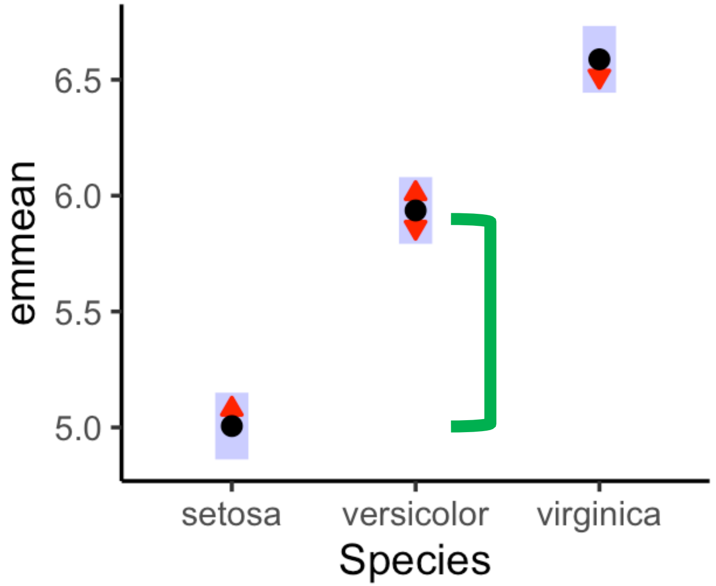

```
$contrasts
 contrast               estimate    SE  df t.ratio p.value
 setosa - versicolor      -0.930 0.103 147  -9.033  <.0001
 setosa - virginica       -1.582 0.103 147 -15.366  <.0001
 versicolor - virginica   -0.652 0.103 147  -6.333  <.0001
```

car::Anova()



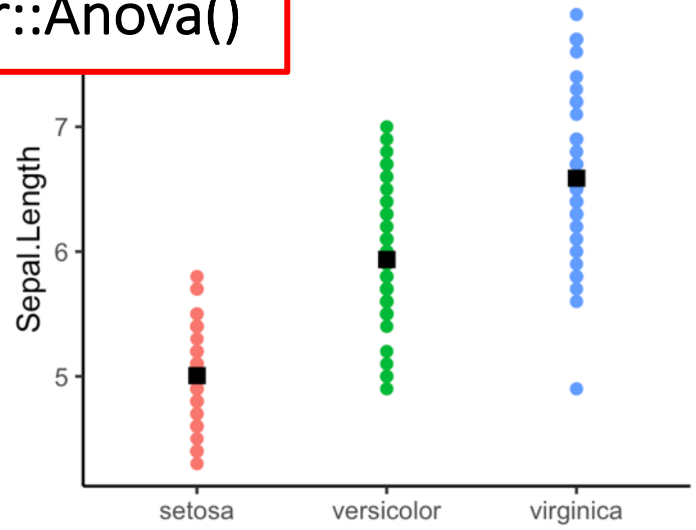

```
Response: Sepal.Length
             Sum Sq  Df F value    Pr(>F)
(Intercept) 1253.00   1 4728.16 < 2.2e-16 ***
Species       63.21   2  119.26 < 2.2e-16 ***
Residuals     38.96 147
```

emmeans()



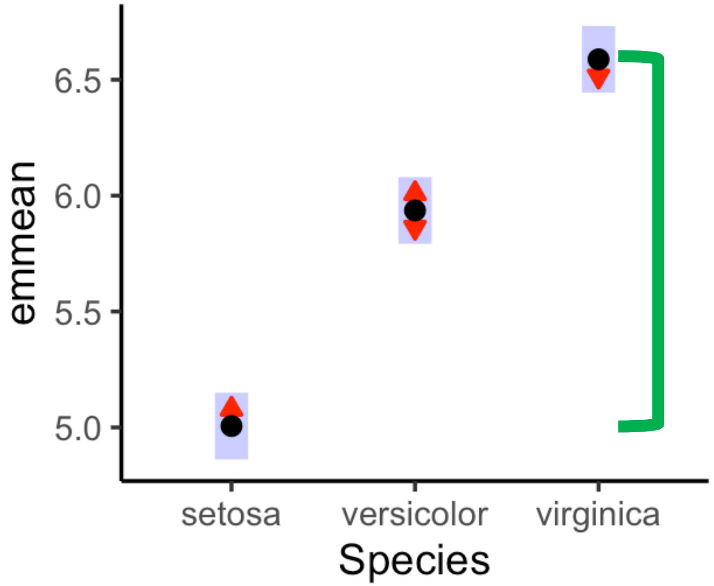

```
$contrasts
 contrast               estimate    SE  df t.ratio p.value
 setosa - versicolor      -0.930 0.103 147  -9.033  <.0001
 setosa - virginica       -1.582 0.103 147 -15.366  <.0001
 versicolor - virginica   -0.652 0.103 147  -6.333  <.0001
```

car::Anova()



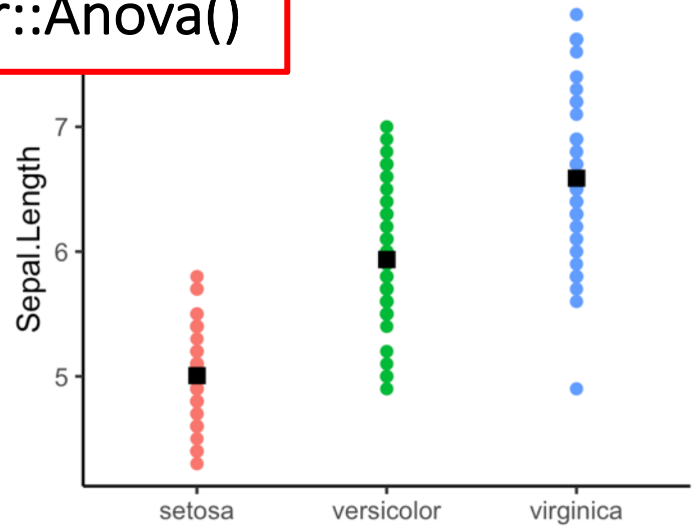

```
Response: Sepal.Length
             Sum Sq  Df F value    Pr(>F)
(Intercept) 1253.00   1 4728.16 < 2.2e-16 ***
Species       63.21   2  119.26 < 2.2e-16 ***
Residuals     38.96 147
```

emmeans()



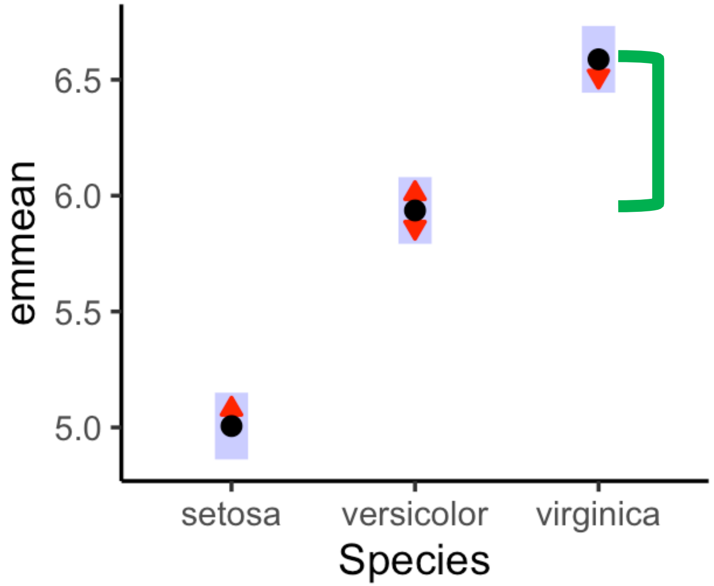

```
$contrasts
 contrast               estimate    SE  df t.ratio p.value
 setosa - versicolor      -0.930 0.103 147  -9.033  <.0001
 setosa - virginica       -1.582 0.103 147 -15.366  <.0001
 versicolor - virginica   -0.652 0.103 147  -6.333  <.0001
```

car::Anova()



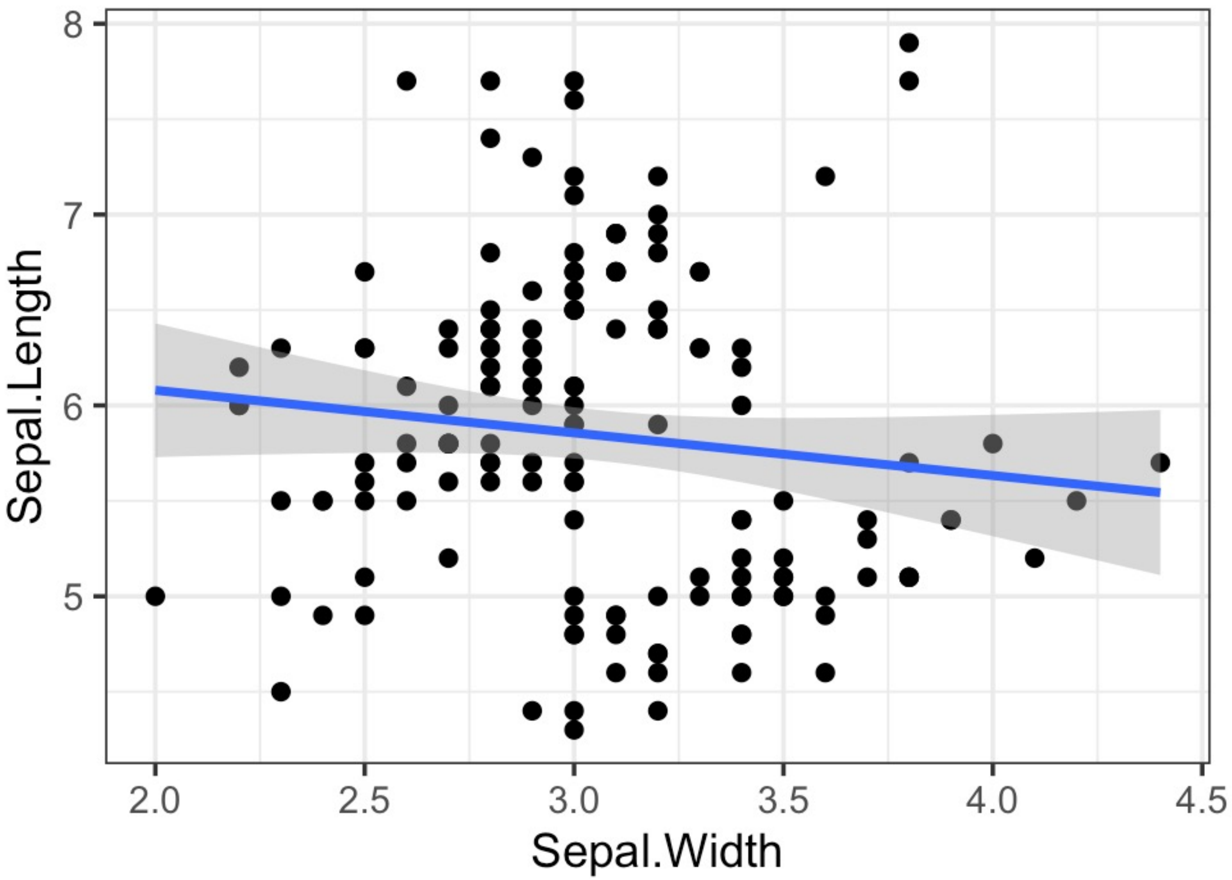

```
Response: Sepal.Length
             Sum Sq  Df F value    Pr(>F)
(Intercept) 1253.00   1 4728.16 < 2.2e-16 ***
Species       63.21   2  119.26 < 2.2e-16 ***
Residuals     38.96 147
```

emmeans()

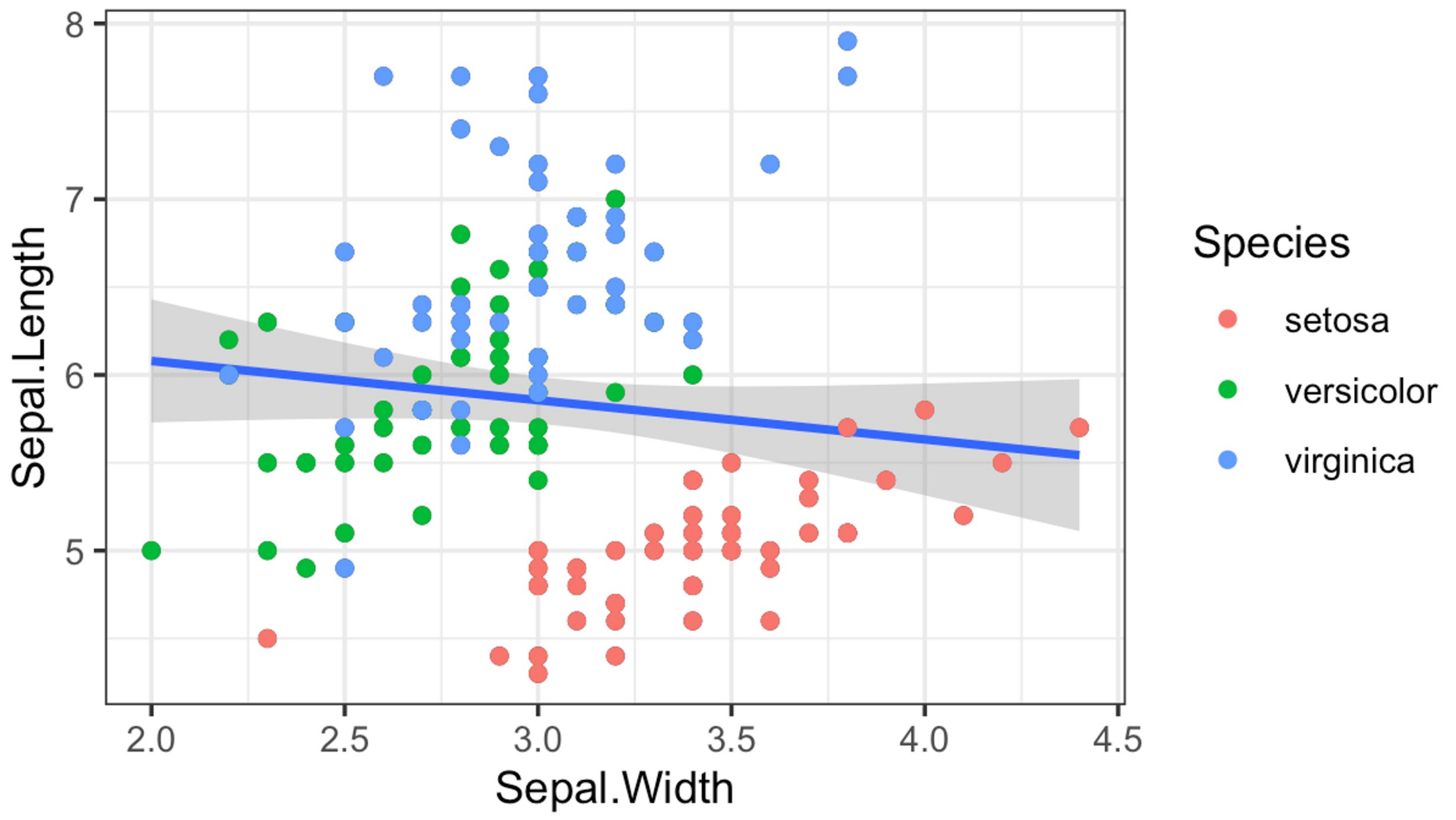



```
$contrasts
 contrast               estimate    SE  df t.ratio p.value
 setosa - versicolor      -0.930 0.103 147  -9.033  <.0001
 setosa - virginica       -1.582 0.103 147 -15.366  <.0001
 versicolor - virginica   -0.652 0.103 147  -6.333  <.0001
```

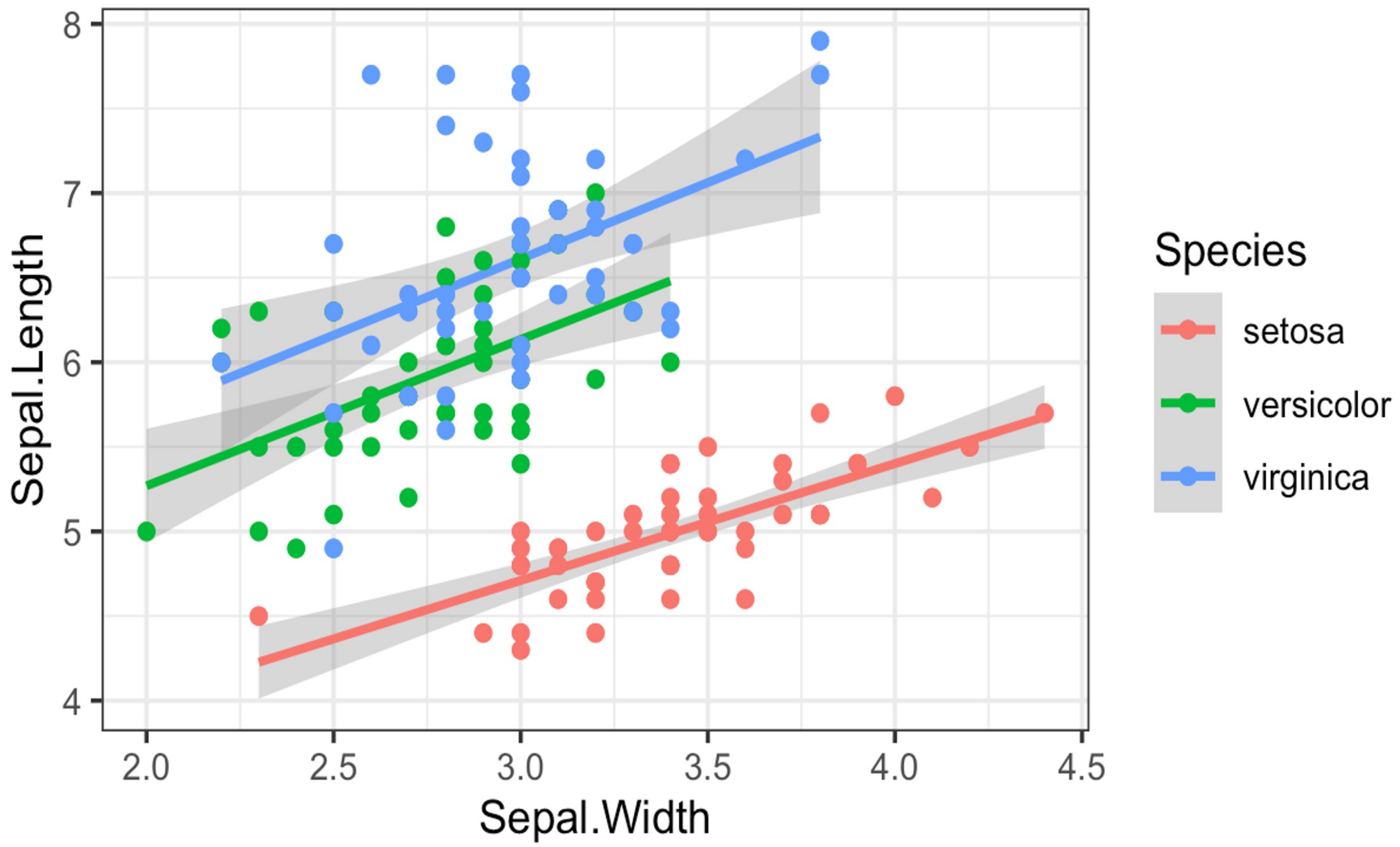
lm(Sepal.Length ~ Sepal.Width, data = iris)



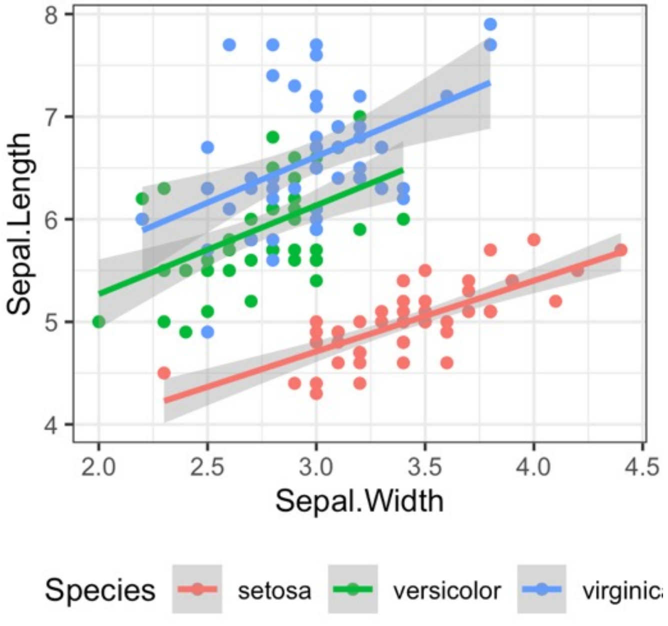
lm(Sepal.Length ~ Sepal.Width, data = iris)



lm(Sepal.Length ~ Sepal.Width **+ Species**, data = iris)

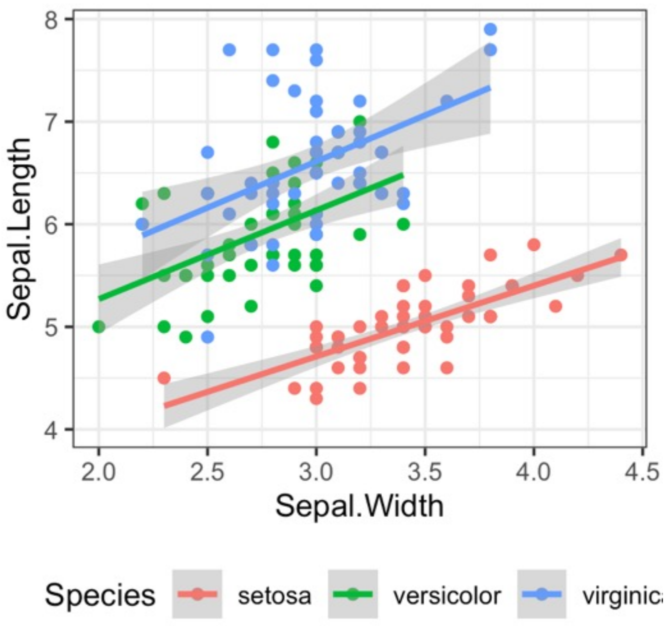


lm(Sepal.Length ~ Sepal.Width + Species, data = iris)




```
Call:
lm(formula = Sepal.Length ~ Sepal.Width + Species, data = iris)

Residuals:
     Min       1Q   Median       3Q      Max 
-1.30711 -0.25713 -0.05325  0.19542  1.41253 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)         2.2514     0.3698   6.089 9.57e-09 ***
Sepal.Width         0.8036     0.1063   7.557 4.19e-12 ***
Speciesversicolor   1.4587     0.1121  13.012  < 2e-16 ***
Speciesvirginica    1.9468     0.1000  19.465  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.438 on 146 degrees of freedom
Multiple R-squared:  0.7259,	Adjusted R-squared:  0.7203 
F-statistic: 128.9 on 3 and 146 DF,  p-value: < 2.2e-16
```

lm(Sepal.Length ~ Sepal.Width + Species, data = iris)




```
Anova Table (Type III tests)

Response: Sepal.Length
             Sum Sq  Df F value    Pr(>F)
(Intercept)  7.111   1  37.075 9.568e-09 ***
Sepal.Width 10.953   1  57.102 4.187e-12 ***
Species     72.752   2 189.651 < 2.2e-16 ***
Residuals   28.004 146
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

interactions in linear model

vs

interactions in linear model

type of supplement

vs

interactions in linear model

dosages

 /  / 

vs

```
lmint <- lm(len ~ supp * dose, data = ToothGrowth)
```

vs


C
VITAMIN



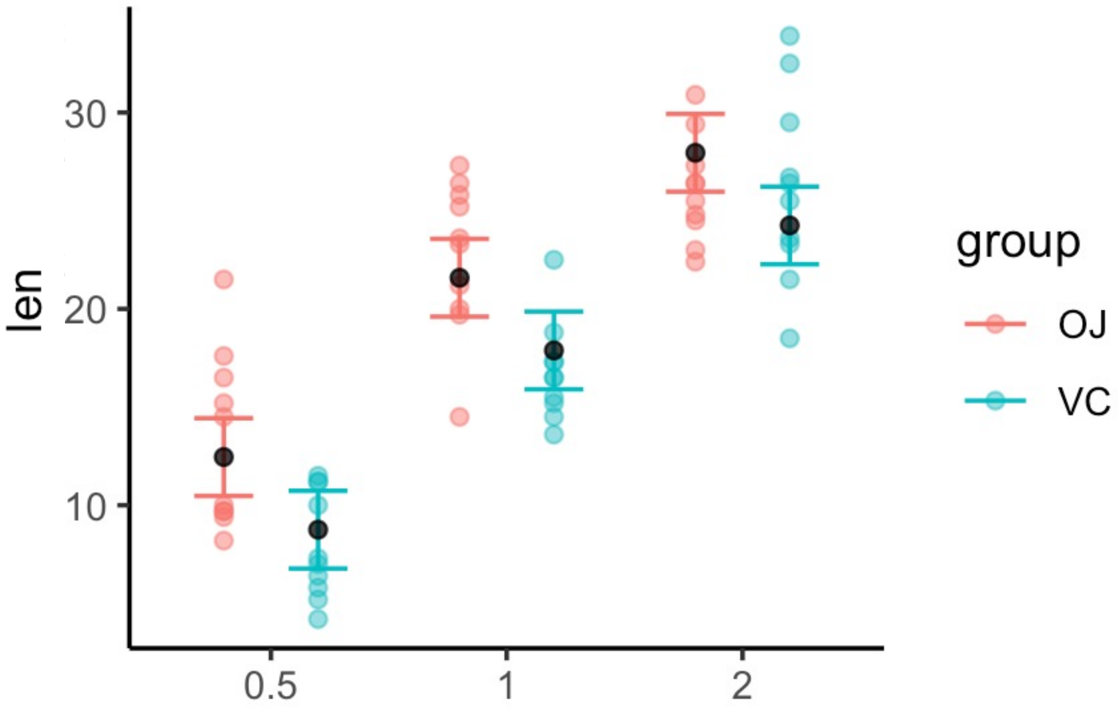




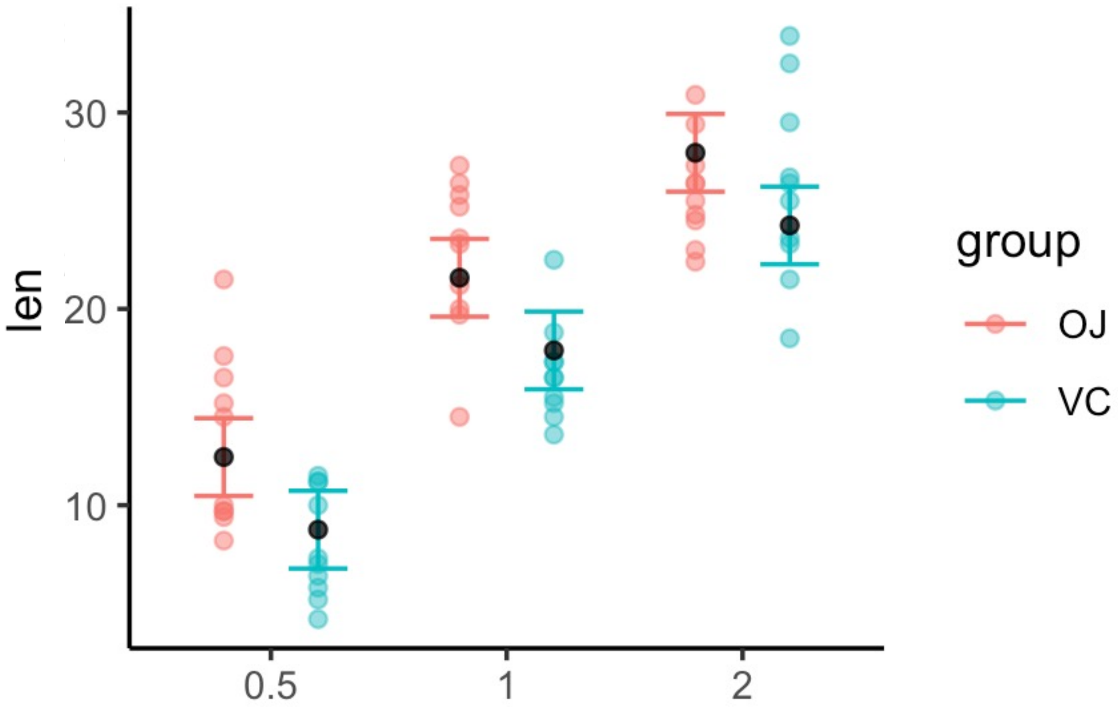





lmint <- lm(len ~ supp + dose, data = ToothGrowth)

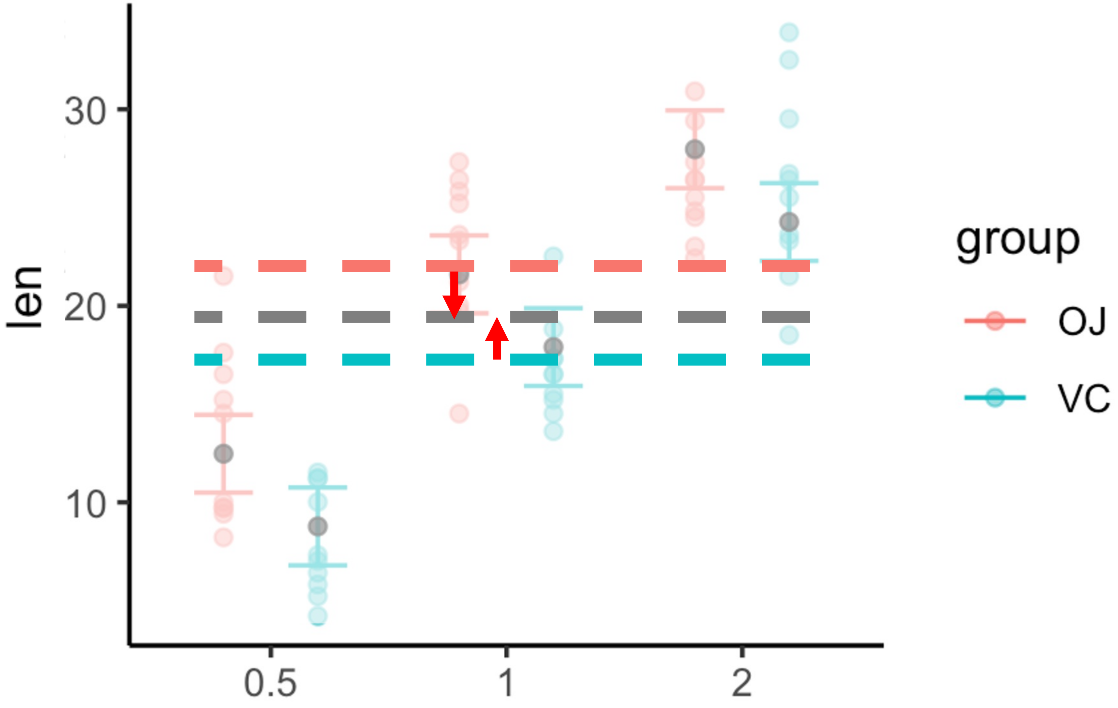


lmint <- lm(len ~ supp + dose, data = ToothGrowth)




```
Anova Table (Type III tests)

Response: len
             Sum Sq Df F value    Pr(>F)
(Intercept) 2326.91  1 158.828 < 2.2e-16 *
supp         205.35  1  14.017 0.0004293 *
dose        2426.43  2  82.811 < 2.2e-16 *
Residuals    820.43 56
```

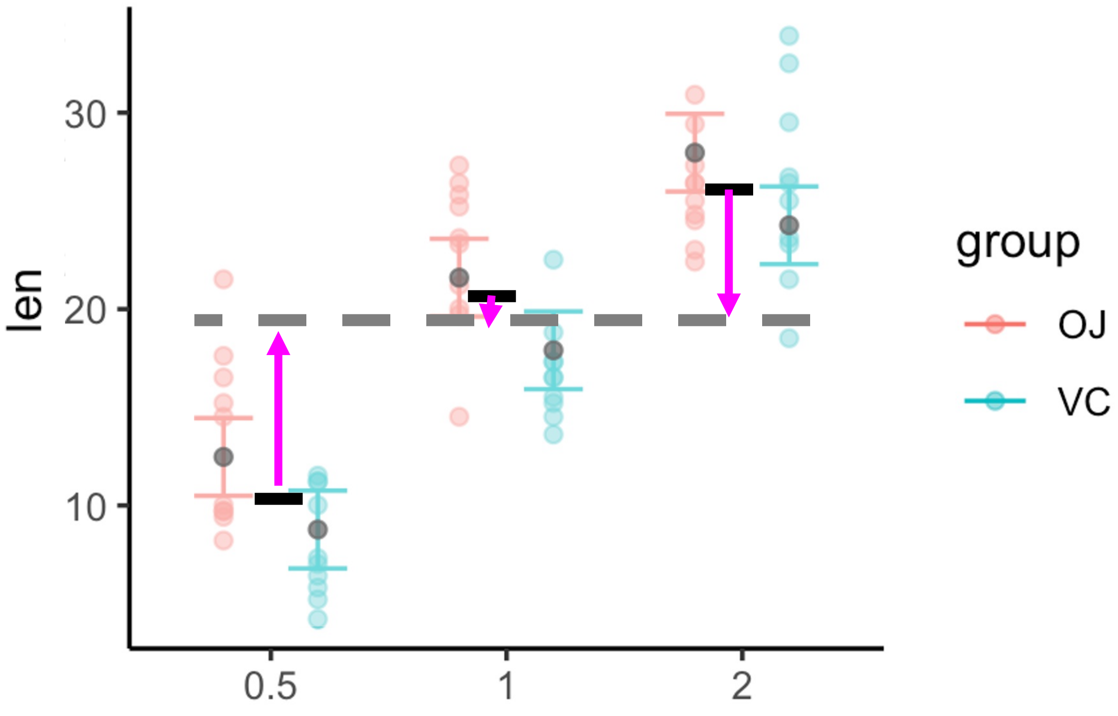
lmint <- lm(len ~ supp + dose, data = ToothGrowth)




```
Anova Table (Type III tests)

Response: len
             Sum Sq Df F value    Pr(>F)
(Intercept) 2326.91  1 158.828 < 2.2e-16 *
supp         205.35  1  14.017 0.0004293 *
dose        2426.43  2  82.811 < 2.2e-16 *
Residuals    820.43 56
```

Redo: Suppose to compare treatment mean to grand mean

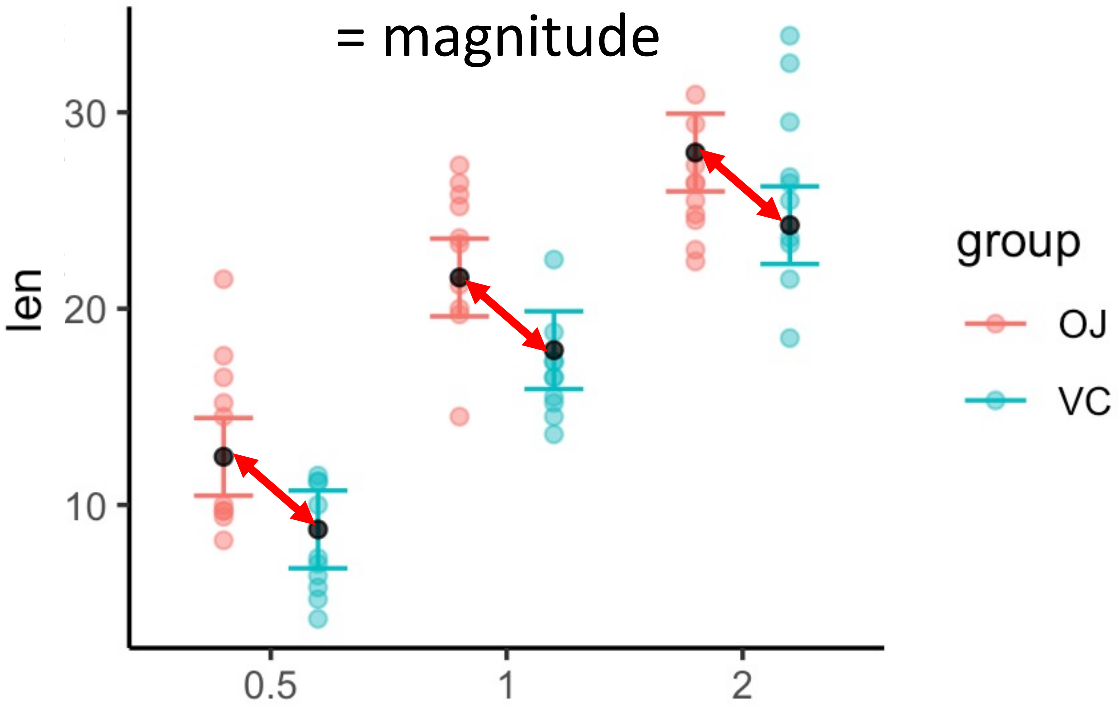
lmint <- lm(len ~ supp + dose, data = ToothGrowth)




```
Anova Table (Type III tests)

Response: len
             Sum Sq Df F value    Pr(>F)
(Intercept) 2326.91  1 158.828 < 2.2e-16 *
supp         205.35  1  14.017 0.0004293 *
dose        2426.43  2  82.811 < 2.2e-16 *
Residuals    820.43 56
```

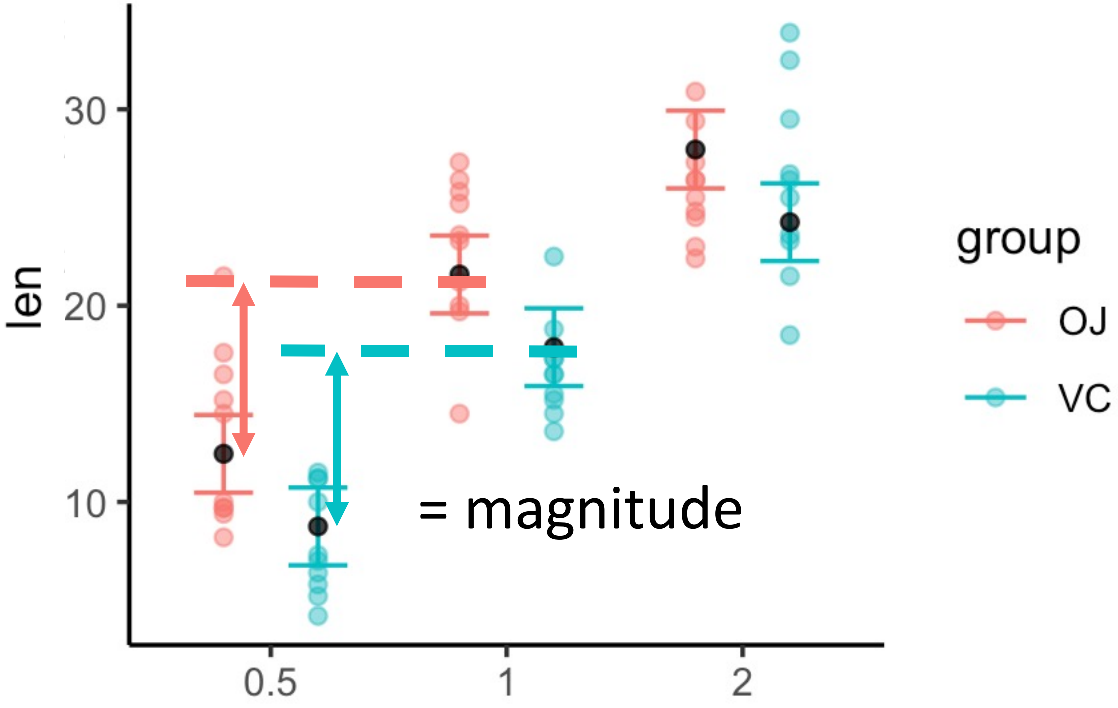
Redo: Suppose to compare treatment mean to grand mean

lmint <- lm(len ~ supp + dose, data = ToothGrowth)



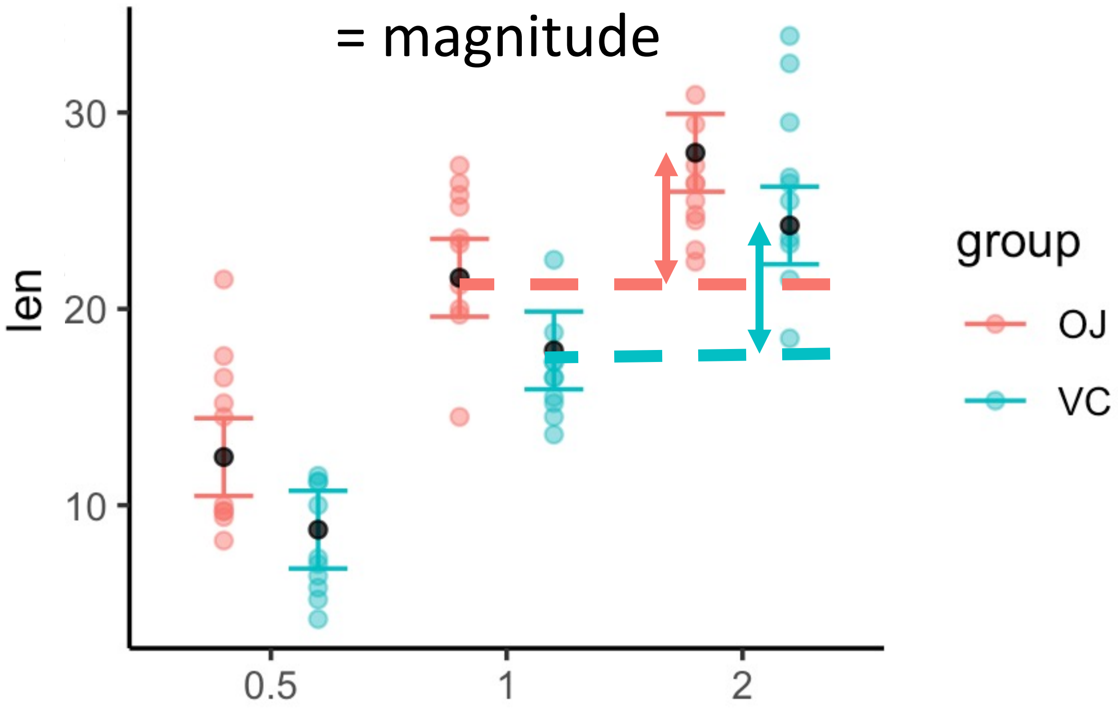

```
Anova Table (Type III tests)

Response: len
            Sum Sq Df F value    Pr(>F)
(Intercept) 2326.91  1 158.828 < 2.2e-16 *
supp         205.35  1  14.017 0.0004293 *
dose        2426.43  2  82.811 < 2.2e-16 *
Residuals    820.43 56
```

lmint <- lm(len ~ supp + dose, data = ToothGrowth)



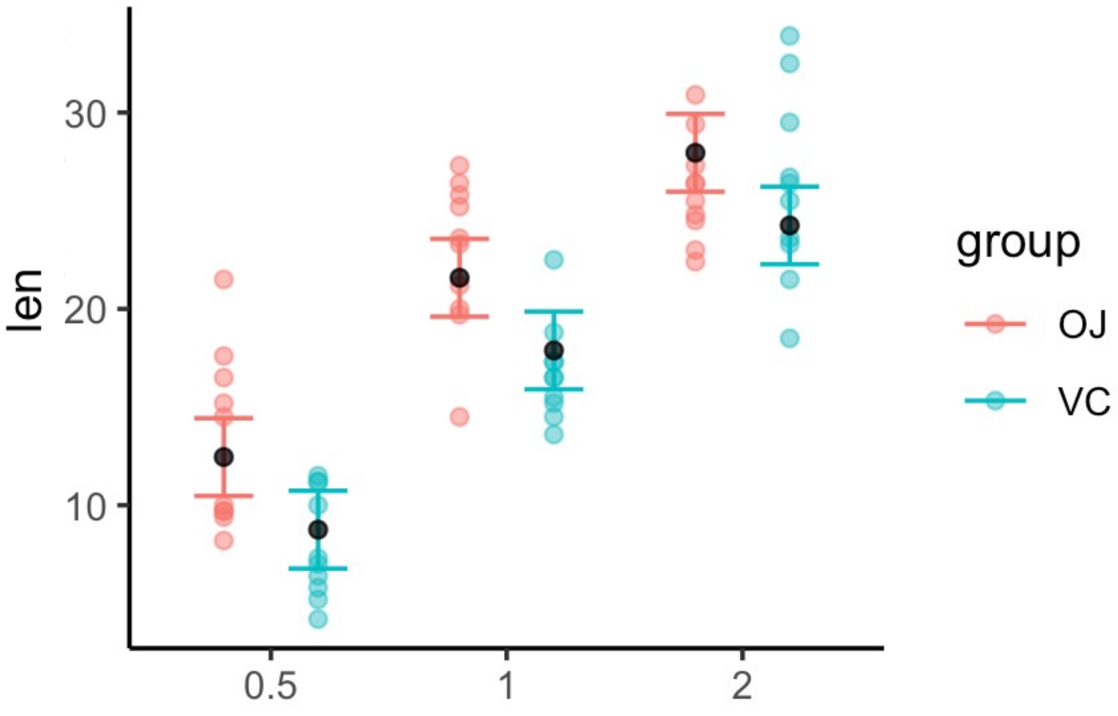

```
Anova Table (Type III tests)

Response: len
             Sum Sq Df F value    Pr(>F)
(Intercept) 2326.91  1 158.828 < 2.2e-16 *
supp         205.35  1  14.017 0.0004293 *
dose        2426.43  2  82.811 < 2.2e-16 *
Residuals    820.43 56
```

lmint <- lm(len ~ supp + dose, data = ToothGrowth)



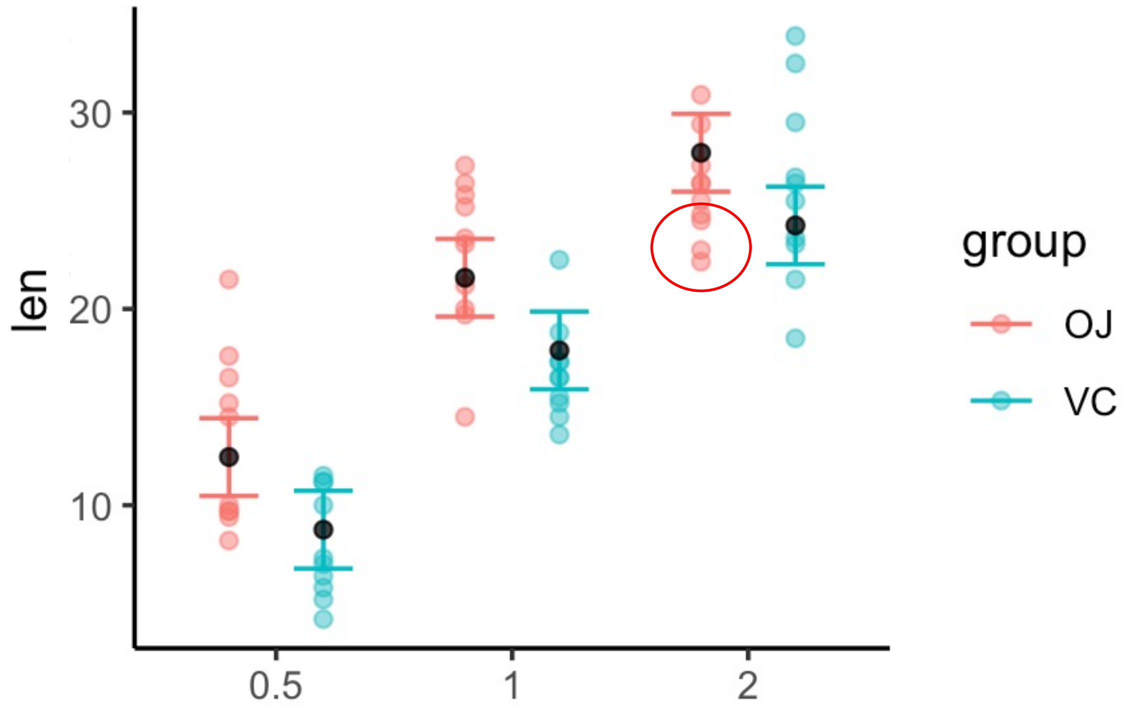

```
Anova Table (Type III tests)

Response: len
             Sum Sq Df F value    Pr(>F)
(Intercept) 2326.91  1 158.828 < 2.2e-16 *
supp         205.35  1  14.017 0.0004293 *
dose        2426.43  2  82.811 < 2.2e-16 *
Residuals    820.43 56
```

```
lmint <- lm(len ~ supp + dose, data = ToothGrowth)
```



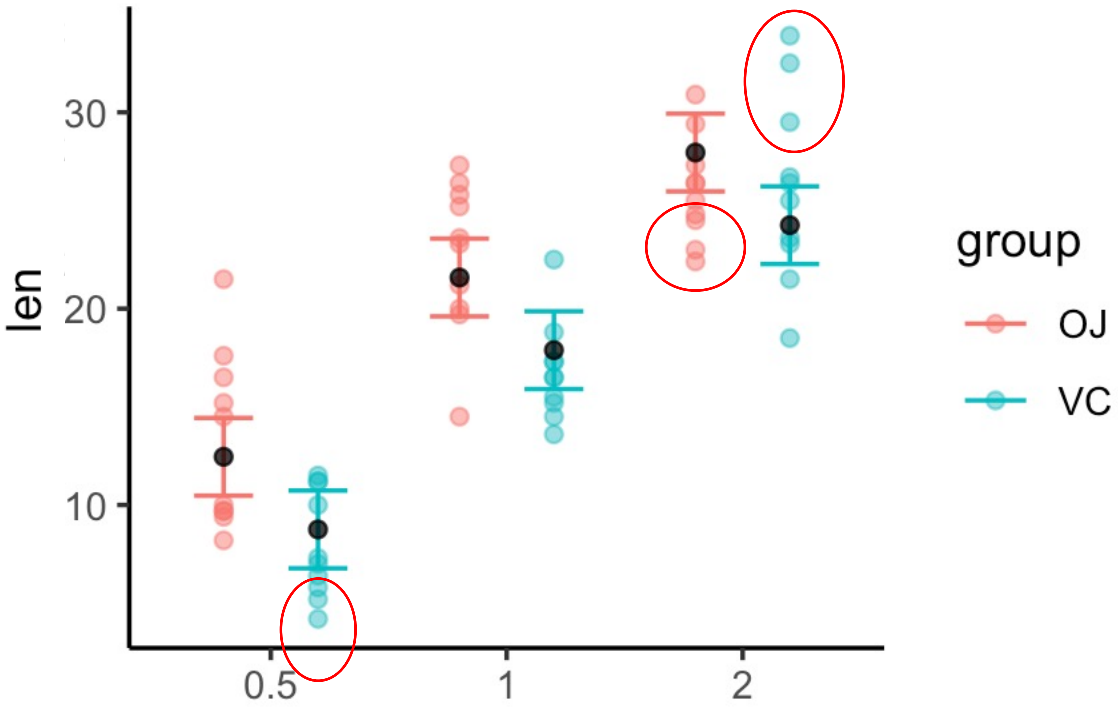
However!

lmint <- lm(len ~ supp + dose, data = ToothGrowth)



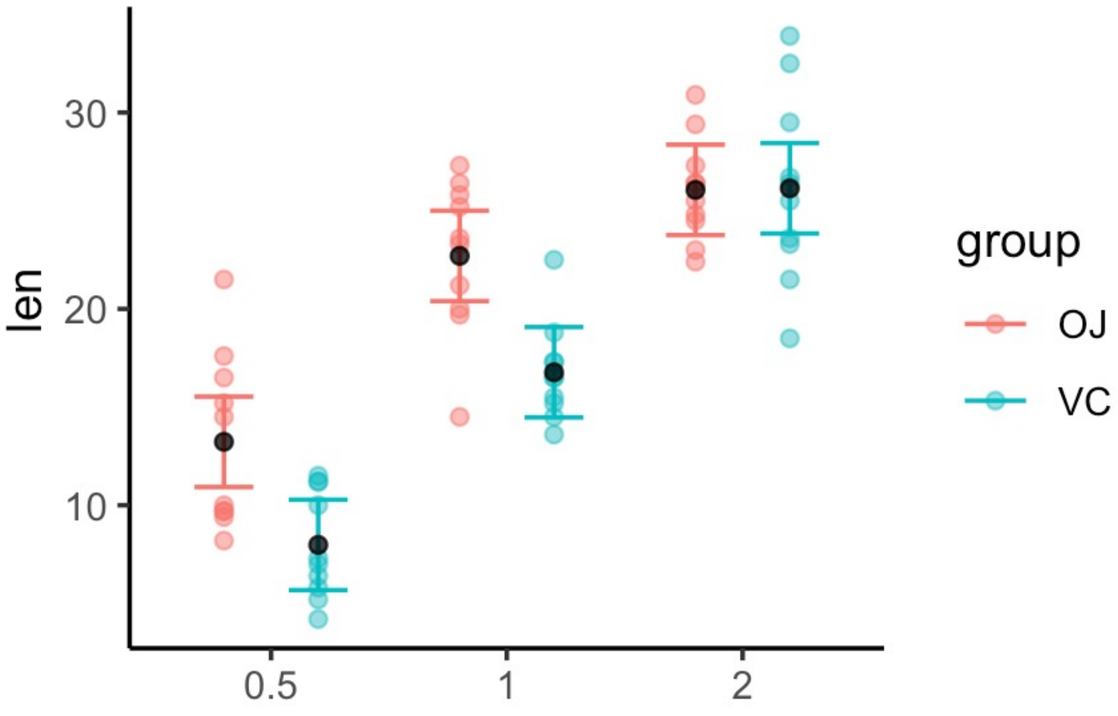
However!

lmint <- lm(len ~ supp + dose, data = ToothGrowth)

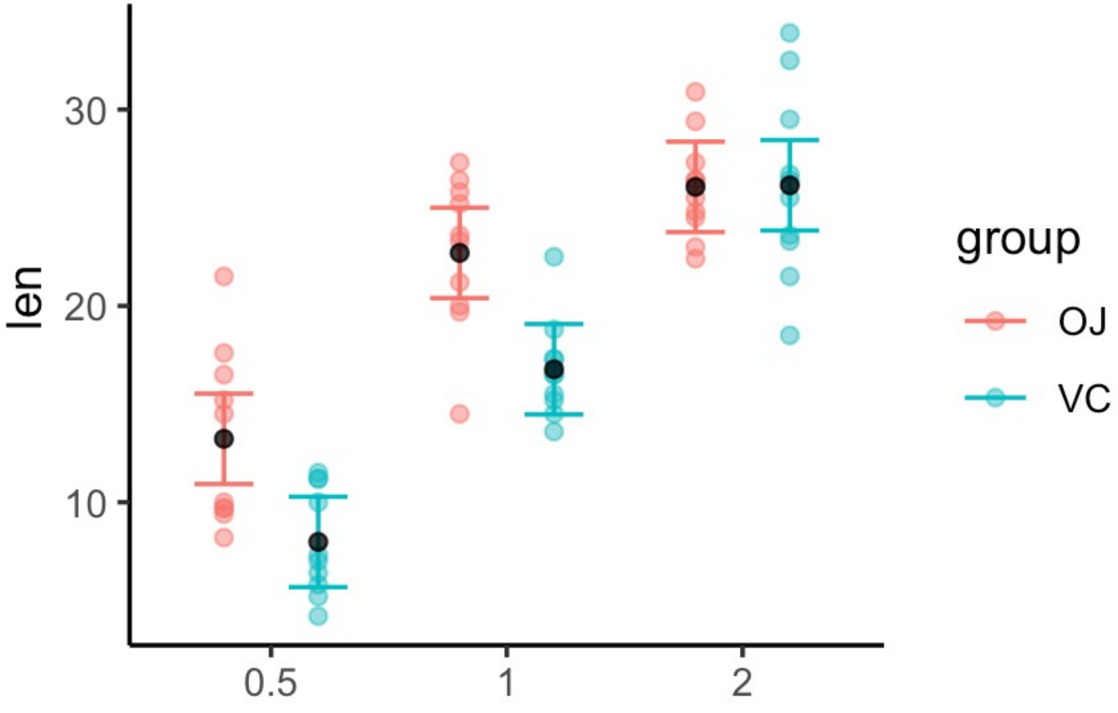


However!

```
lmint <- lm(len ~ supp * dose, data = ToothGrowth)
```

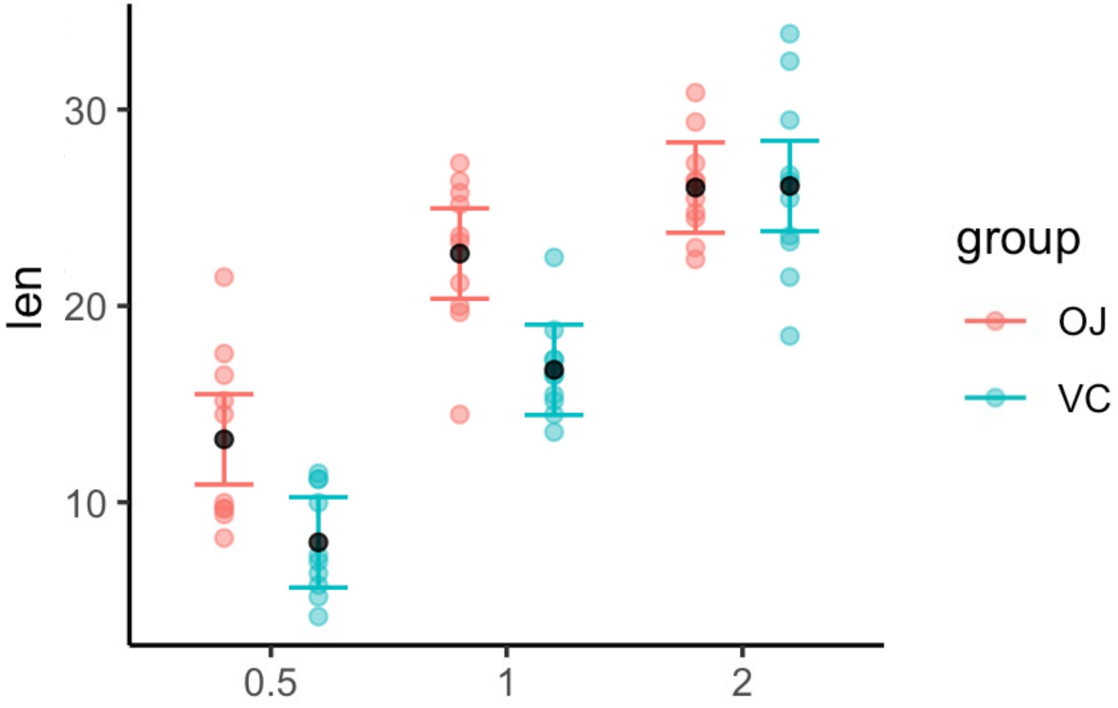


lmint <- lm(len ~ supp * dose, data = ToothGrowth)




```
Anova Table (Type III tests)

Response: len
            Sum Sq Df F value    Pr(>F)
(Intercept) 1750.33  1 132.730 3.603e-16 **
supp         137.81  1  10.450  0.002092 **
dose         885.26  2  33.565 3.363e-10 **
supp:dose    108.32  2   4.107  0.021860 *
Residuals    712.11 54
```

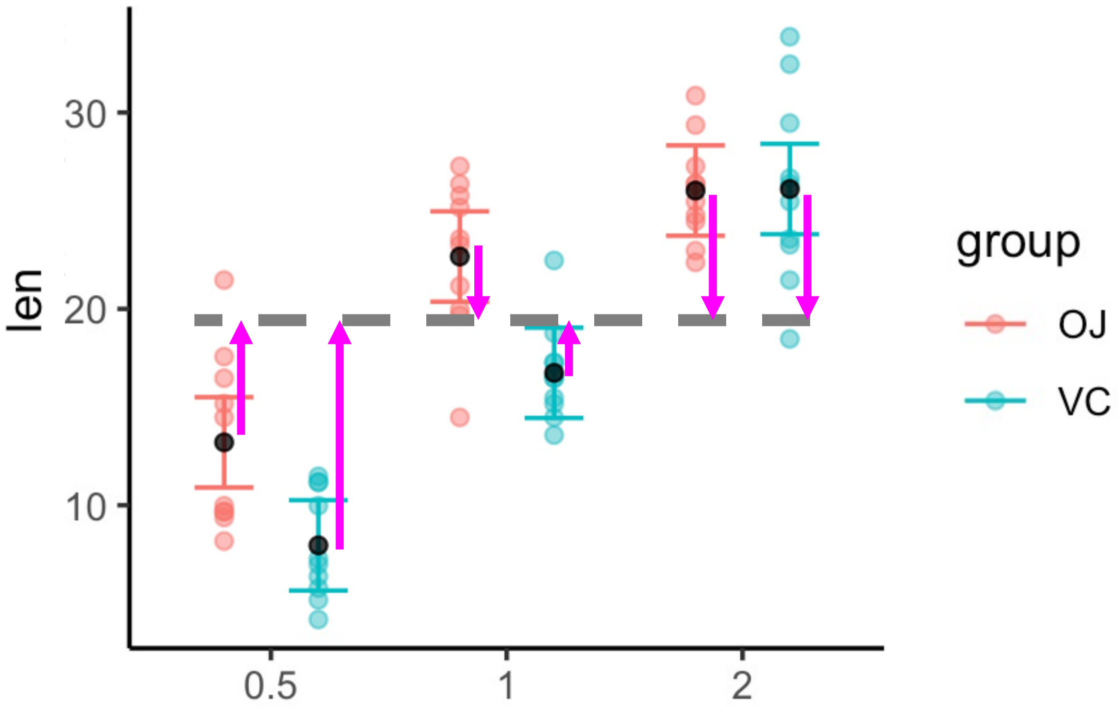
lmint <- lm(len ~ supp * dose, data = ToothGrowth)




```
Anova Table (Type III tests)

Response: len
              Sum Sq Df F value    Pr(>F)
(Intercept) 1750.33  1 132.730 3.603e-16 **
supp         137.81  1  10.450  0.002092 **
dose         885.26  2  33.565 3.363e-10 **
supp:dose    108.32  2   4.107  0.021860 *
Residuals    712.11 54
```

$$(\bar{y}_{AB} - \bar{y}_T) - (\bar{y}_A - \bar{y}_T) - (\bar{y}_B - \bar{y}_T)$$

lmint <- lm(len ~ supp * dose, data = ToothGrowth)



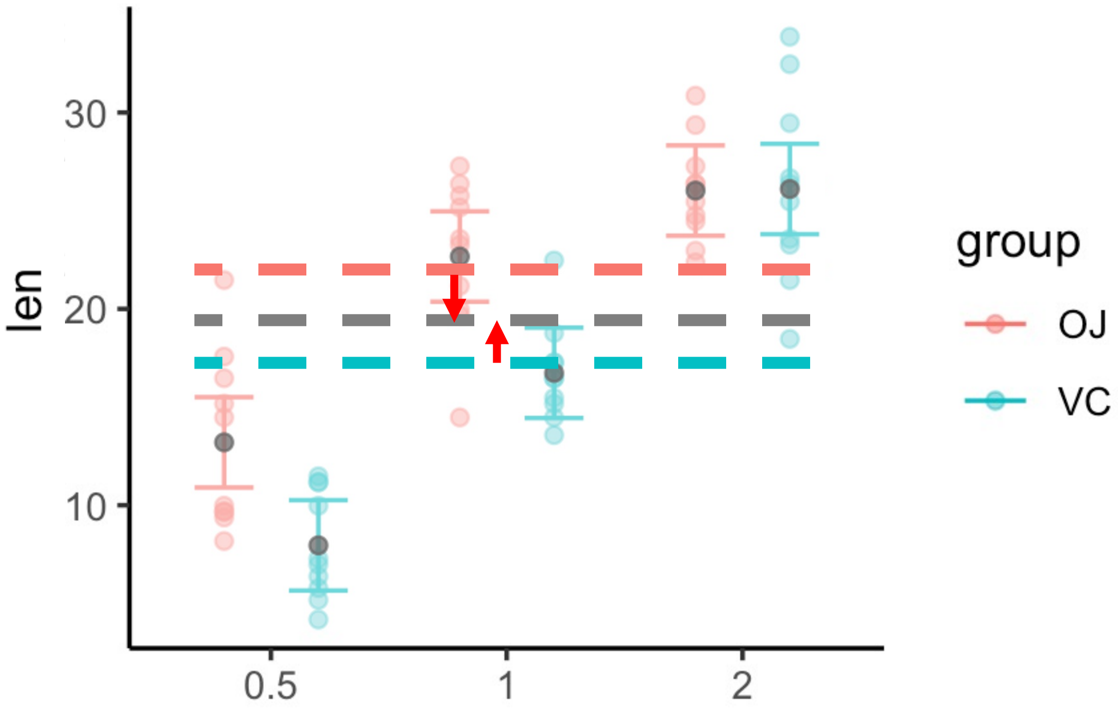
Anova Table (Type III tests)

Response: len

	Sum Sq	Df	F value	Pr(>F)	
(Intercept)	1750.33	1	132.730	3.603e-16	**
supp	137.81	1	10.450	0.002092	**
dose	885.26	2	33.565	3.363e-10	**
supp:dose	108.32	2	4.107	0.021860	*
Residuals	712.11	54			

$$(\bar{y}_{AB} - \bar{y}_T) - (\bar{y}_A - \bar{y}_T) - (\bar{y}_B - \bar{y}_T)$$

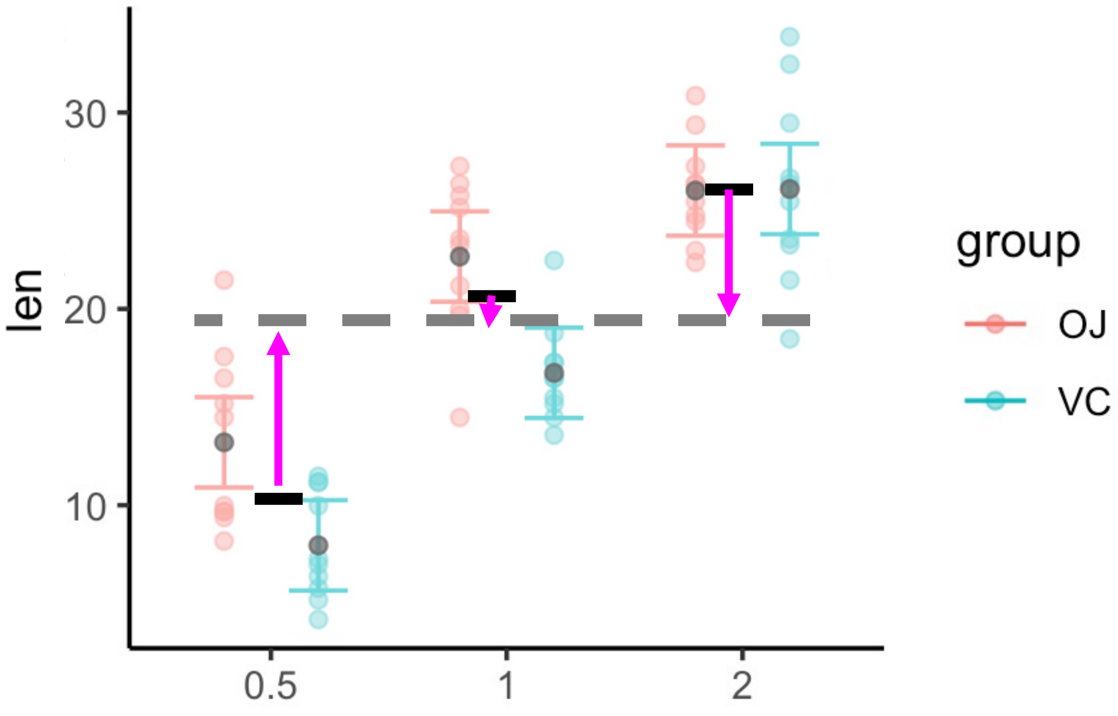
lmint <- lm(len ~ supp * dose, data = ToothGrowth)




```
Anova Table (Type III tests)

Response: len
            Sum Sq Df F value    Pr(>F)
(Intercept) 1750.33  1 132.730 3.603e-16 **
supp         137.81  1  10.450  0.002092 *
dose         885.26  2  33.565 3.363e-10 *
supp:dose    108.32  2   4.107  0.021860 *
Residuals    712.11 54
```

$$(\bar{y}_{AB} - \bar{y}_T) - (\bar{y}_A - \bar{y}_T) - (\bar{y}_B - \bar{y}_T)$$

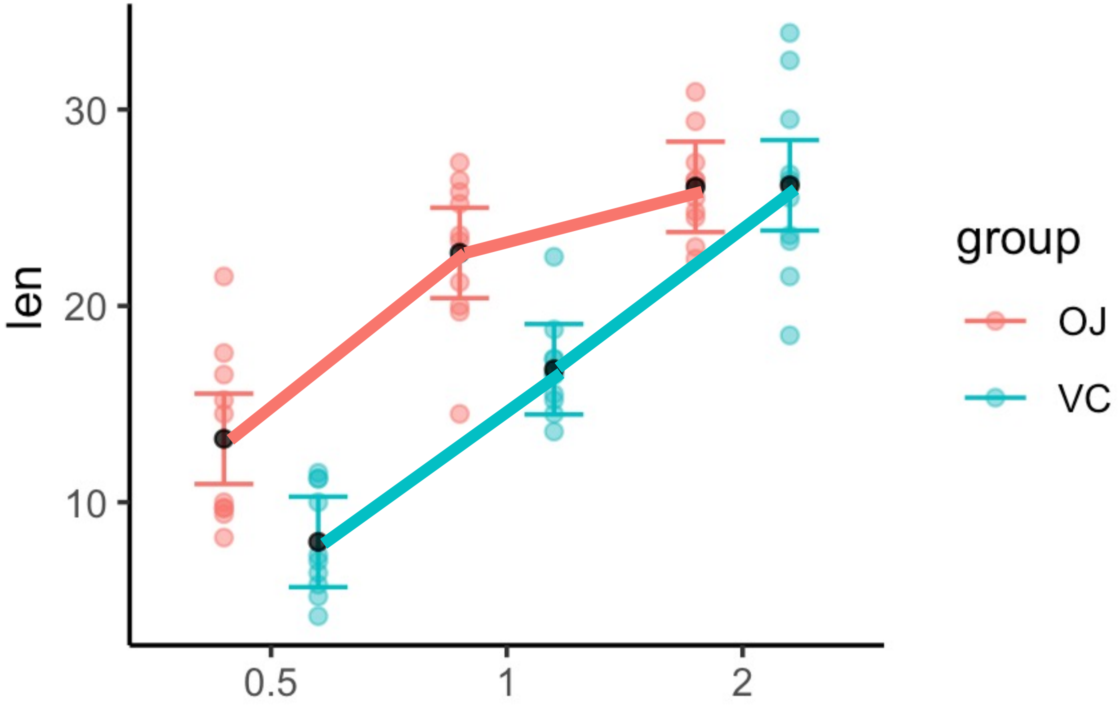
lmint <- lm(len ~ supp * dose, data = ToothGrowth)




```
Anova Table (Type III tests)

Response: len
             Sum Sq Df F value    Pr(>F)
(Intercept) 1750.33  1 132.730 3.603e-16 **
supp         137.81  1  10.450  0.002092 **
dose         885.26  2  33.565 3.363e-10 **
supp:dose    108.32  2   4.107  0.021860 *
Residuals    712.11 54
```

$$(\bar{y}_{AB} - \bar{y}_T) - (\bar{y}_A - \bar{y}_T) - (\bar{y}_B - \bar{y}_T)$$

lmint <- lm(len ~ supp * dose, data = ToothGrowth)

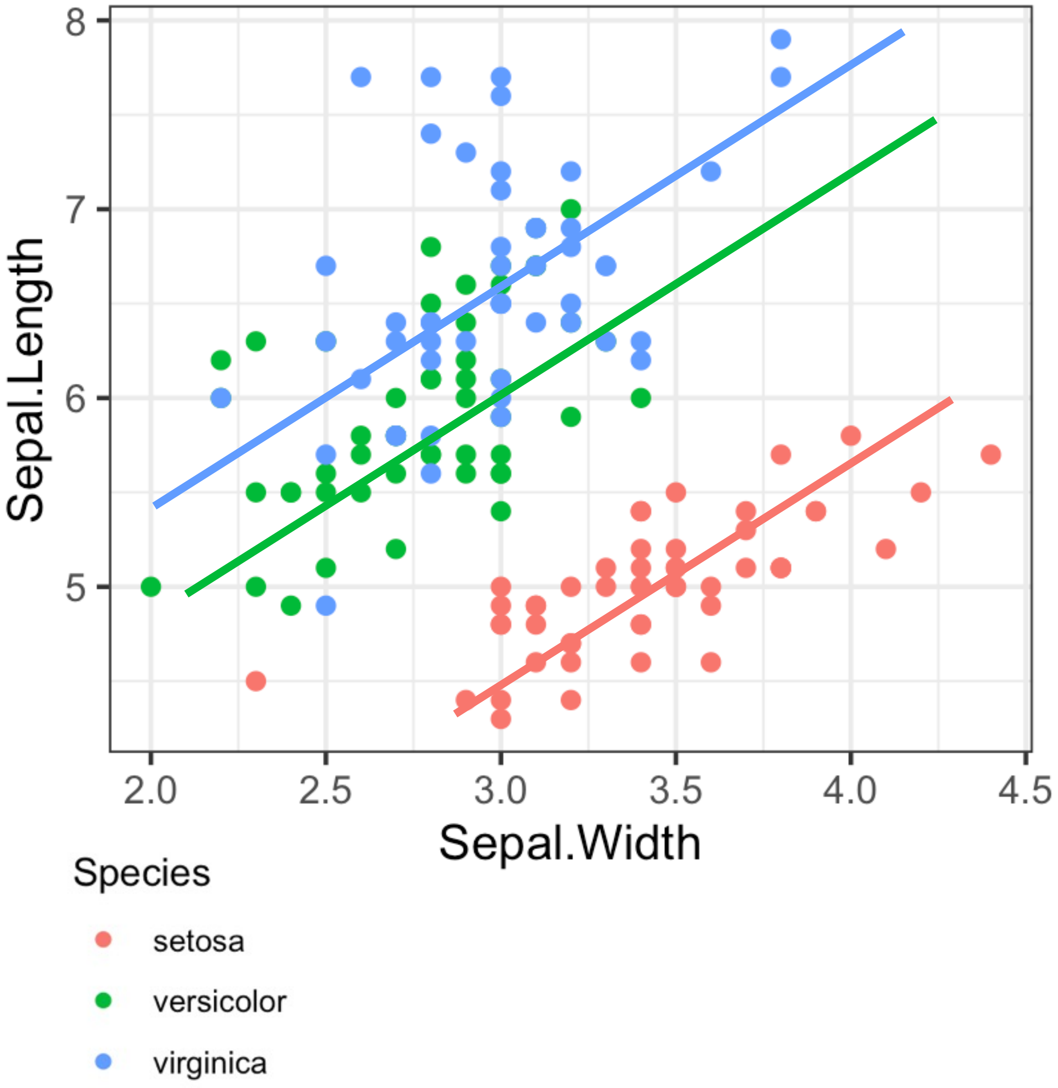



```
Anova Table (Type III tests)

Response: len
             Sum Sq Df F value    Pr(>F)
(Intercept) 1750.33  1 132.730 3.603e-16 **
supp         137.81  1  10.450  0.002092 **
dose         885.26  2  33.565 3.363e-10 **
supp:dose    108.32  2   4.107  0.021860 *
Residuals    712.11 54
```

type of interactions

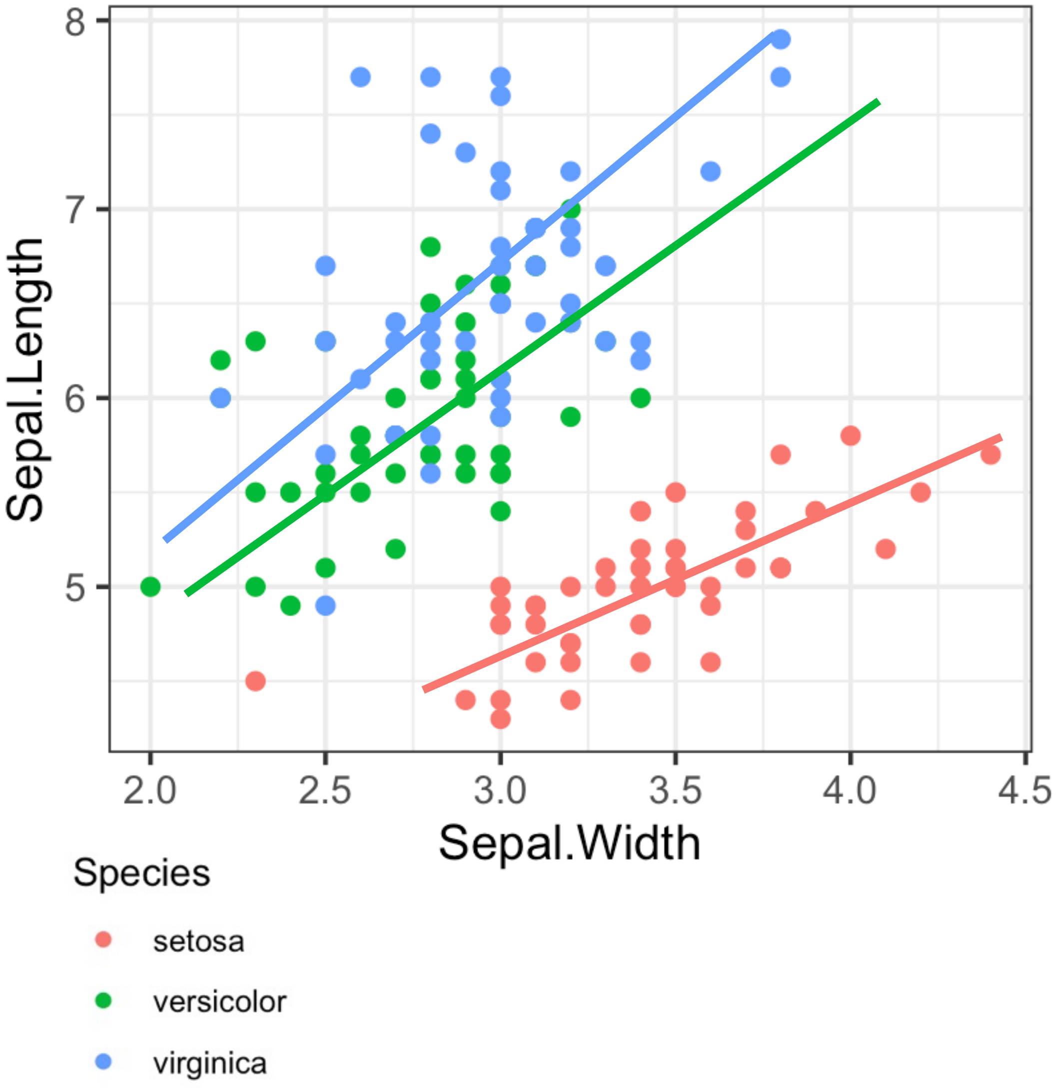
- categorical x categorical ✓
- categorical x continuous
- continuous x continuous

lm(Sepal.Length ~ Sepal.Width + Species, data = iris)



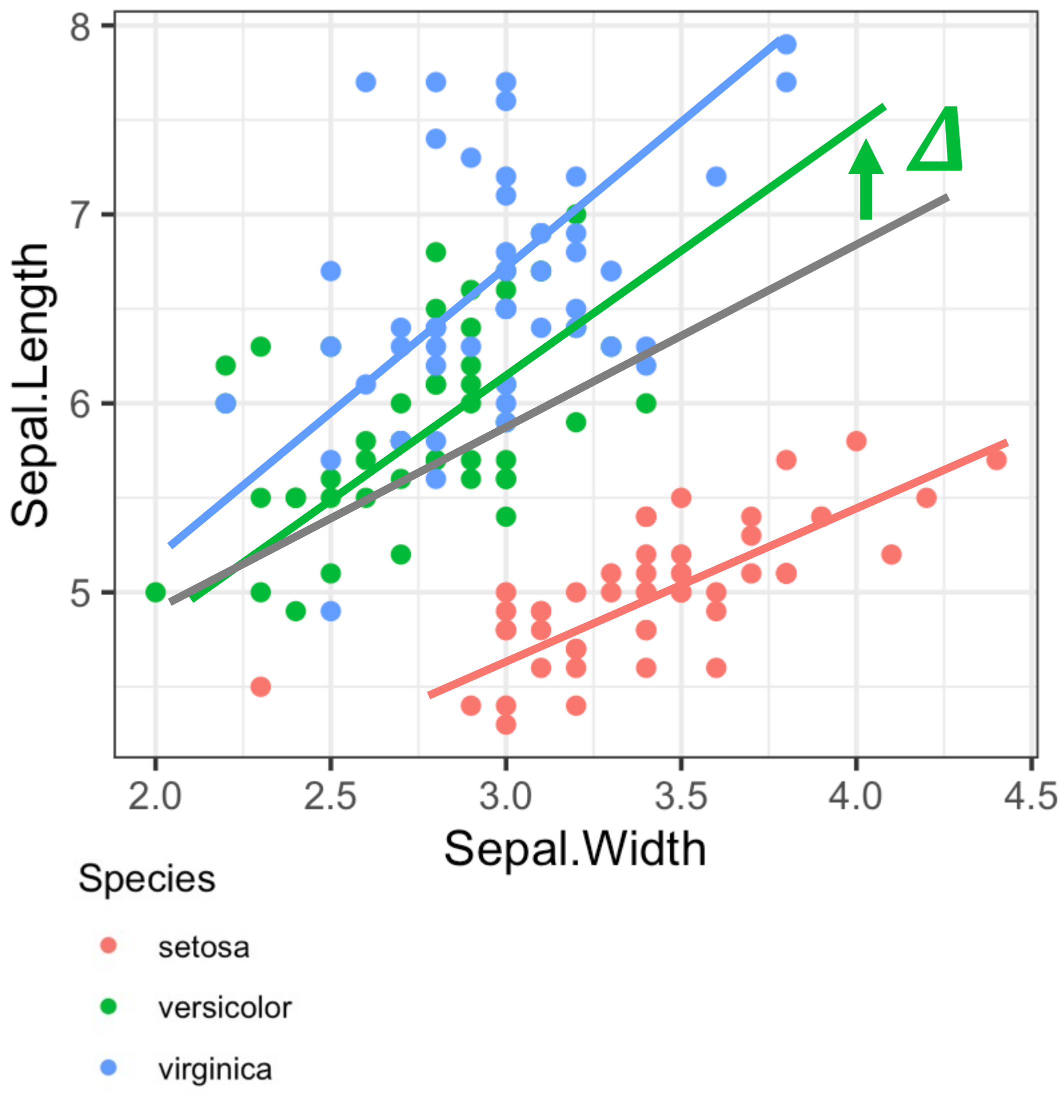
	Estimate
(Intercept)	2.2514
Sepal.Width	$\boxed{0.8036}$
Speciesversicolor	1.4587
Speciesvirginica	1.9468

lm(Sepal.Length ~ Sepal.Width * Species, data = iris)



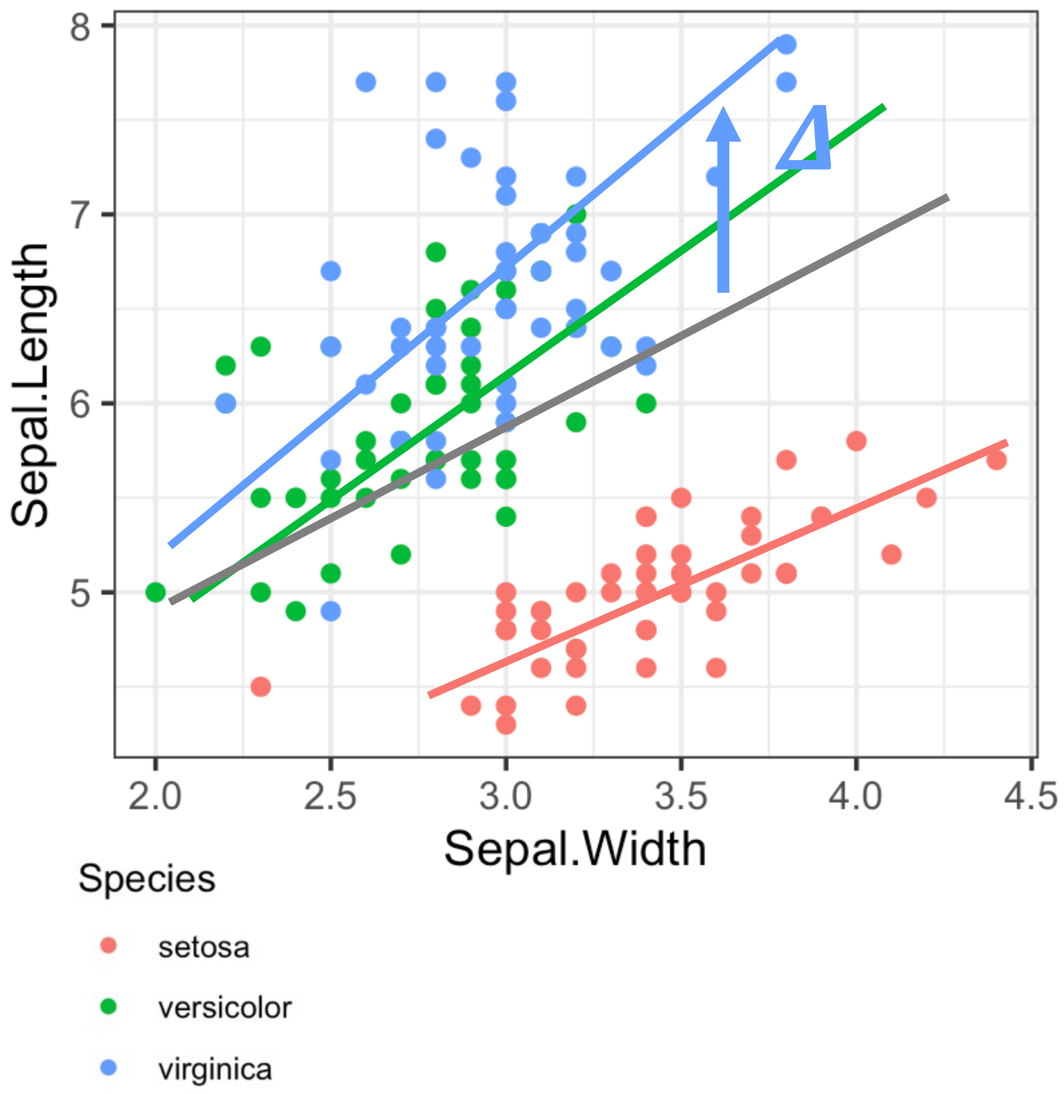
	Estimate
(Intercept)	2.6390
Sepal.Width	0.6905
Speciesversicolor	0.9007
Speciesvirginica	1.2678
Sepal.Width:Speciesversicolor	0.1746
Sepal.Width:Speciesvirginica	0.2110

lm(Sepal.Length ~ Sepal.Width * Species, data = iris)



	Estimate
(Intercept)	2.6390
Sepal.Width	$\boxed{0.6905}$
Speciesversicolor	0.9007
Speciesvirginica	1.2678
Sepal.Width:Speciesversicolor	$\boxed{0.1746}$
Sepal.Width:Speciesvirginica	0.2110

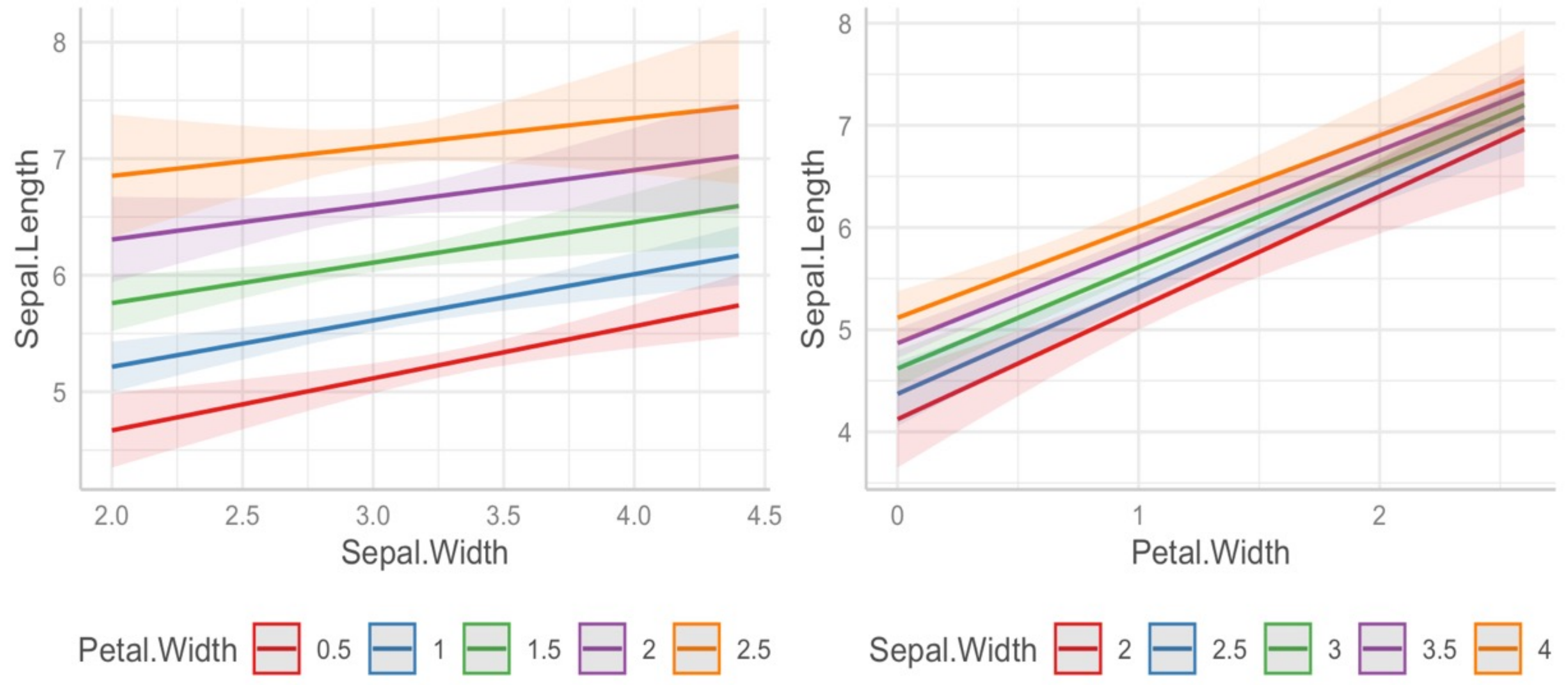
lm(Sepal.Length ~ Sepal.Width * Species, data = iris)



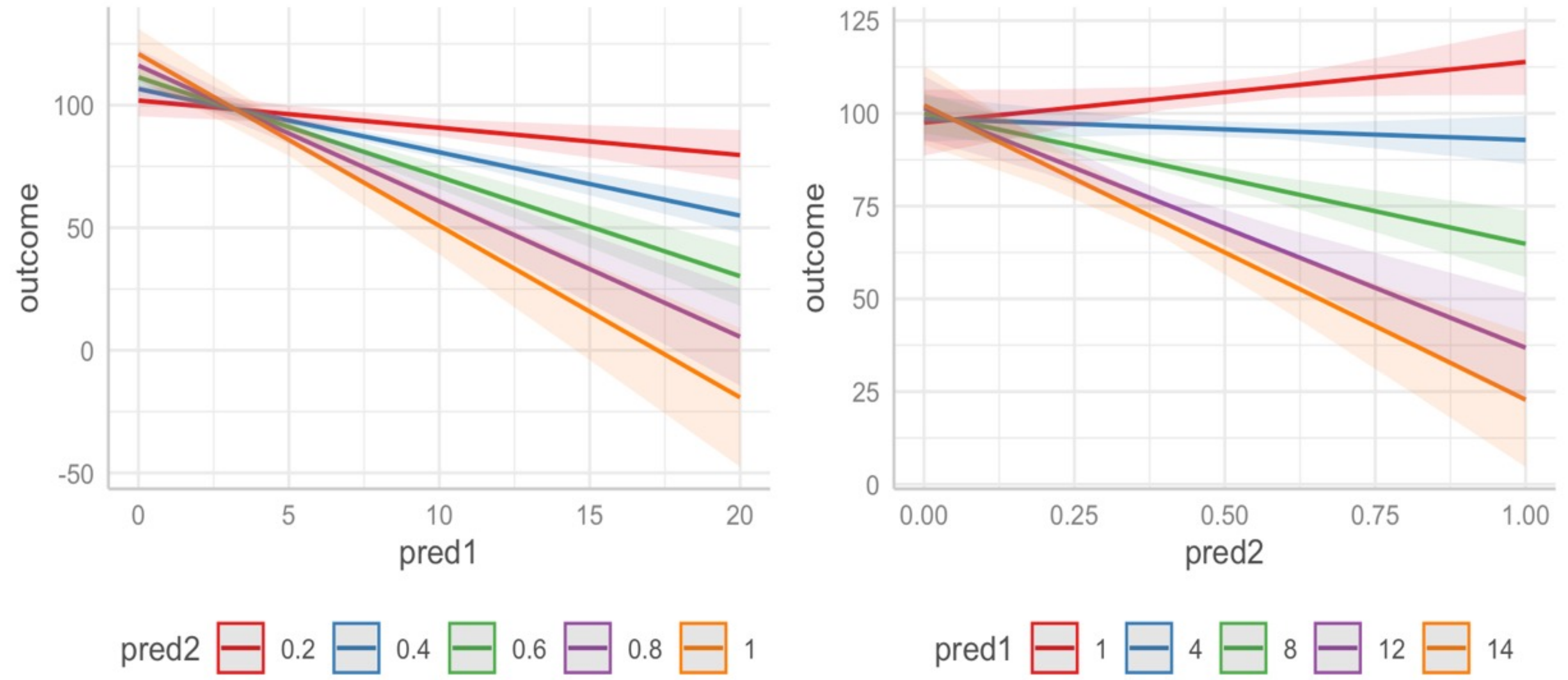
	Estimate
(Intercept)	2.6390
Sepal.Width	$\boxed{0.6905}$
Speciesversicolor	0.9007
Speciesvirginica	1.2678
Sepal.Width:Speciesversicolor	0.1746
Sepal.Width:Speciesvirginica	$\boxed{0.2110}$

type of interactions

- categorical x categorical ✓
- categorical x continuous ✓
- continuous x continuous

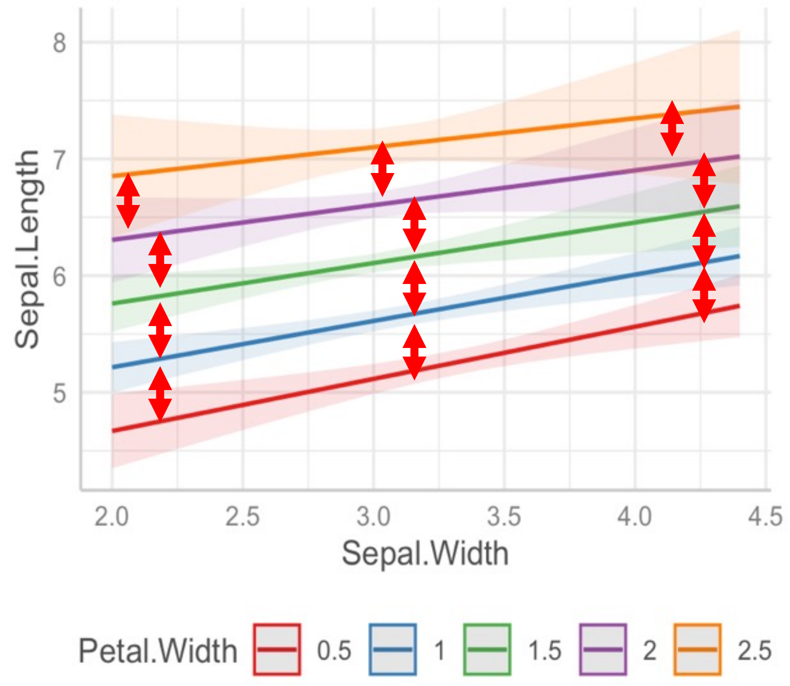


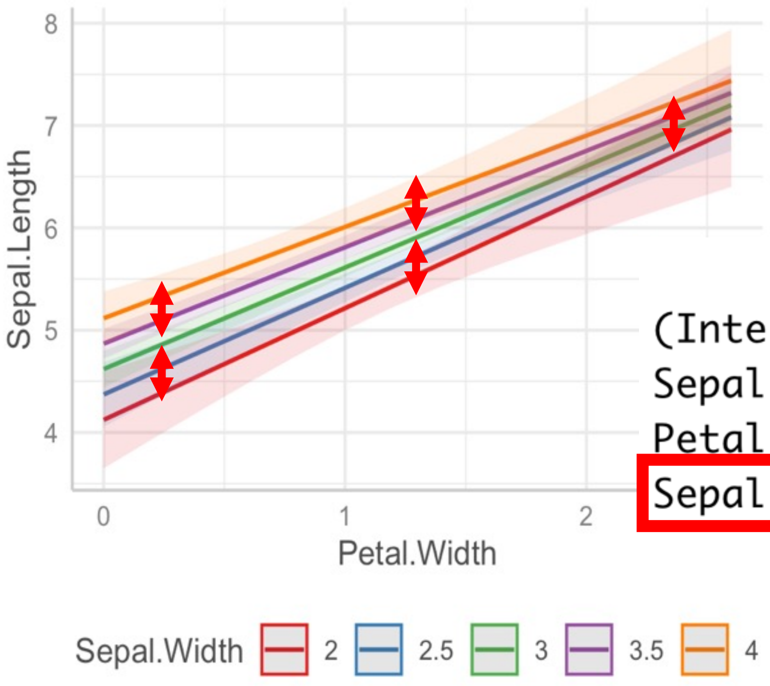
Sepal.length ~ Sepal.width * Petal.width



outcome ~ pred1 * pred2

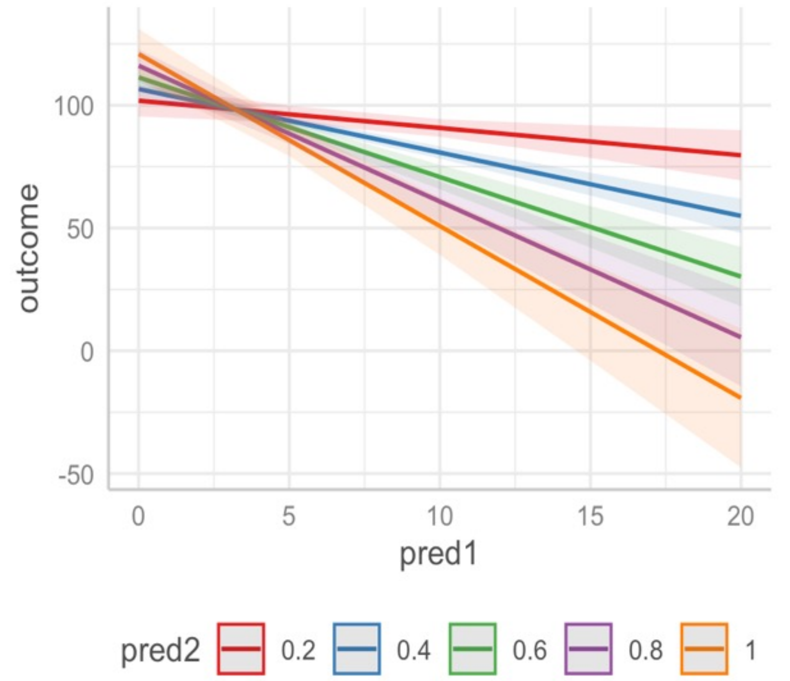
Sepal.length ~ Sepal.width * Petal.width

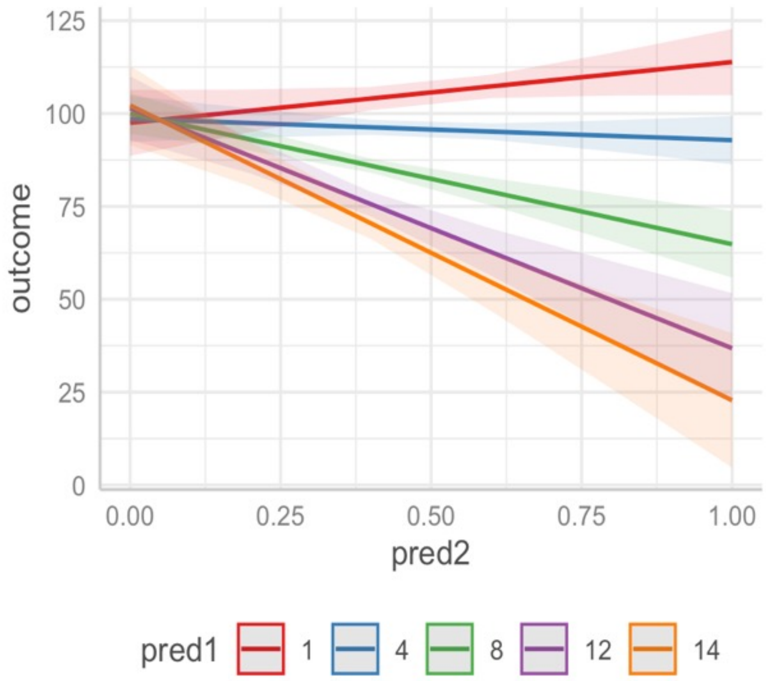


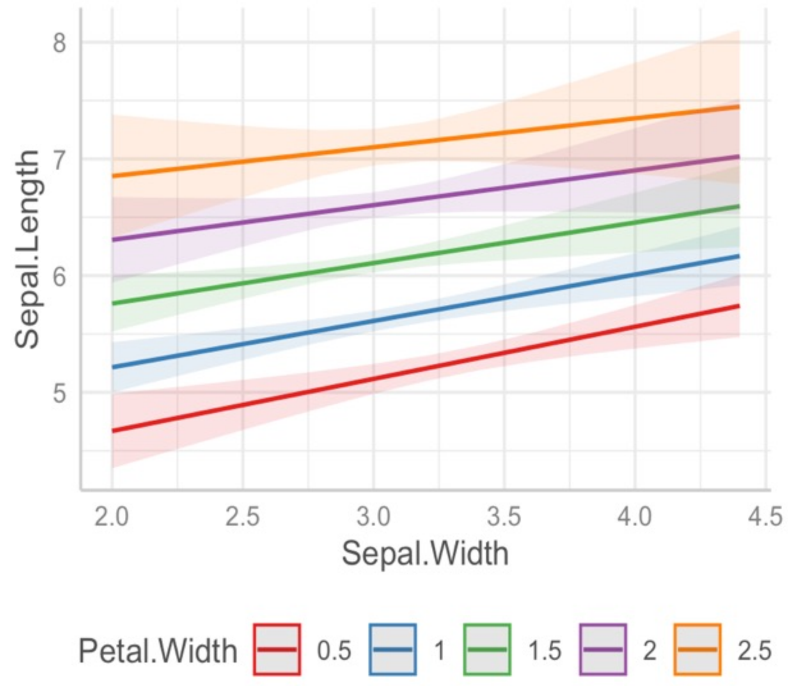


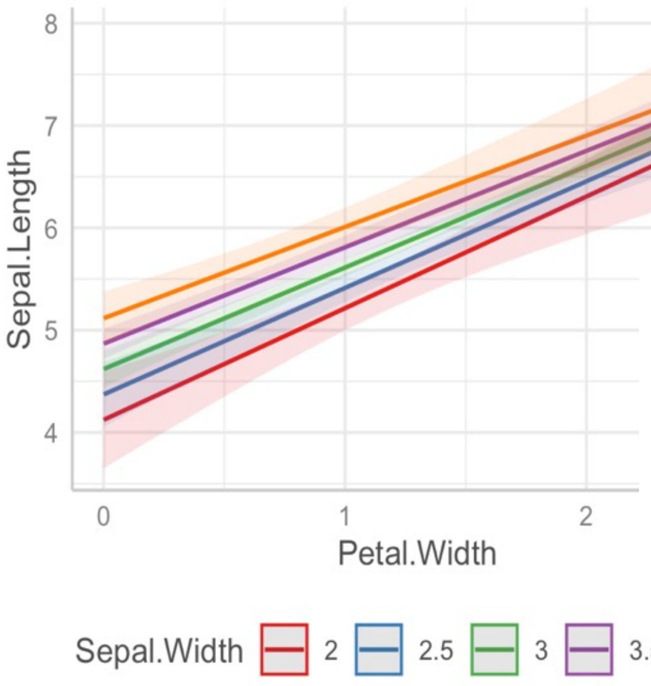
	Estimate	Pr(>|t|)	
(Intercept)	3.13046	2.1e-07	***
Sepal.Width	0.49642	0.00418	**
Petal.Width	1.29057	0.00729	**
Sepal.Width:Petal.Width	-0.09946	0.50038	

outcome ~ pred1 * pred2



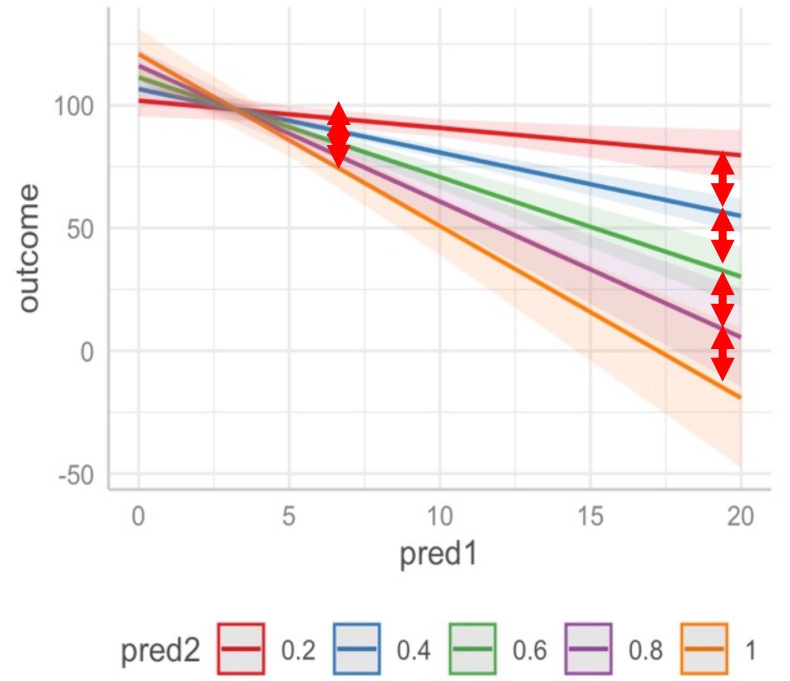


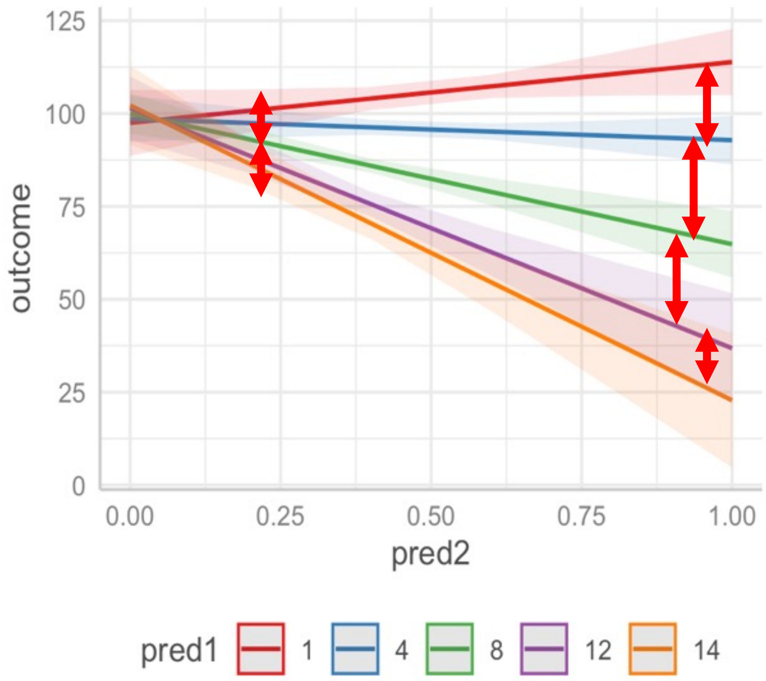




Sepal.length ~ Sepal.width * Petal.width

	Estimate	Pr(>|t|)	
(Intercept)	3.13046	2.1e-07	***
Sepal.Width	0.49642	0.00418	**
Petal.Width	1.29057	0.00729	**
Sepal.Width:Petal.Width	-0.09946	0.50038	





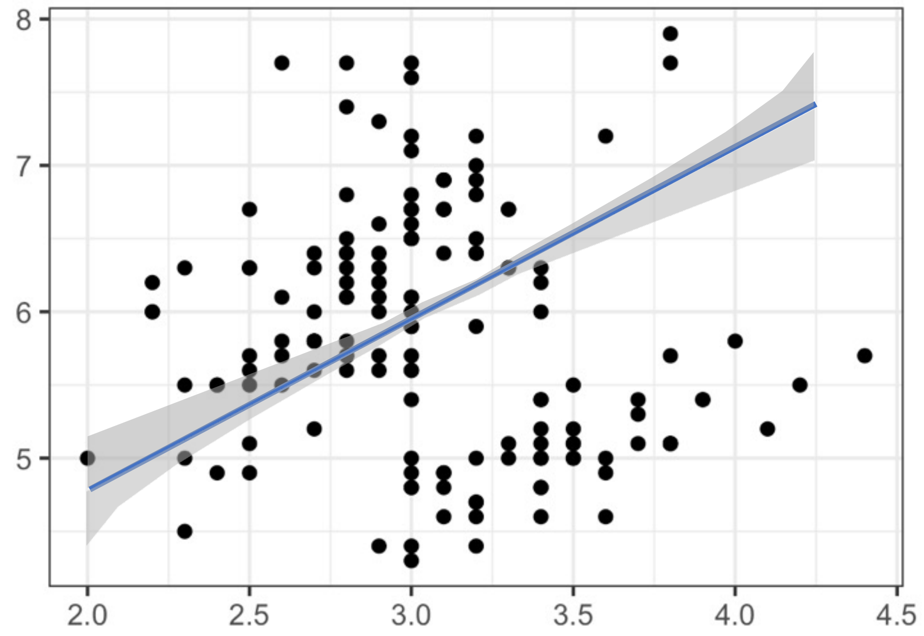
outcome ~ pred1 * pred2

	Estimate	Pr(>|t|)	
(Intercept)	97.1423	< 2e-16	***
pred1	0.3617	0.5743	
pred2	23.7294	0.0155	*
pred1:pred2	-7.3663	9.6e-07	***

type of interactions

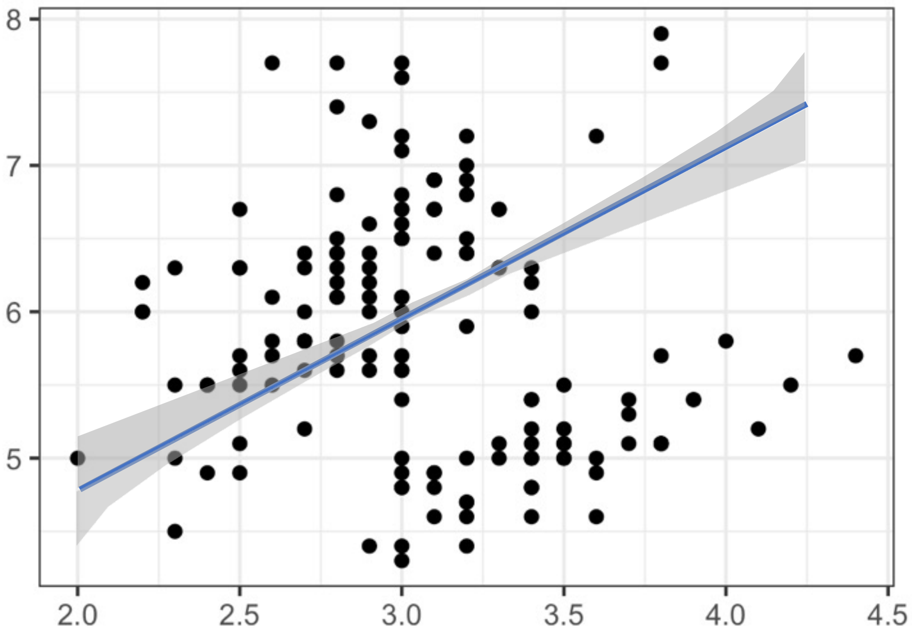
- categorical x categorical ✓
- categorical x continuous ✓
- continuous x continuous ✓

Mixed Model



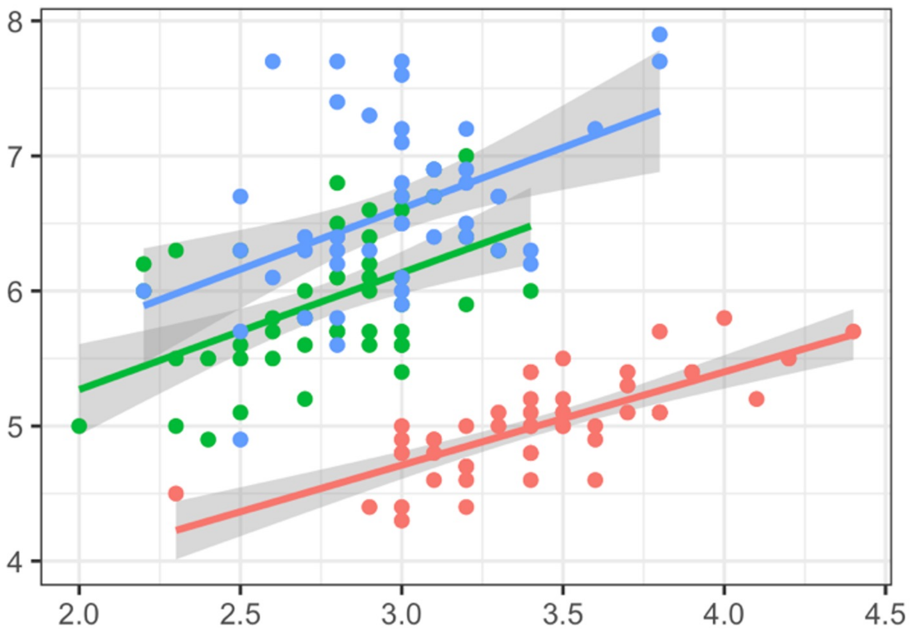
Simple linear regression

$$y \sim \beta_0 + \beta x + \varepsilon$$



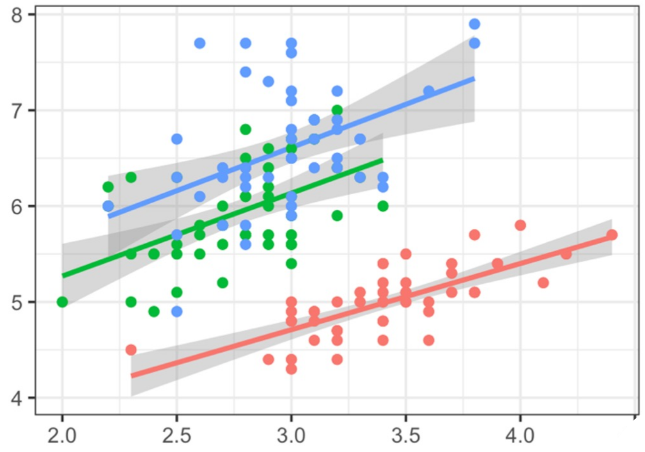
Simple linear regression

$$y \sim \beta_0 + \beta x + \varepsilon$$



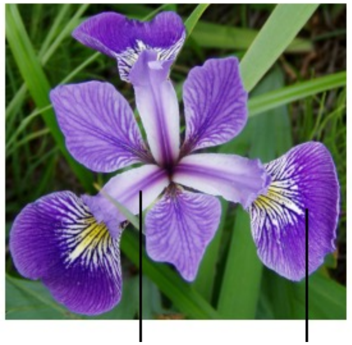
Multiple linear regression

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_N x_N + \varepsilon$$

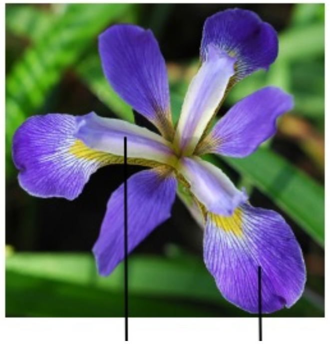


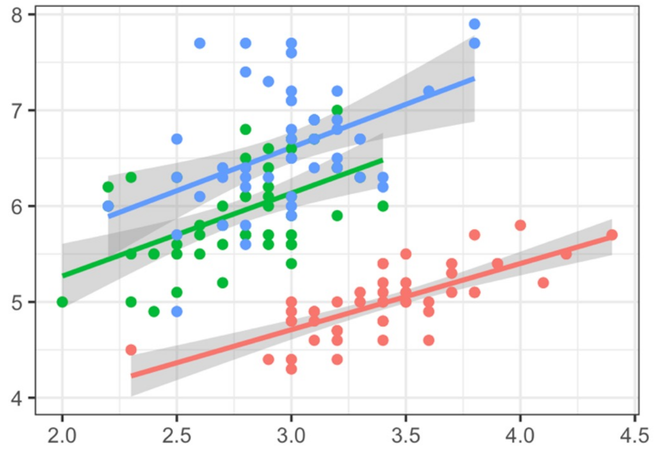
iris setosa

iris versicolor



iris virginica



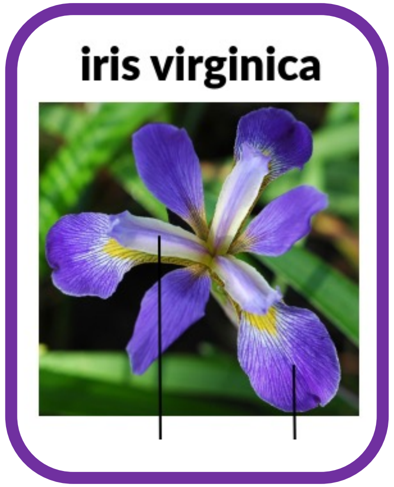


8
7
6
5
4
2.0
2.5
3.0
3.5
4.0
4.5

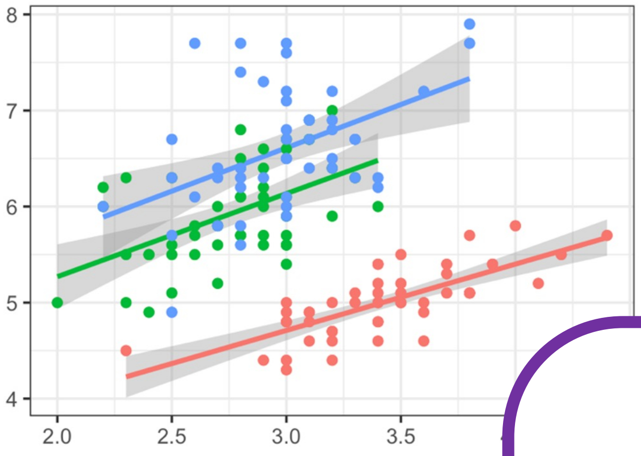
Species

iris setosa

iris versicolor



iris virginica

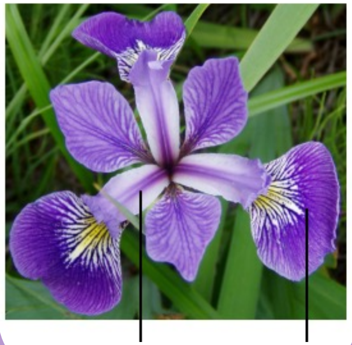


Genus Iris

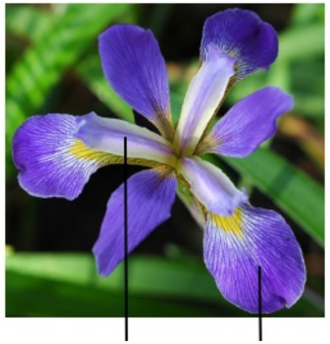
Species

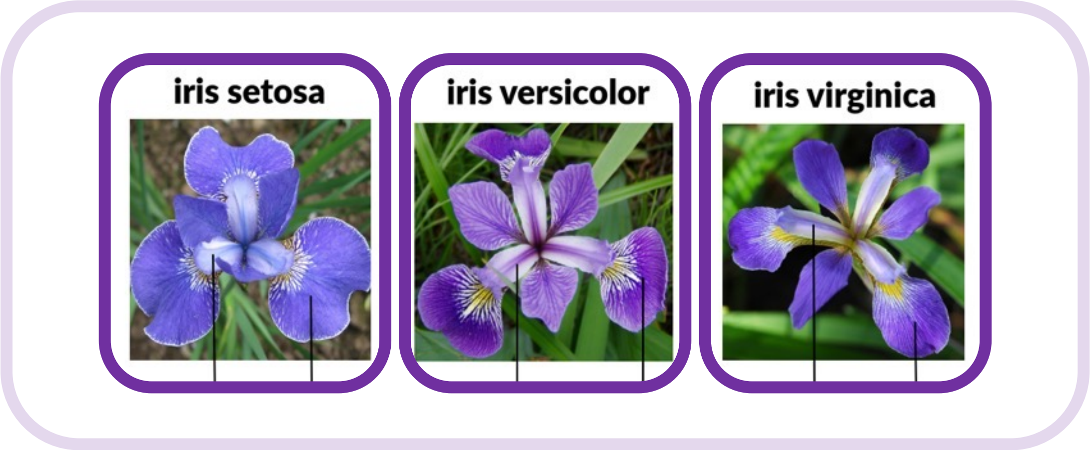
iris setosa

iris versicolor



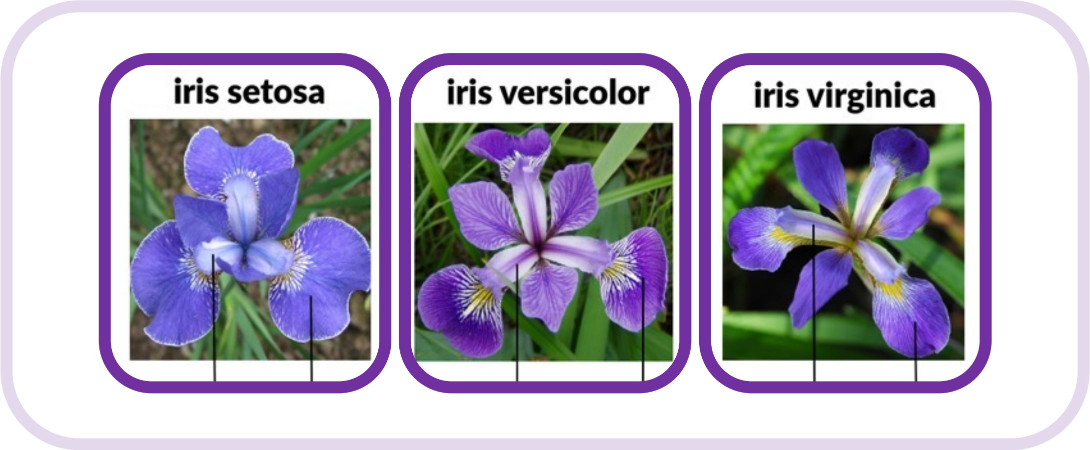
iris virginica





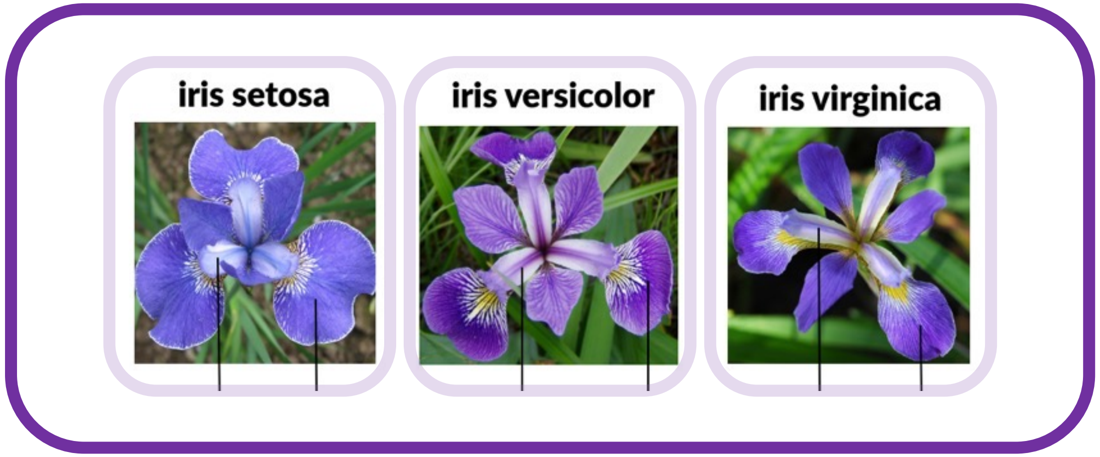
Species as

Fixed Effects



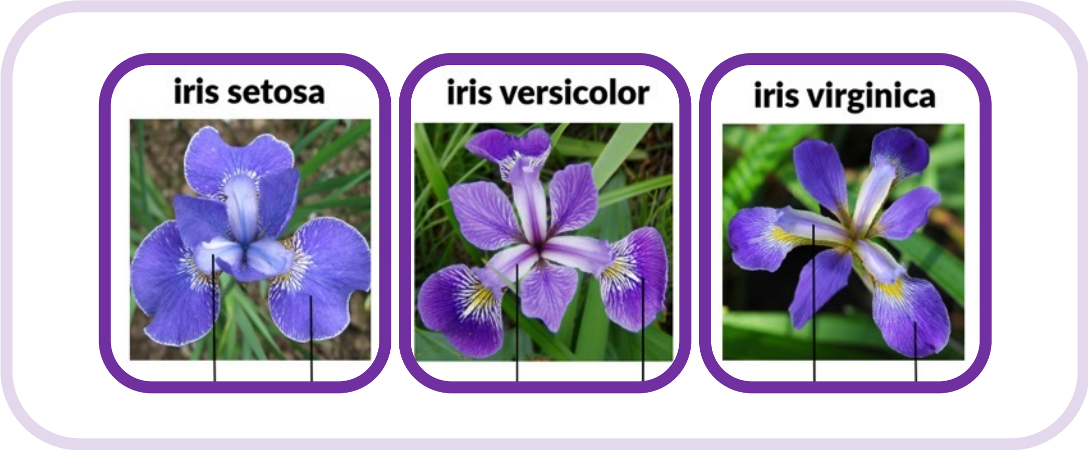
Species as

Fixed Effects

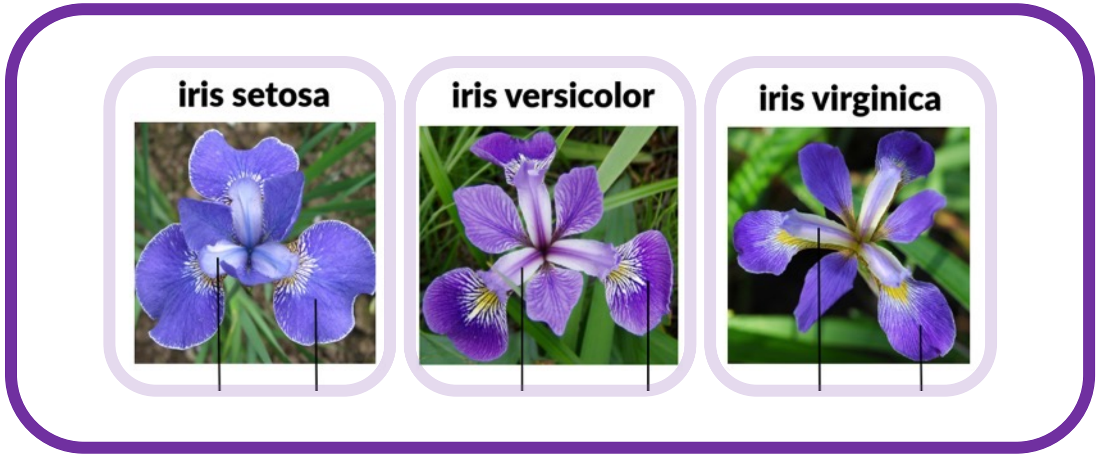


Species as

Random Effects

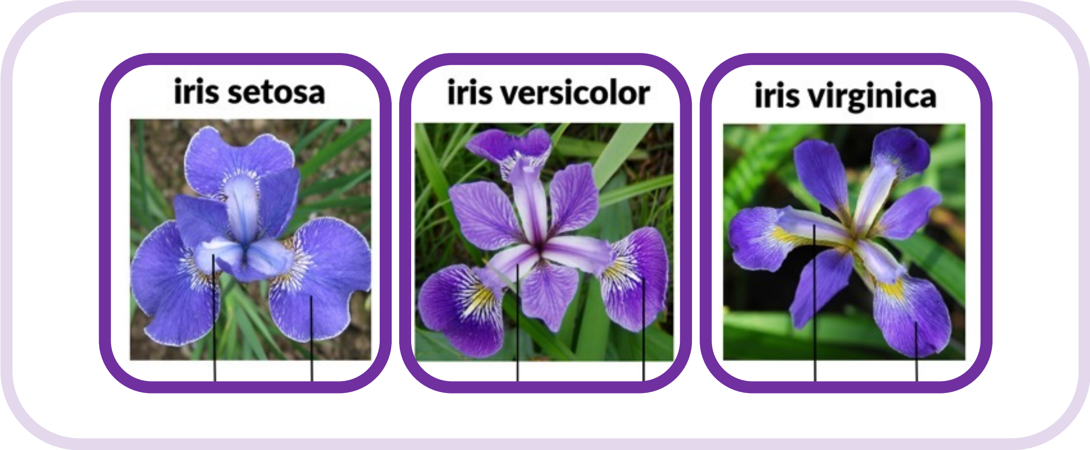


$y \sim x +$ Species

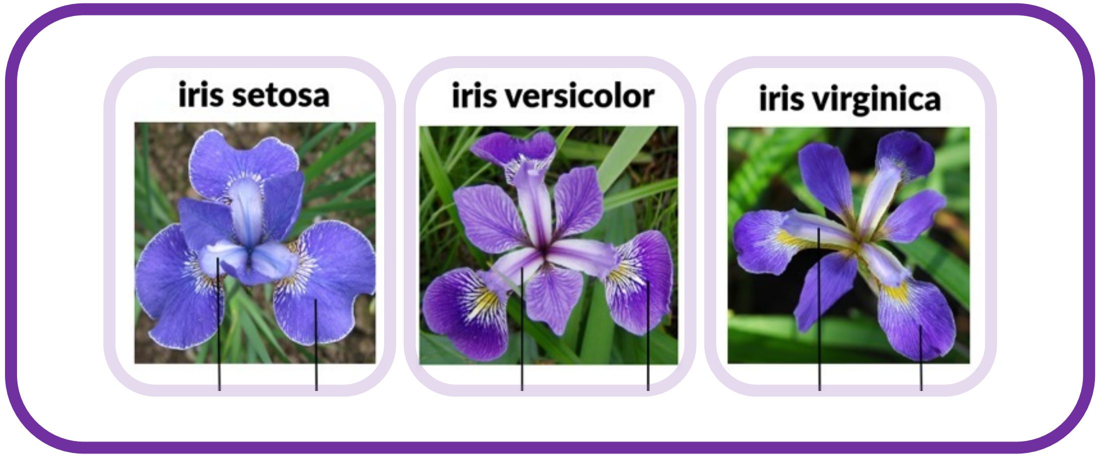


Species as

Random Effects



y ~ x + Species



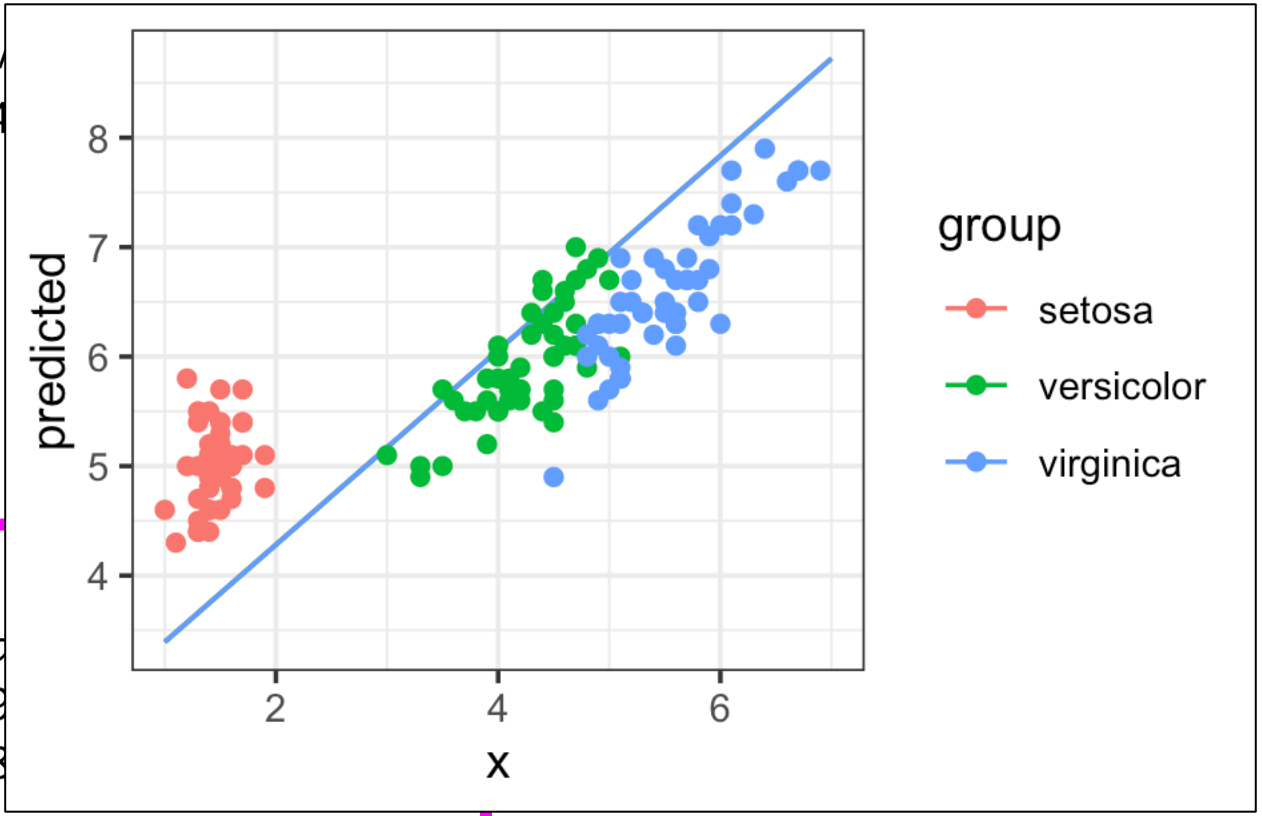
y ~ x + (1 | Species)

Sepal.Length ~ Petal.Length + (1 | Species)

```
Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.23334 -0.67332 -0.01684  0.67680  3.04429

Random effects:
 Groups   Name        Variance Std.Dev.
 Species  (Intercept) 1.1617   1.0778
 Residual             0.1143   0.3381
Number of obs: 150, groups:  Species, 3

Fixed effects:
              Estimate Std. Error        df t value Pr(>|t|)
(Intercept)    2.50446    0.66743   2.45496   3.752   0.0463 *
Petal.Length   0.88847    0.06379 144.56089  13.927   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

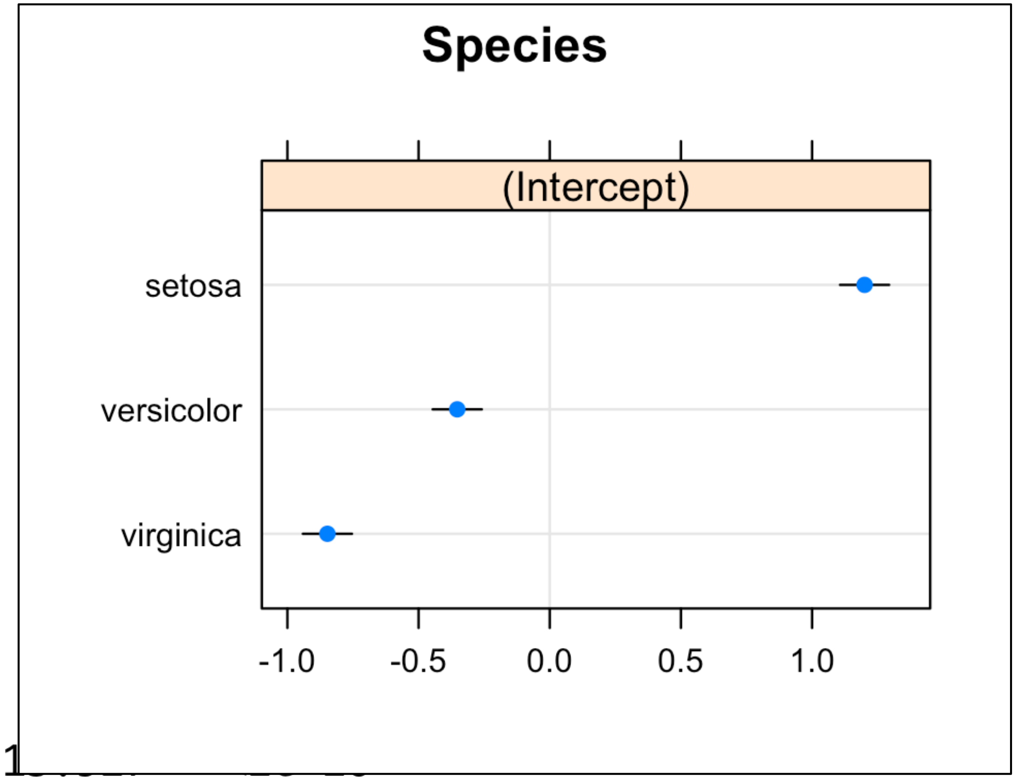
Sepal.Length ~ Petal.Length + (1 | Species)

```
Scaled residuals:
     Min       1Q   Median       3Q      M
-2.23334 -0.67332 -0.01684  0.67680  3.044

Random effects:
 Groups   Name        Variance Std.Dev.
 Species  (Intercept) 1.1617   1.0778
 Residual             0.1143   0.3381
Number of obs: 150, groups:  Species, 3

Fixed effects:
              Estimate Std. Error       d
(Intercept)    2.50446    0.66743   2.4549
Petal.Length   0.88847    0.06379 144.5608
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



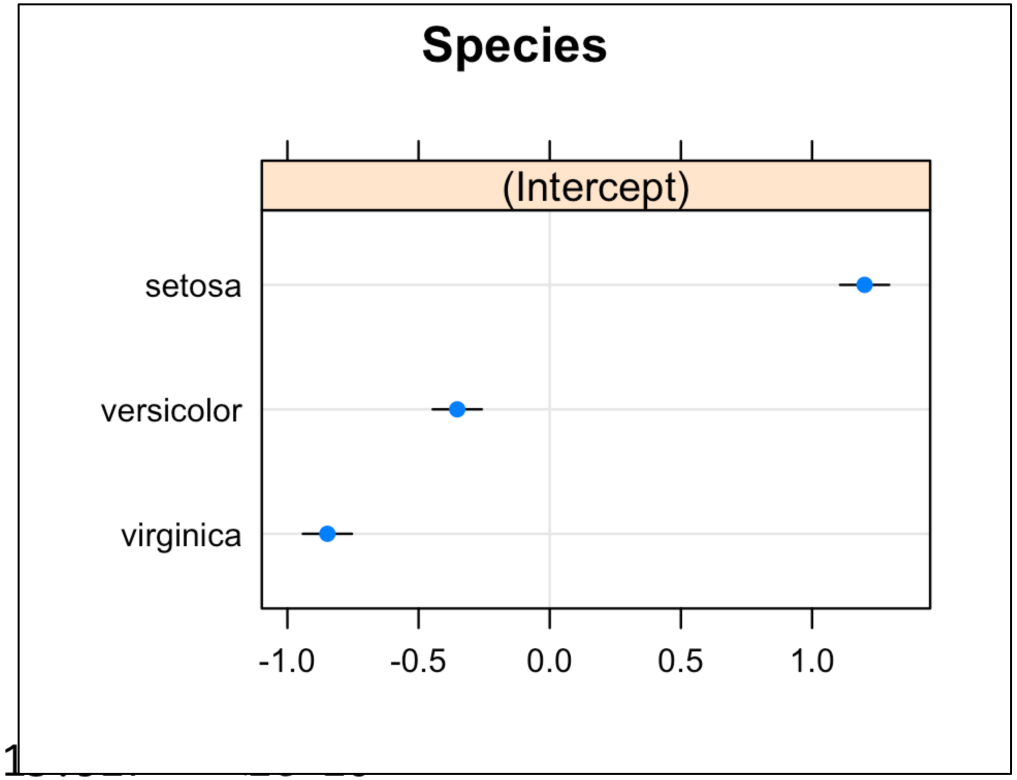
Sepal.Length ~ Petal.Length + (1 | Species)

```
Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.23334 -0.67332 -0.01684  0.67680  3.04429

Random effects:
 Groups   Name        Variance Std.Dev.
 Species  (Intercept) 1.1617   1.0778
 Residual             0.1143   0.3381
Number of obs: 150, groups:  Species, 3

Fixed effects:
              Estimate Std. Error        df t
(Intercept)    2.50446    0.66743   2.45496
Petal.Length   0.88847    0.06379 144.56089
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

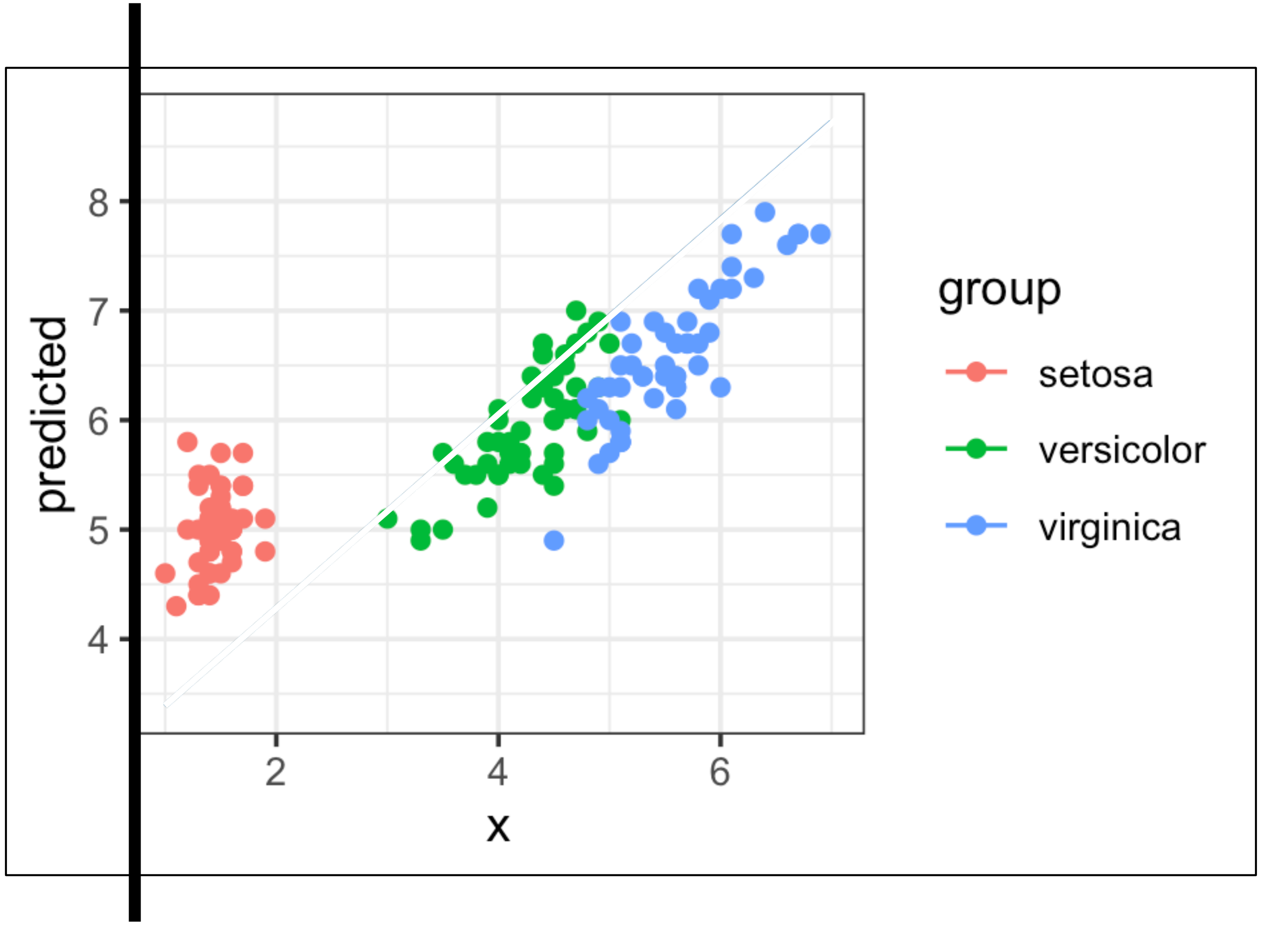


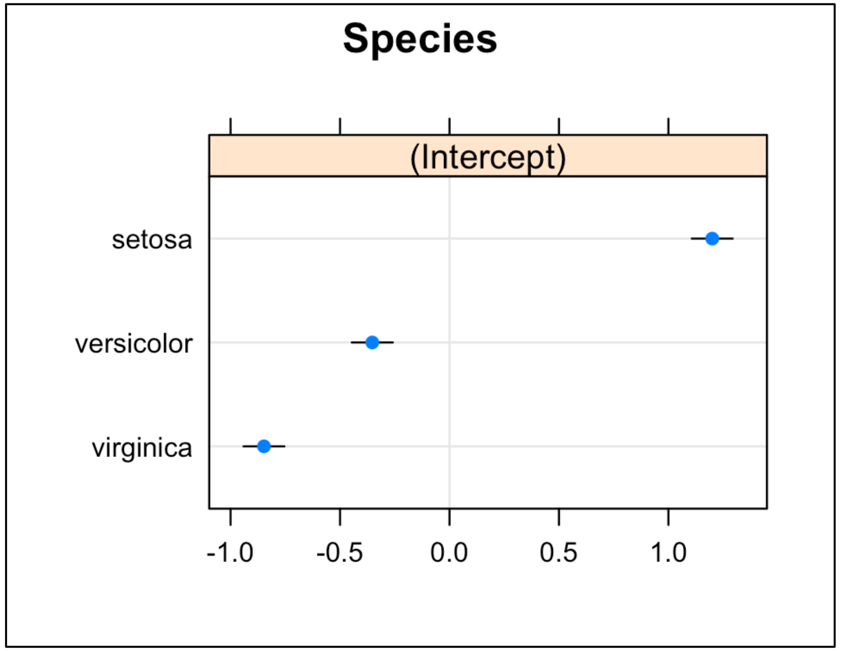
Sepal.Length ~ Petal.Length + (1 | Species)

```
Scaled residuals:
     Min       1Q   Median       3Q      Max
-2.23334 -0.67332 -0.01684  0.67680  3.04429

Random effects:
 Groups   Name        Variance Std.Dev.
 Species  (Intercept) 1.1617   1.0778
 Residual             0.1143   0.3381
Number of obs: 150, groups:  Species, 3

Fixed effects:
              Estimate Std. Error        df t
(Intercept)    2.50446    0.66743   2.45496
Petal.Length   0.88847    0.06379 144.56089  1
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

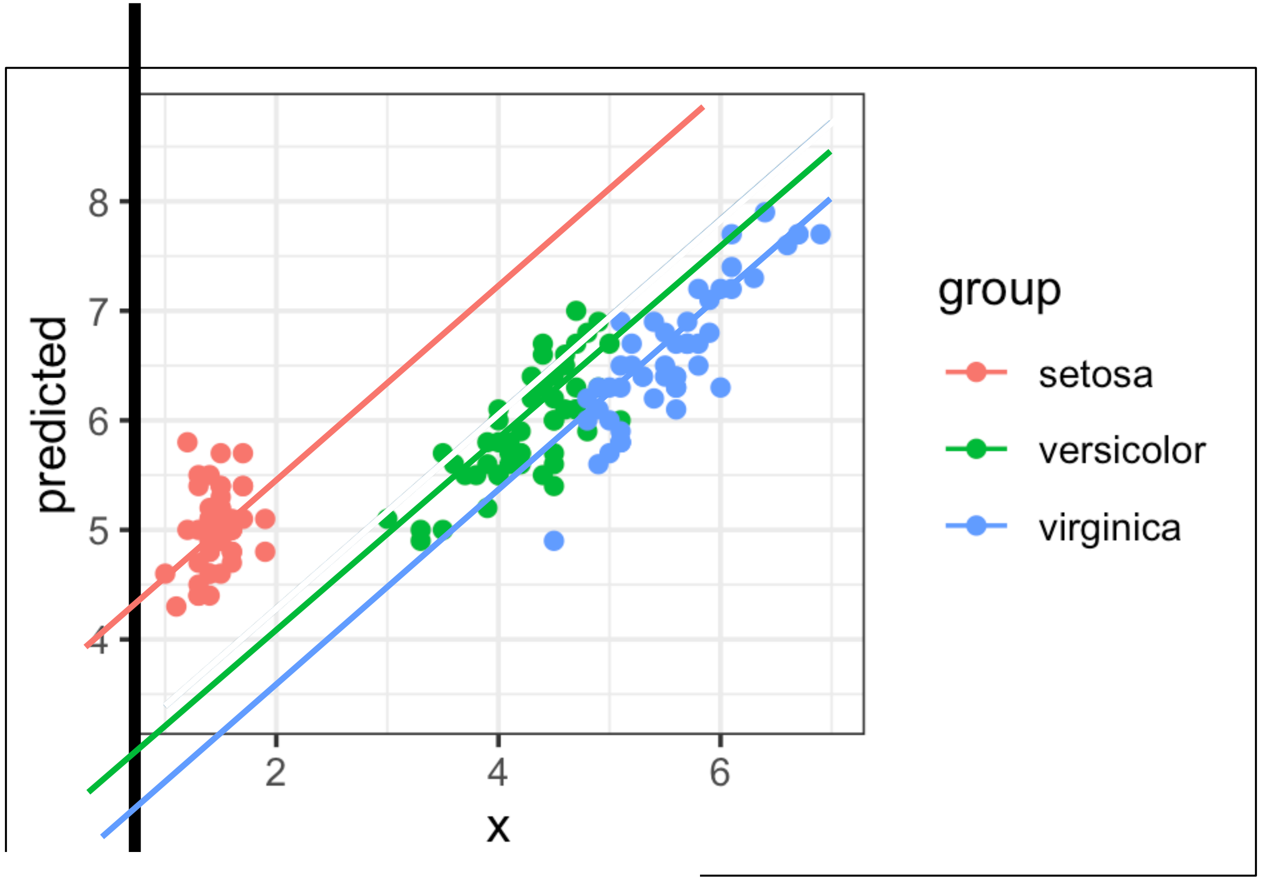


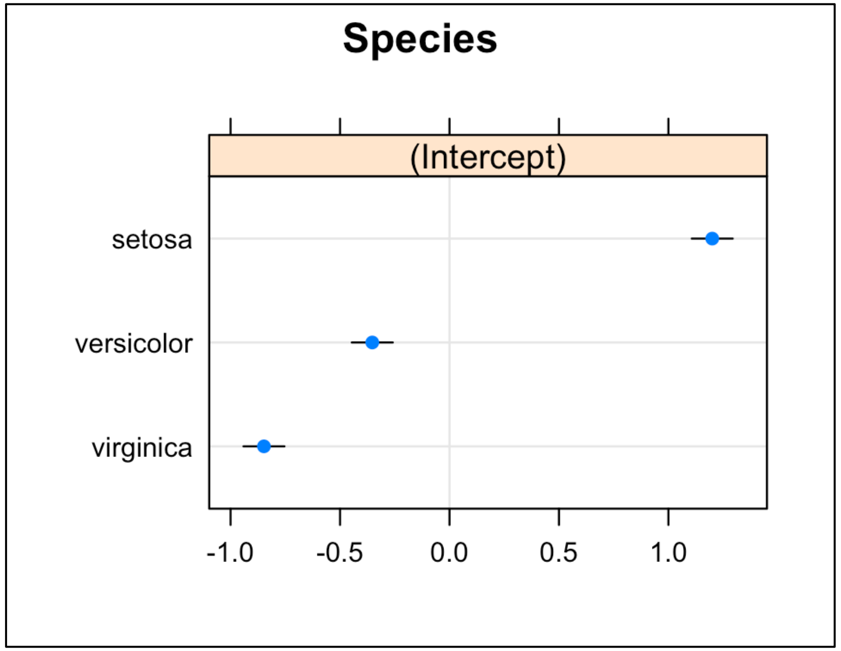
Sepal.Length ~ Petal.Length + (1 | Species)





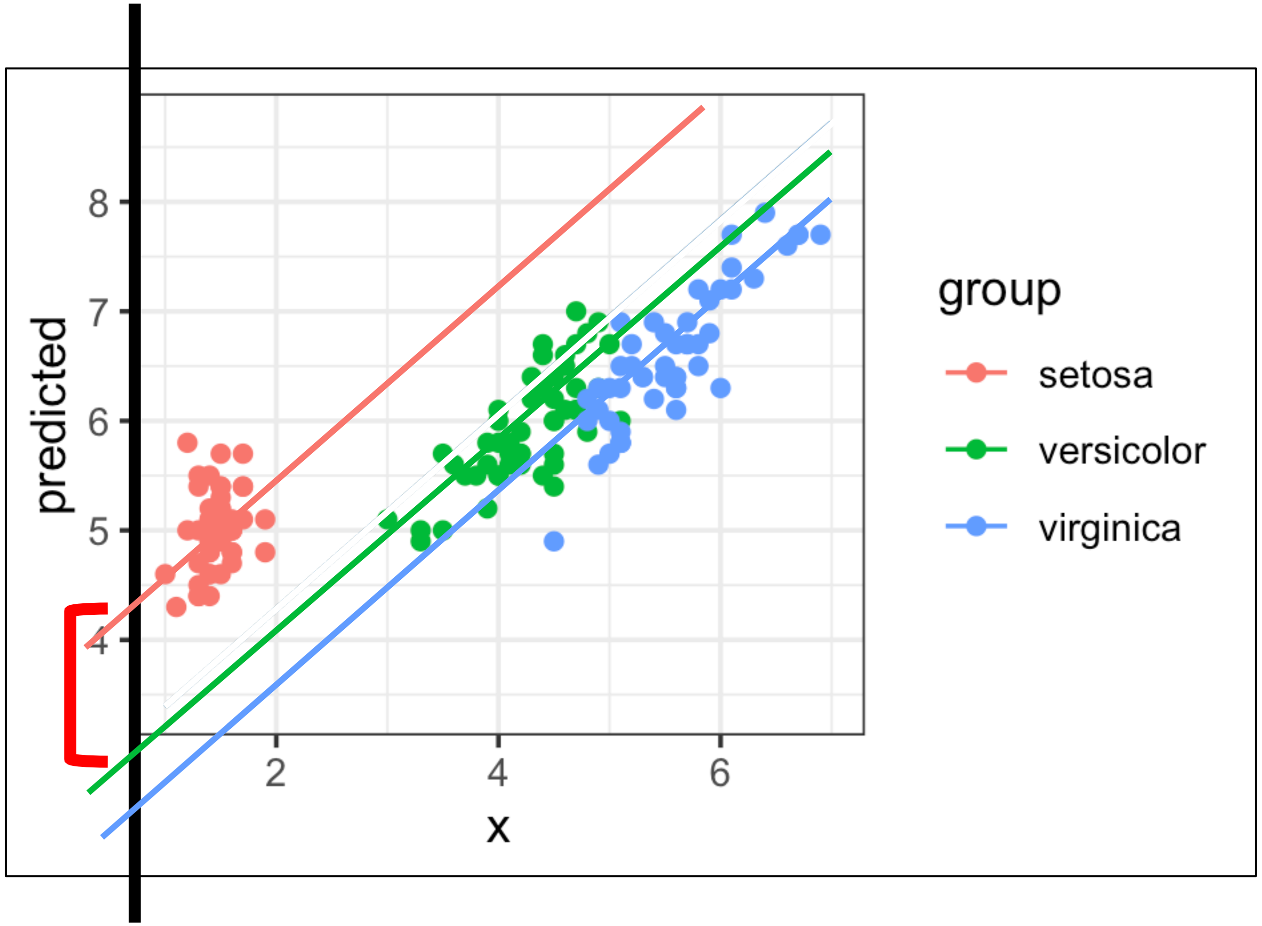
Sepal.Length ~ Petal.Length + (1 | Species)

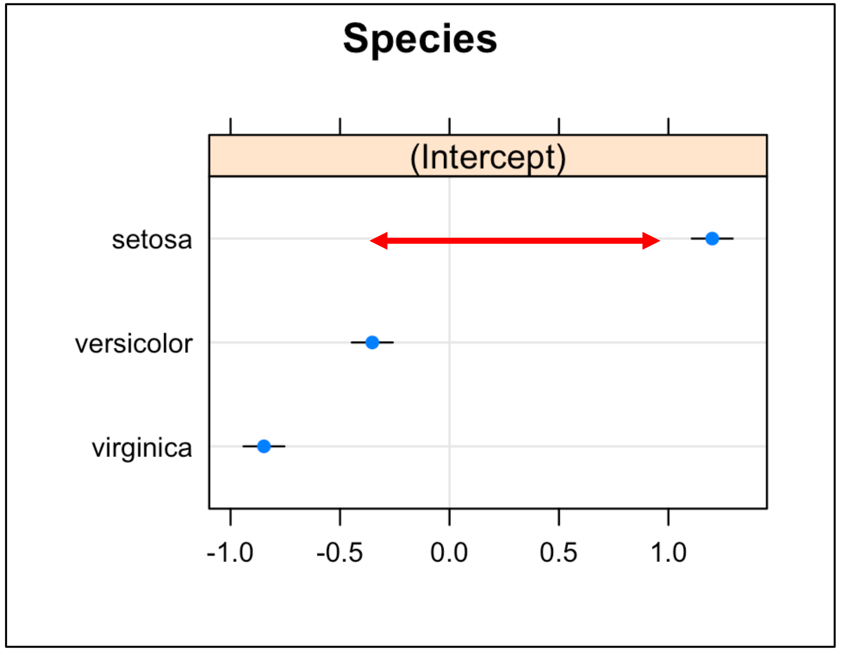




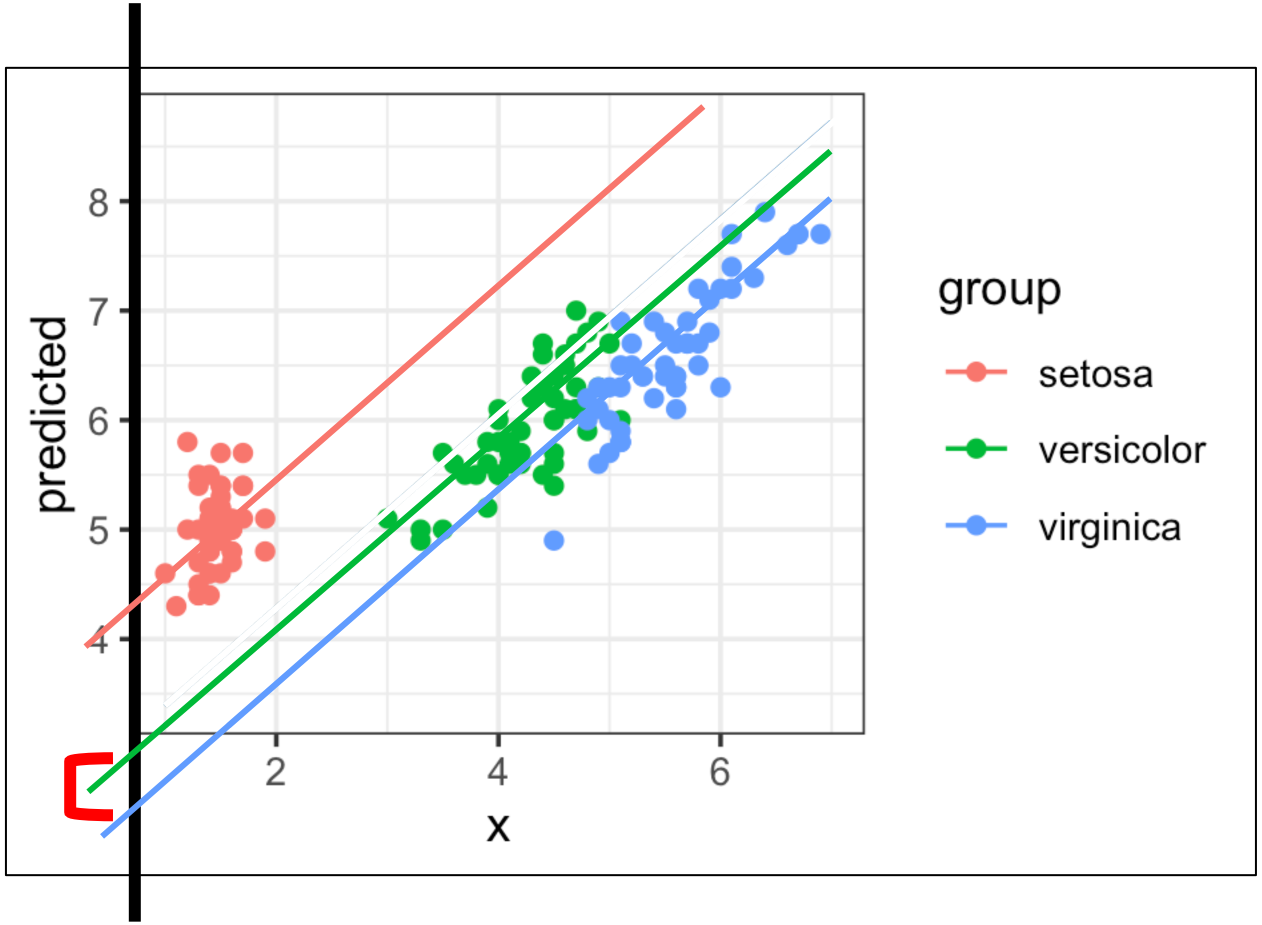
Random Intercepts

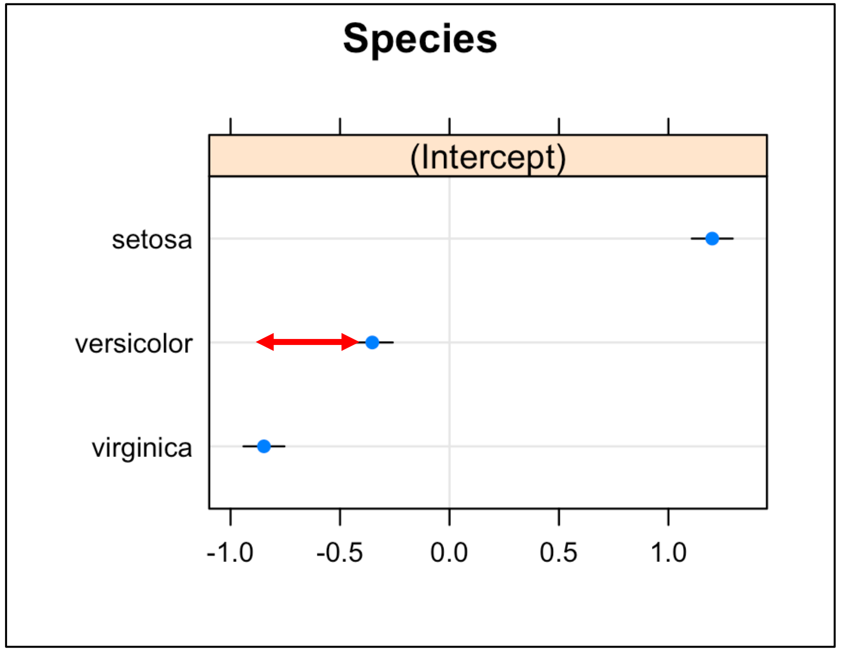
Sepal.Length ~ Petal.Length + (1 | Species)



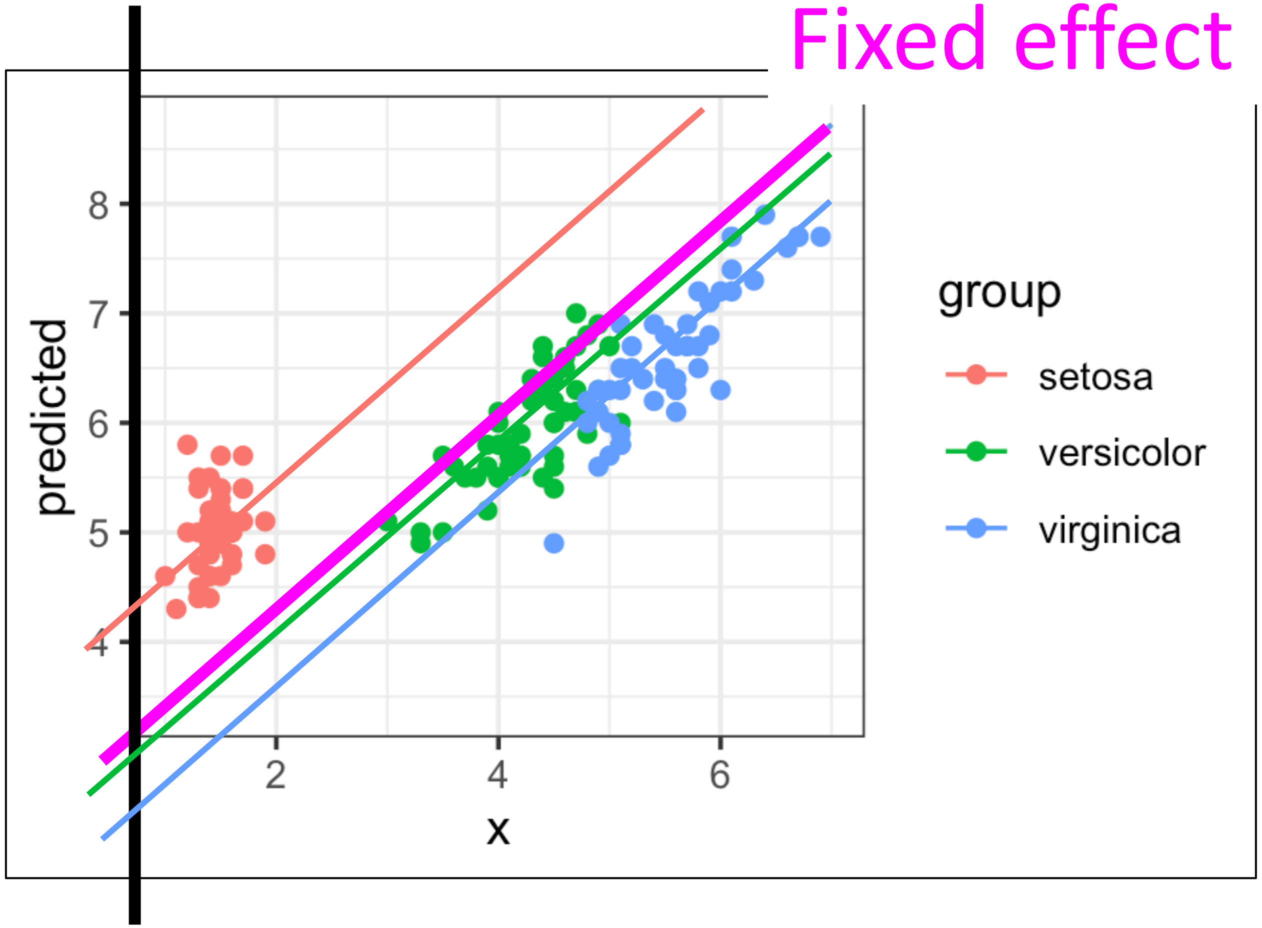


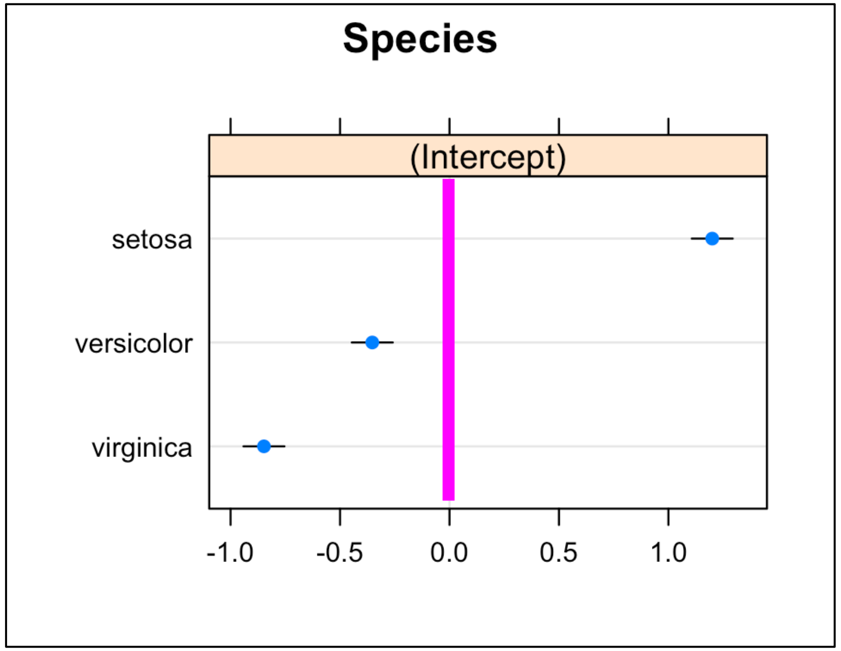
Sepal.Length ~ Petal.Length + (1 | Species)





Sepal.Length ~ Petal.Length + (1 | Species)





Random effects are variables based on grouping definition:

Random effects are variables based on grouping definition:

- Plant and plant fertilizer across areas
- Exam scores and SES across schools
- RT and condition across participants

Random effects are variables based on grouping definition:

- Plant and plant fertilizer across areas
- Exam scores and SES across schools
- RT and condition across participants

They are mostly for repeated measures:

- Repeatedly measure plant growth within each area
- Repeatedly measure exam scores within each school
- Repeatedly measure RT within each participant

Summary

- Linear regression with categorical variable
- Multiple linear regression with categorical variable
- Linear regression with interactions
- emmeans
- ggeffect
- Linear mixed model

Thank you!

junho.chai@chosun.ac.kr