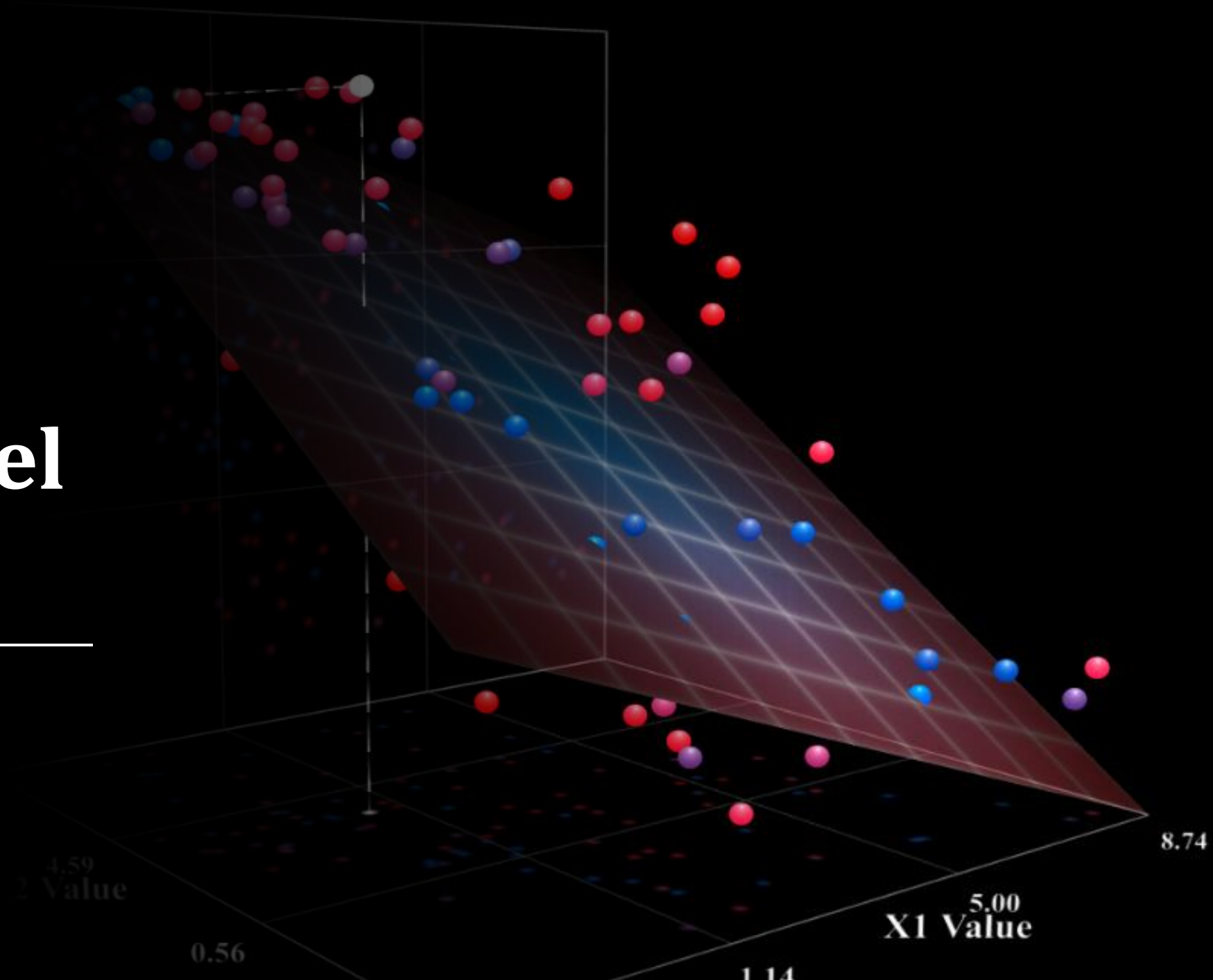


# General Linear Model Approach

---



# The General Linear Model

- The **General Linear Model** is a general framework for analyzing the effects of **continuous** and **categorical** predictor variables on a **continuous outcome variable** using a multiple linear regression model (or simple linear regression in the case of only one predictor).

# The Basic Equation for Data Analysis

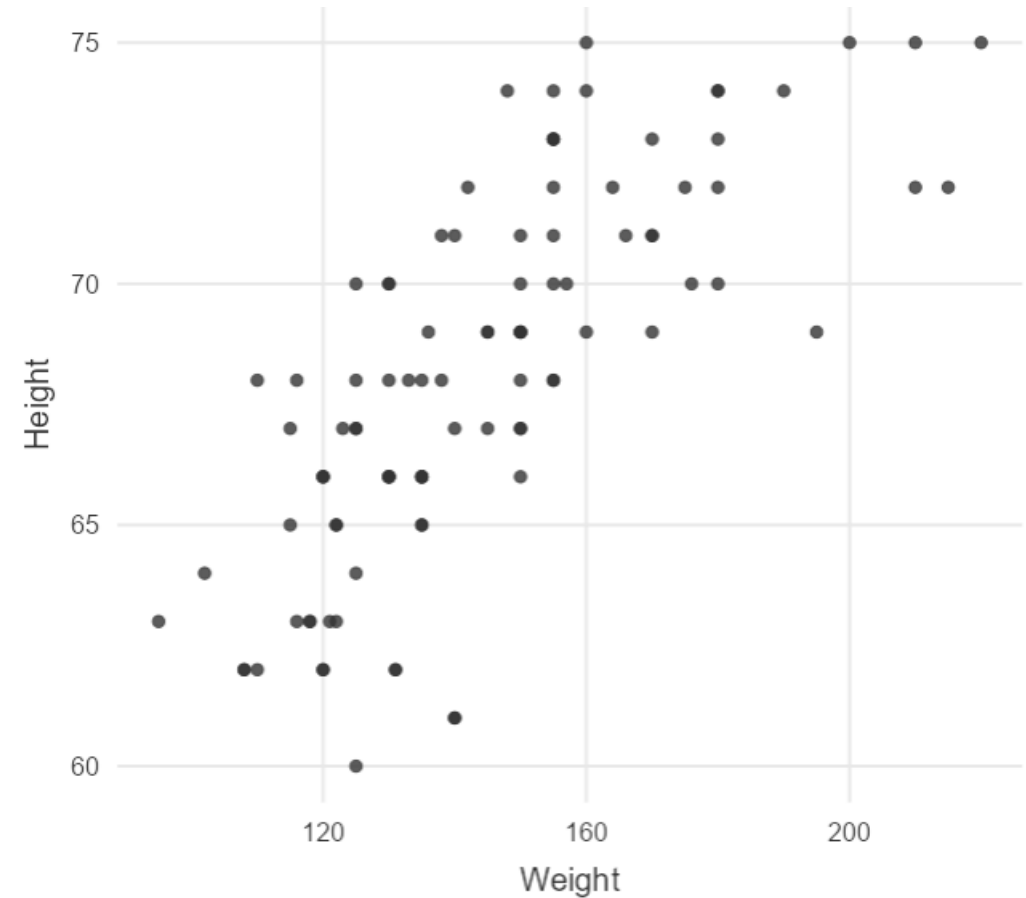
$$\text{Data} = \text{Model} + \text{Error}$$

- ❑ Data = our empirical observations
- ❑ Model = an algebraic expression that attempts to compactly represent the data
- ❑ Error = the part of our data left unexplained by the model

# Example of Data

Data:

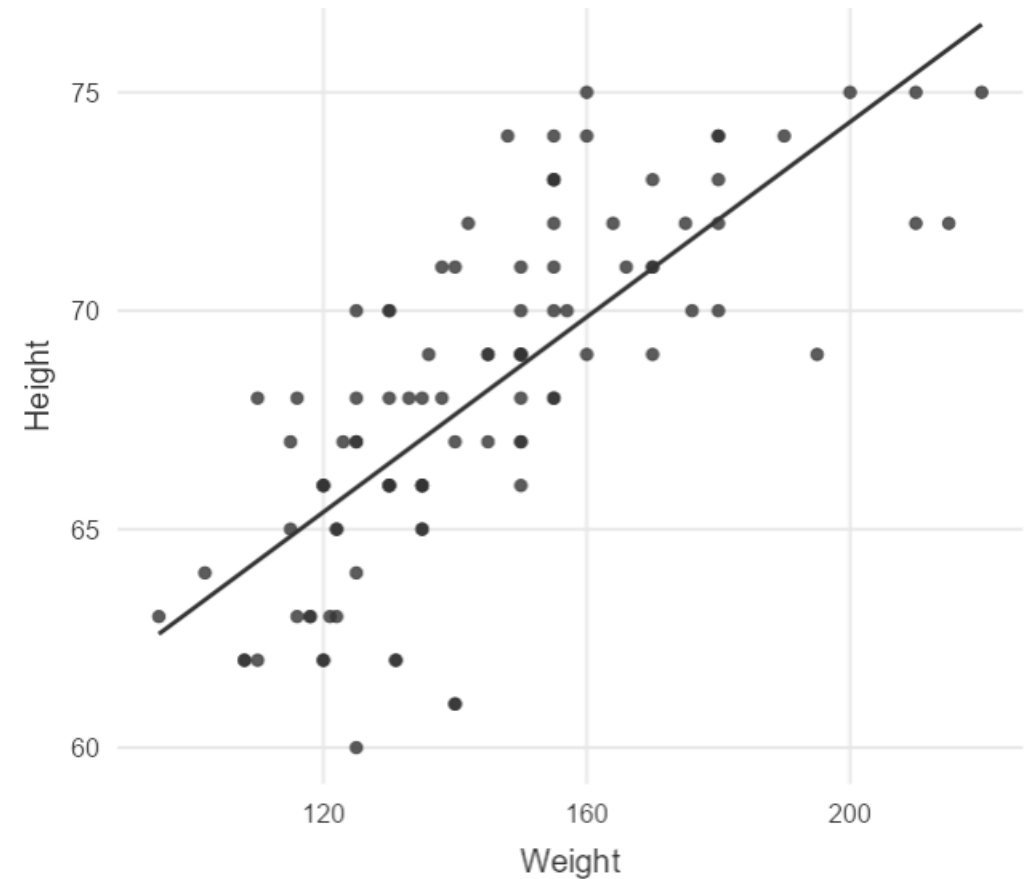
- Say I asked 100 people to report their height and weight. Their *data* is shown in the scatterplot to the right.



# Example of a Model

Model:

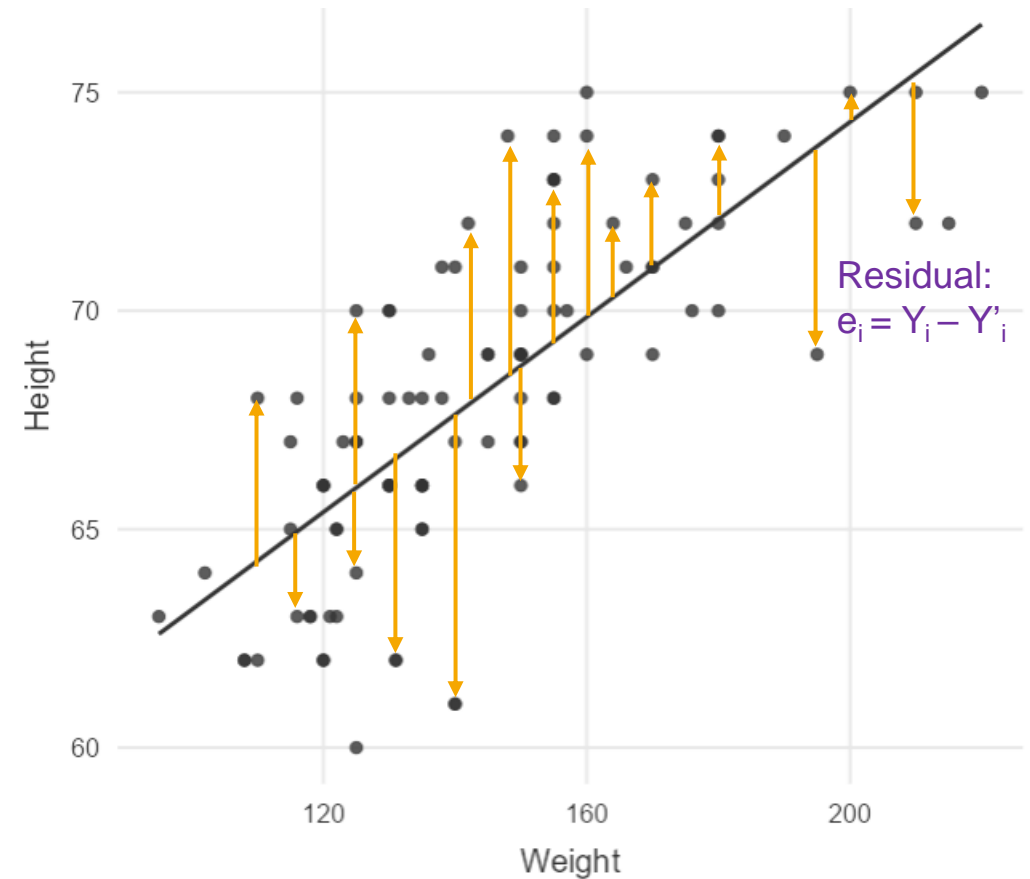
- Now I've added my *model* to the data.
- I'm using the geometric shape of a line to attempt to compactly represent the pattern in the data
- This model can be expressed using the equation:  $Y' = 0.11 * \text{Weight} + 52$ 
  - $Y'$  means predicted values on  $Y$  (but you can also use  $Y\text{-hat}$ ,  $\hat{Y}$ )



# Example of Error

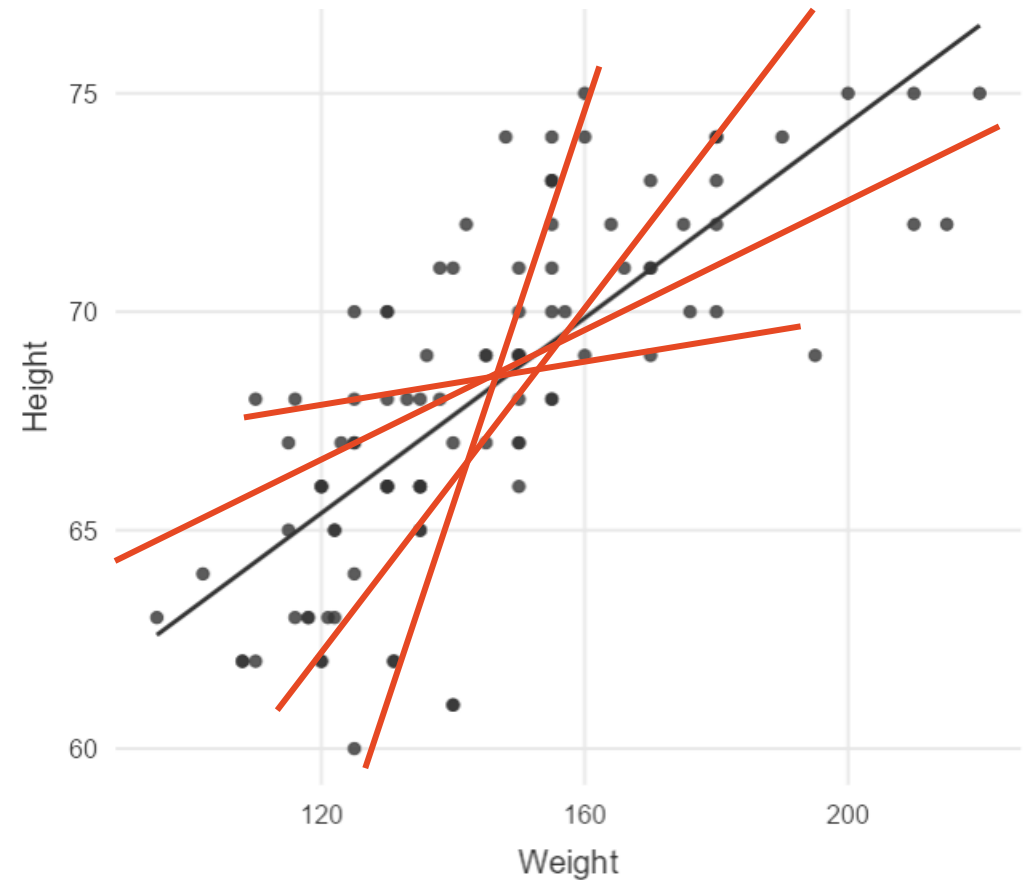
## Error:

- The part of the data left unexplained by my model.
- An individual error is also called a **residual**.
  - $e_i = Y_i - Y'_i$
  - The difference between a participant's actual score and the score predicted for them by the model.



# What makes a good model?

- There are many possible models that we *could* fit to the data.
- How do I choose the *best* model?
  - The one that does best at *minimizing error*.
  - This depends on how we choose to calculate *total error*.



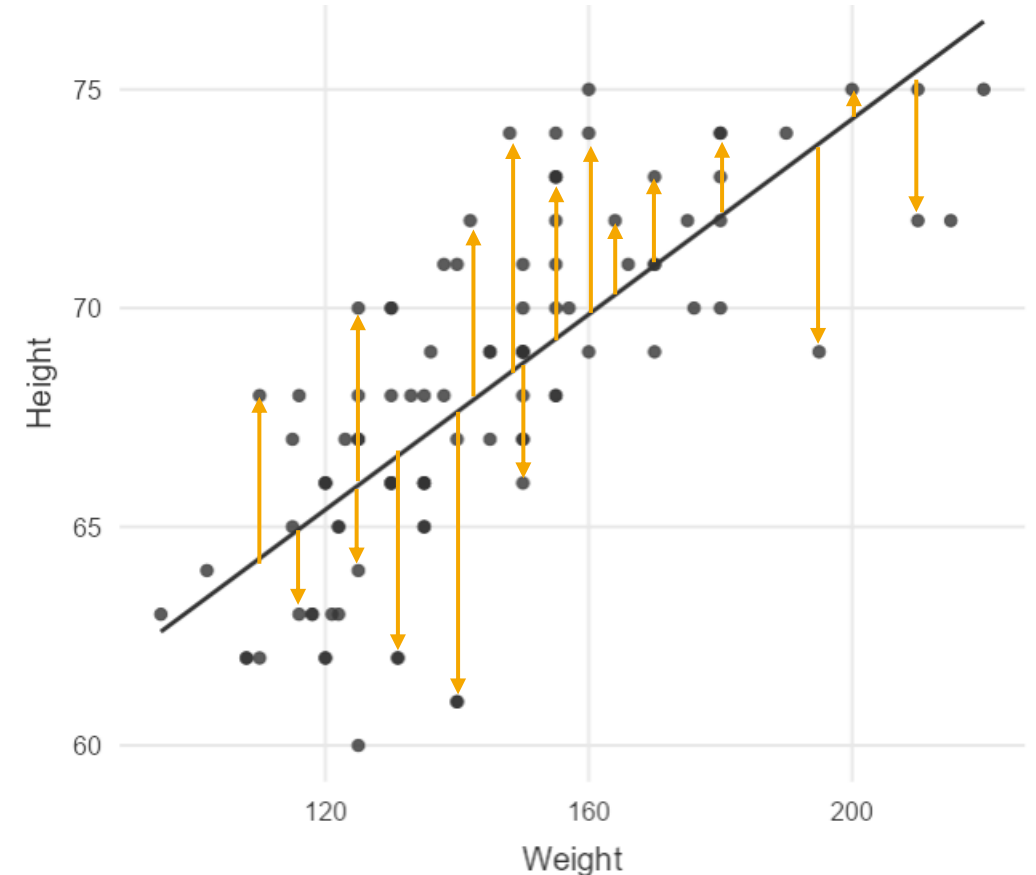
# Ways of Calculating Total Error

- Count of errors
  - Did an error occur? Yes/No
  - Cons: Limited amount of information
- Sum of absolute errors
  - Take the absolute value of each residual and then sum them
  - Cons: Small and large errors contribute the *same* amount to total error



## Sum of squared errors

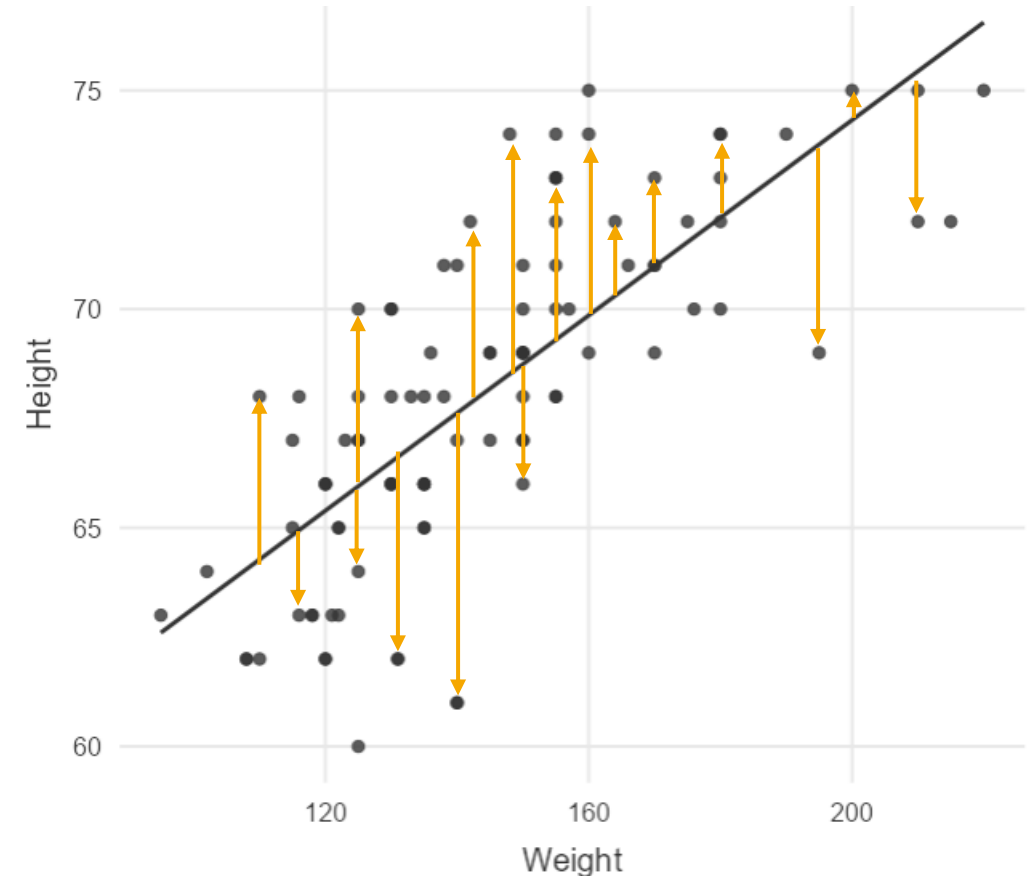
- Square each residual and then sum them
- Cons: Highly affected by outliers





# The Best-Fitting Model

- The best-fitting model is the one that **minimizes the sum of squared errors (SSE)**.
- Or in other words, the model that does best at minimizing *the sum of the squared distances between each participant's actual score and the score predicted by the linear model*.



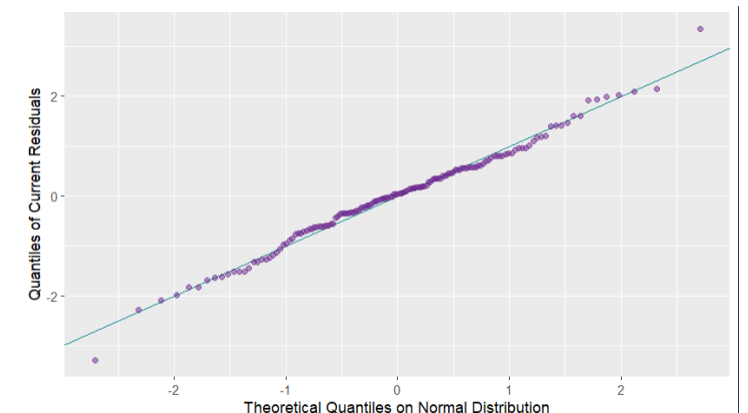
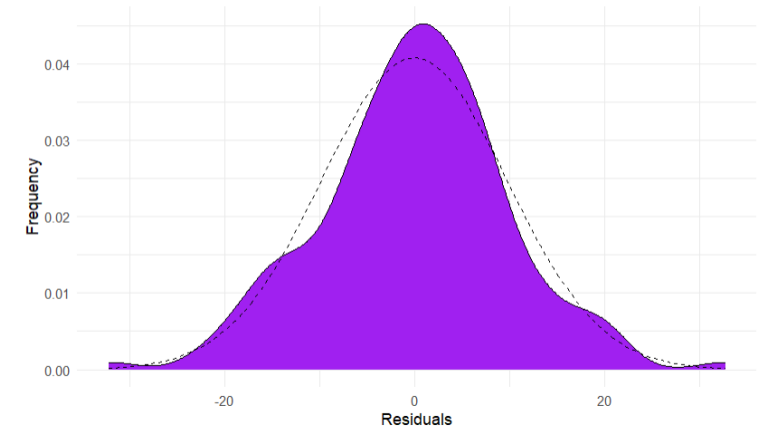
# Underlying Assumptions about Errors

## 1. Errors are normally distributed

- This assumption is robust to violations

### ○ Ways of assessing:

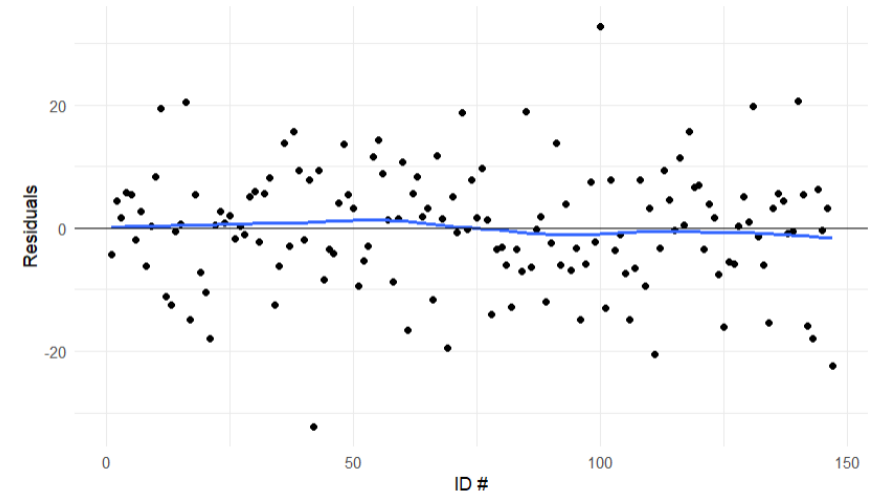
- Distribution of residuals
  - Do the residuals look approximately normally distributed?
- Q-Q plot
  - Are the points close to lying on the diagonal reference line?



# Underlying Assumptions about Errors

## 2. Independence of errors

- Each participant's error is unrelated to every other participant's error
- This assumption is **not** robust to violations
- If you violate this assumption, using a test that assumes independence of errors would give biased results
  - If violated, use an analysis intended for within-subjects, or nested, data instead
- Ways of assessing:
  - Ensure random sample during data collection
  - Plot residuals against a variable they should *not* be associated with (e.g., ID #)
    - Is there a pattern in the residuals that shouldn't be there?



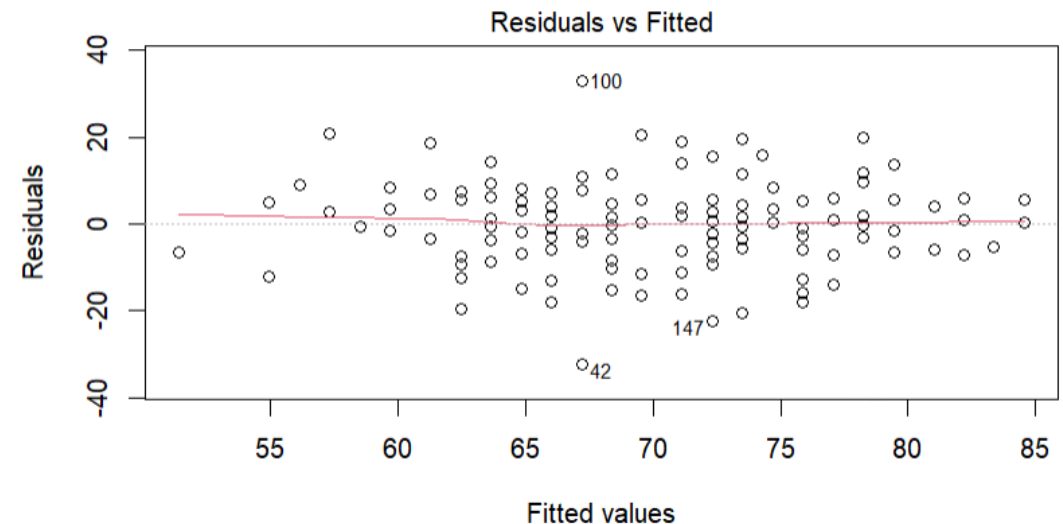
# Underlying Assumptions about Errors

## 3. Homogeneity of variance

- How spread out the residuals are is approximately the same across all points of the model
  - Also called **homoscedasticity**
- This assumption is fairly robust to violations
  - Violating this assumption means there is **heteroscedasticity**

### ○ Ways of assessing:

- A residuals plot (fitted values vs residuals)
  - Is the amount of scatter in the residuals approximately the same across the range of fitted values?



# The General Equation for a Linear Model

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_3 X_{i3} + \dots + \beta_j X_{ij} + \varepsilon_i$$

- $Y_i$  = Participant  $i$ 's score on the outcome variable
- $\beta$ s are population-level model parameters that must be estimated from the data
  - $\beta_0$  = the model intercept
  - $\beta_1$  = the slope contribution corresponding to  $X_1$
  - $\beta_2$  = the slope contribution corresponding to  $X_2$
  - $\beta_j$  = the slope contribution corresponding to  $X_j$
- Xs are predictor variables included in the model
  - $X_{i1}$  = Participant  $i$ 's score on predictor variable 1
  - $X_{i2}$  = Participant  $i$ 's score on predictor variable 2
  - $X_{ij}$  = Participant  $i$ 's score on predictor variable  $j$
- $\varepsilon_i$  = The error (or residual) for participant  $i$

# Estimate of the Linear Model

$$Y_i = b_0 + b_1X_{i1} + b_2X_{i2} + b_3X_{i3} + \dots + b_jX_{ij} + e_i$$

- $Y_i$  = Participant  $i$ 's score on the outcome variable
- $bs$  are estimates of the model parameters based on the data
  - $b_0$  = the model intercept
  - $b_1$  = estimate of the slope contribution corresponding to  $X_1$
  - $b_2$  = estimate of the slope contribution corresponding to  $X_2$
  - $b_j$  = estimate of the slope contribution corresponding to  $X_j$
- $Xs$  are predictor variables included in the model
  - $X_{i1}$  = Participant  $i$ 's score on predictor variable 1
  - $X_{i2}$  = Participant  $i$ 's score on predictor variable 2
  - $X_{ij}$  = Participant  $i$ 's score on predictor variable  $j$
- $e_i$  = Estimate of the error (or residual) for participant  $i$

# Zero Parameter Model (aka, The Simplest Model)

$$Y_i = B_0 + \varepsilon_i$$

- $B_0$  is an *a priori value* that we expect to match the data based on previous research
- It is *not* a model parameter that is estimated from the data
  - This is why this model is called the **Zero Parameter Model**

# One Parameter Model (aka, The Simple Model)

$$Y_i = \beta_0 + \varepsilon_i$$

- The one parameter model has a single parameter that is used to predict people's actual scores
  - i.e., The one parameter model uses *a single numerical value* to predict the data
- This single parameter is estimated from the data
  - Remember the criteria for choosing the best model: minimizing SSE
    - Q: What single numerical value do you think does best at minimizing SSE in a set of data?
      - The sample mean



# Using Model Comparisons to Test Hypotheses

- The GLM approach uses **model comparisons** to test hypotheses
- Fit a model representing the **null hypothesis** and a model representing the **alternative hypothesis**
- Which model fits the empirical data better?
  - i.e., Which model does best at minimizing the SSE?

# One Parameter vs Zero Parameter Model



# One- vs Zero-Parameter Model Comparison

**Example:** NASA's astrobiology department is interested in understanding the biological properties of material found on other astronomical objects. Let's say an astrobiologist is studying the biological properties of 20 water samples that were obtained from nearby asteroids. The researcher is interested in whether the properties of the water samples obtained from these asteroids mimic the properties of water found on Earth. Water on Earth freezes at  $0^{\circ}\text{C}$ . The astrobiologist tests the temperature at which her 20 asteroid water samples freeze to compare to this Earth standard.



# Specify the Models

The model describing the **null hypothesis** is called Model C:

$$\text{Model C: } Y_i = 0 + \varepsilon_i$$

The model describing the **alternative hypothesis** is called Model A:

$$\text{Model A: } Y_i = \beta_0 + \varepsilon_i$$

# The Model Comparison

## Model Comparison:

$$\text{Model C: } Y_i = 0 + \varepsilon_i \quad \text{PC} = 0$$

$$\text{Model A: } Y_i = \beta_0 + \varepsilon_i \quad \text{PA} = 1$$

## Null & Alternative Hypotheses:

$$H_0: \beta_0 = 0$$

$$H_1: \beta_0 \neq 0$$

# State your Decision Rule

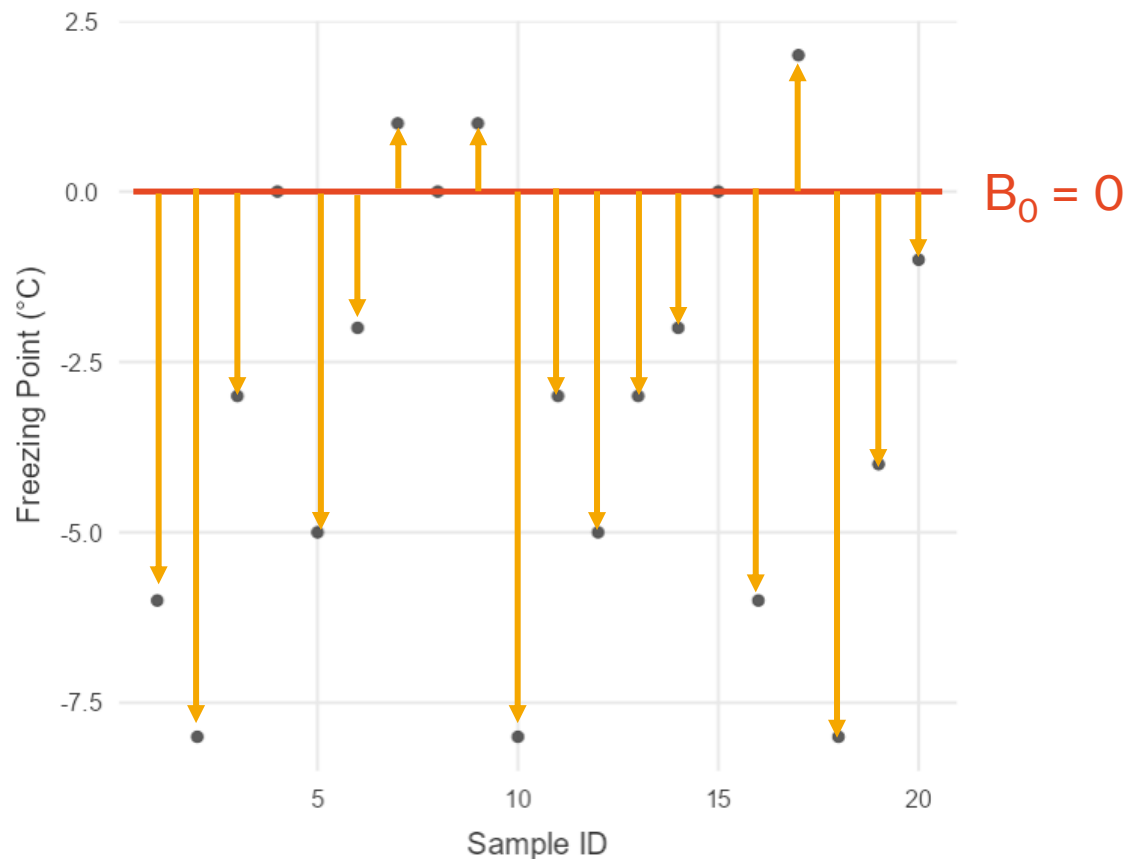
- Willingness to make a Type I error = 5%
  - $\alpha = .05$

# The Collected Data

- **Data:**
  - Here are the freezing points for the 20 asteroid water samples the astrobiologist measured.
- Which model does better at predicting these real-world empirical observations?

Sample ID	Freezing Point (°C)
1	-6
2	-8
3	-3
4	0
5	-5
6	-2
7	1
8	0
9	1
10	-8
11	-3
12	-5
13	-3
14	-2
15	0
16	-6
17	2
18	-8
19	-4
20	-1

# Fit Model C to the Data



- Model C:  $Y_i = 0 + \varepsilon_i$ 
  - How well does a predicted freezing point of 0 fit the data?
- Calculate the SSE
  - Calculate each residual:
    - $e_i = Y_i - Y'_i$
  - Square each residual:
    - $(e_i)^2 = (Y_i - Y'_i)^2$
  - Sum the squared residuals
    - $SSE = \sum (Y_i - Y'_i)^2$



# Evaluate Fit of Model C

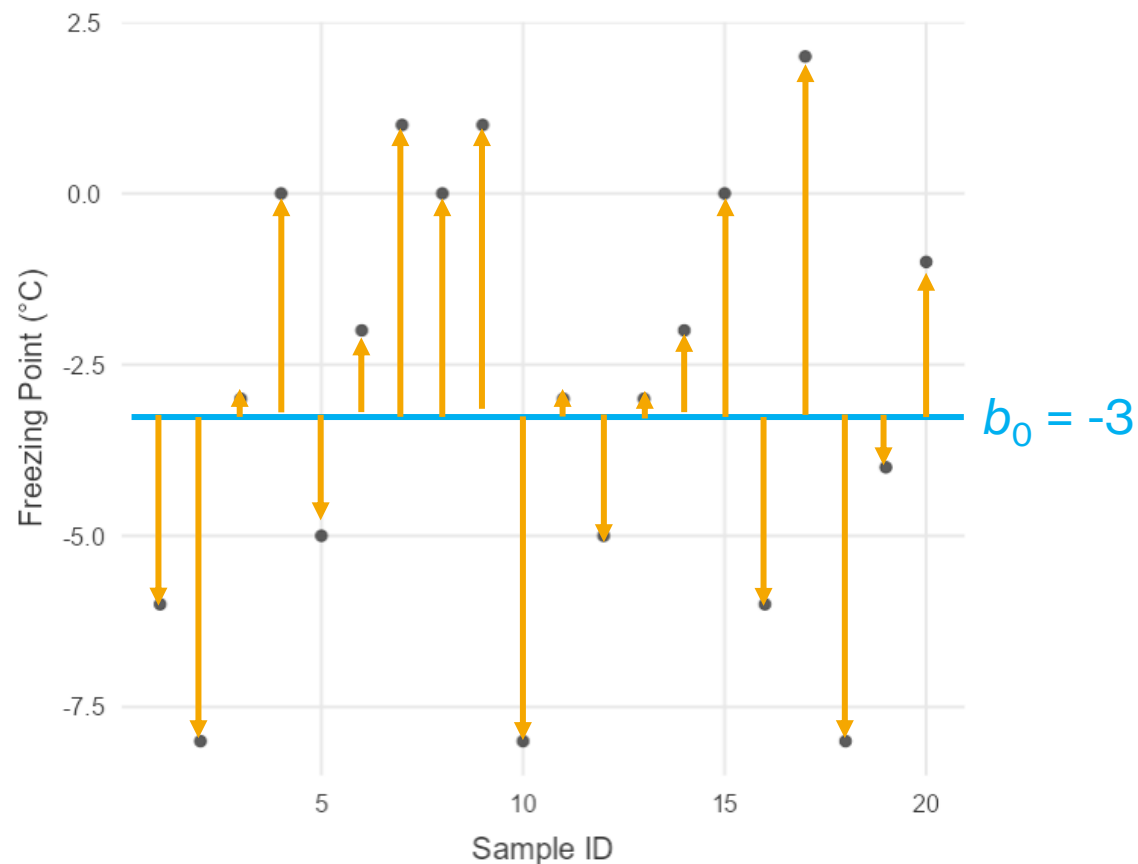
Sample ID	$Y_i$ Freezing Point (°C)	$Y'$ (Value Predicted by Model)	$e_i = Y_i - Y'_i$	$(e_i)^2 = (Y_i - Y'_i)^2$
1	-6	0	-6	36
2	-8	0	-8	64
3	-3	0	-3	9
4	0	0	0	0
5	-5	0	-5	25
6	-2	0	-2	4
7	1	0	1	1
8	0	0	0	0
9	1	0	1	1
10	-8	0	-8	64
11	-3	0	-3	9
12	-5	0	-5	25
13	-3	0	-3	9
14	-2	0	-2	4
15	0	0	0	0
16	-6	0	-6	36
17	2	0	2	4
18	-8	0	-8	64
19	-4	0	-4	16
20	-1	0	-1	1

$$SSE = \sum (Y_i - Y'_i)^2$$

SSE for Model C:

- $SSE(C) = 372$

# Fit Model A to the Data



Model A:  $Y_i = \beta_0 + \varepsilon_i$

- Estimate  $\beta_0$  from the data
  - $Y_i = b_0 + \varepsilon_i$ 
    - $b_0$  is our estimate of  $\beta_0$
    - When using a single numerical value, the mean does best at minimizing SSE
- $Y_i = -3.00 + \varepsilon_i$
- Calculate the SSE
  - $SSE = \sum (Y_i - Y'_i)^2$

# Evaluate Fit of Model A

Sample ID	$Y_i$ Freezing Point (°C)	$Y'$ (Value Predicted by Model)	$e_i = Y_i - Y'_i$	$(e_i)^2 = (Y_i - Y'_i)^2$
1	-6	-3	-3	9
2	-8	-3	-5	25
3	-3	-3	0	0
4	0	-3	3	9
5	-5	-3	-2	4
6	-2	-3	1	1
7	1	-3	4	16
8	0	-3	3	9
9	1	-3	4	16
10	-8	-3	-5	25
11	-3	-3	0	0
12	-5	-3	-2	4
13	-3	-3	0	0
14	-2	-3	1	1
15	0	-3	3	9
16	-6	-3	-3	9
17	2	-3	5	25
18	-8	-3	-5	25
19	-4	-3	-1	1
20	-1	-3	2	4

$$SSE = \sum (Y_i - Y'_i)^2$$

SSE for Model A:

- $SSE(A) = 192$

# Compare Model C and Model A's Fits

Sum of Squares Reduced (SSR):

- $SSR = SSE(C) - SSE(A) = 372 - 192 = 180$

Proportional Reduction in Error (PRE):

- $PRE = \frac{SSE(C) - SSE(A)}{SSE(C)} = \frac{SSR}{SSE(C)} = \frac{(372 - 192)}{372} = \frac{180}{372} = 0.48$

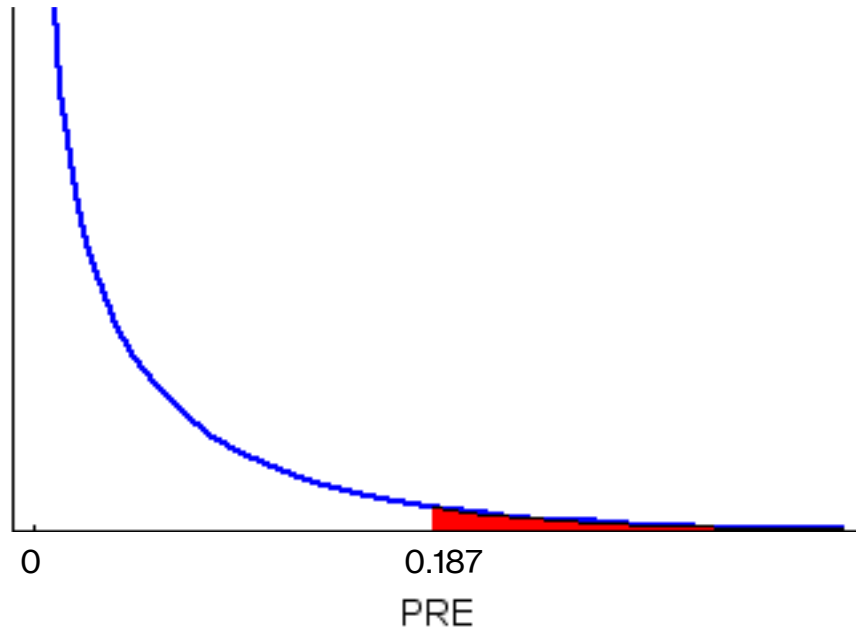
- Model A accounts for 48% more error compared to Model C.

# Sampling Error

- The same phenomenon of sampling error occurs here with PRE
  - PRE is an estimate of the true proportional reduction in error, called  $\eta^2$ 
    - Due to sampling error, PRE values calculated using sample data can vary around  $\eta^2$
- Is  $PRE = 0.48$  large enough to be unlikely due to chance when the null hypothesis is true?
  - To decide, construct a sampling distribution representing the results (i.e., the PRE-values) you would expect to obtain *if the null hypothesis is true*, and then
  - Determine the probability ( $p$ -value) of your results on this sampling distribution

# Sampling Distribution of PRE

## Sampling Distribution of PRE



- Remember that a **sampling distribution** is a tool that's used to represent the results one would expect to obtain *if the null hypothesis is true*
- The sampling distribution of PRE can be used to represent the PRE values one would expect to obtain *if the null hypothesis is true*

$$PRE = \frac{SSE(C) - SSE(A)}{SSE(C)}$$

- What value would you expect most often if the null hypothesis is true?

# Sampling Distribution of PRE

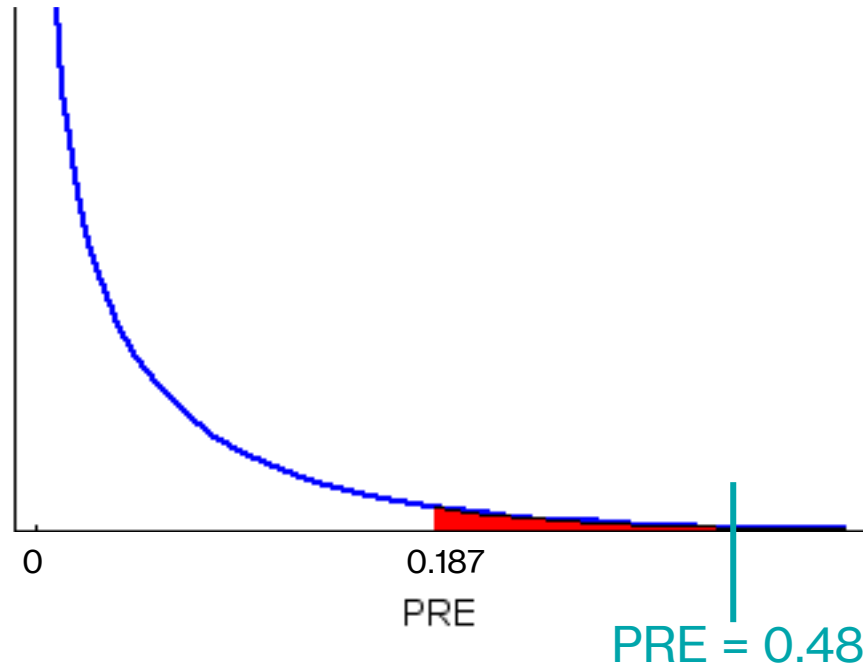
FIGURE A.1 Critical values of PRE for  $PA - PC = 1$

$n - PA$	$\alpha$							
	.5	.2	.1	.05	.025	.01	.005	.001
1	.500	.905	.976	.994	.998	1.000	1.000	1.000
2	.250	.640	.810	.902	.951	.980	.990	.998
3	.163	.472	.649	.771	.853	.919	.949	.982
4	.121	.370	.532	.658	.753	.841	.887	.949
5	.096	.303	.448	.569	.667	.765	.820	.904
6	.079	.257	.386	.499	.595	.696	.756	.855
7	.067	.222	.339	.444	.536	.636	.699	.807
8	.059	.196	.302	.399	.486	.585	.647	.761
9	.052	.175	.272	.362	.445	.540	.602	.717
10	.047	.158	.247	.332	.410	.501	.562	.678
11	.042	.145	.227	.306	.379	.467	.526	.642
12	.039	.133	.209	.283	.353	.437	.495	.608
13	.036	.123	.194	.264	.330	.411	.467	.578
14	.033	.114	.181	.247	.310	.388	.441	.550
15	.031	.107	.170	.232	.292	.367	.419	.525
16	.029	.100	.160	.219	.277	.348	.398	.502
17	.027	.095	.151	.208	.262	.331	.379	.480
18	.026	.090	.143	.197	.249	.315	.362	.461
19	.024	.085	.136	.187	.238	.301	.346	.443
20	.023	.081	.129	.179	.227	.288	.332	.426
22	.021	.074	.118	.164	.208	.265	.307	.395
24	.019	.067	.109	.151	.192	.246	.285	.369
26	.018	.062	.101	.140	.179	.229	.266	.346
28	.016	.058	.094	.130	.167	.214	.249	.325
30	.015	.054	.088	.122	.157	.201	.234	.307
35	.013	.046	.075	.105	.135	.175	.204	.269
40	.011	.041	.066	.093	.119	.155	.181	.240
45	.010	.036	.059	.083	.107	.138	.162	.216
50	.009	.033	.053	.075	.096	.125	.147	.196
55	.008	.030	.048	.068	.088	.115	.135	.180
60	.008	.027	.044	.063	.081	.106	.124	.166
80	.006	.020	.033	.047	.061	.080	.094	.127
100	.005	.016	.027	.038	.049	.065	.076	.103
150	.003	.011	.018	.025	.033	.043	.051	.070
200	.002	.008	.013	.019	.025	.033	.039	.053
500	.001	.003	.005	.008	.010	.013	.016	.021

- Remember we chose  $\alpha = .05$ 
  - Determines the size of the “rejection region”
  - Can find the critical value corresponding to alpha level using a PRE-critical value table
    - Appendix of your book
- For our example:
  - $PA-PC = (1 - 0) = 1$ ,  $n-PA = (20 - 1) = 19$ ,  $\alpha = .05$
  - $PRE_{CV} = .187$

# Sampling Distribution of PRE

## Sampling Distribution of PRE



- Remember we chose  $\alpha = .05$ 
  - Determines the size of the “rejection region”
  - Can find the critical value corresponding to alpha level using a PRE-critical value table
    - Appendix of your book
- For our example:
  - $PA-PC = (1 - 0) = 1$ ,  $n-PA = (20 - 1) = 19$ ,  $\alpha = .05$
  - $PRE_{CV} = .187$

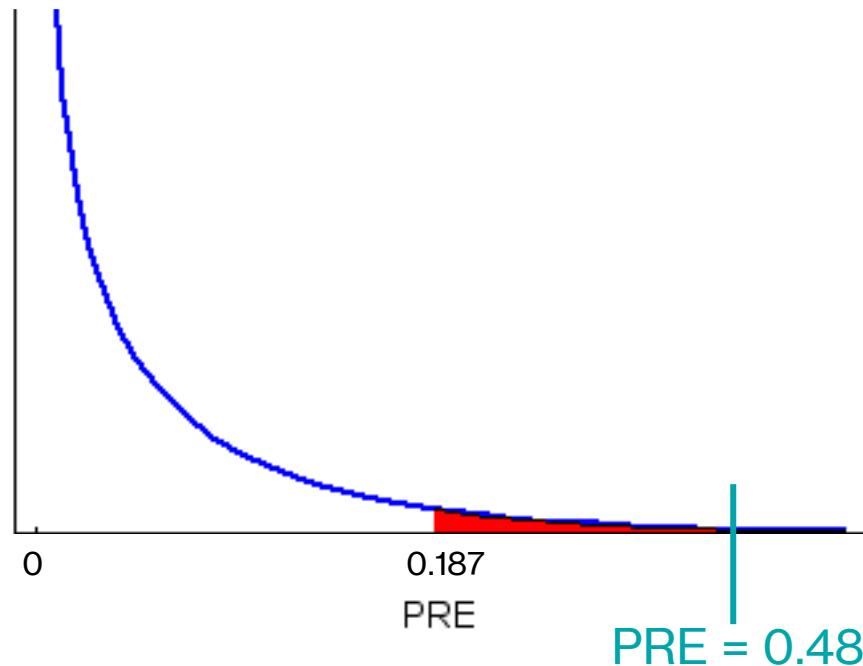
Deciding whether results are significant:

- If  $PRE > PRE_{CV} \rightarrow$  **Reject  $H_0$  (Results are significant)**
- If  $PRE < PRE_{CV} \rightarrow$  Fail to reject  $H_0$  (Results are non-significant)



# Sampling Distribution of PRE

## Sampling Distribution of PRE



- Drawbacks of the Sampling Distribution of PRE:
  - It is uncommon to use this type of sampling distribution
    - The  $p$ -values you obtain using R most often correspond to more frequently reported test statistics, like  $F$  and  $t$
  - In order to determine whether one's results are significant, or non-significant, using a Sampling Distribution of PRE...
    - You would need to carry your book around and use the PRE-critical value table in the back of it
- Instead, let's talk about a more commonly used sampling distribution

# F-Statistic

- $F$ -statistics are much more commonly reported test statistics in statistical software's outputs
- $p$ -values in R often correspond to where your results landed on a sampling distribution of the  $F$ -statistics one would expect to obtain if the null hypothesis is true
  - (or the closely related  $t$ -distribution which, when  $PA-PC = 1$ , is just the square root of the  $F$ -statistic)
- The  $F$ -statistic for one's results can be easily obtained using PRE and/or the SSE already calculated

# F-Statistic

$$F = \frac{\frac{SSR}{(PA - PC)}}{\frac{SSE(A)}{(n - PA)}} = \frac{\frac{PRE}{(PA - PC)}}{\frac{(1 - PRE)}{(n - PA)}}$$

## Numerator of the *F*-statistic:

- Variance in the dependent variable explained by the model
  - Numerator df:
    - $df_{\text{Reduced}} = PA - PC$

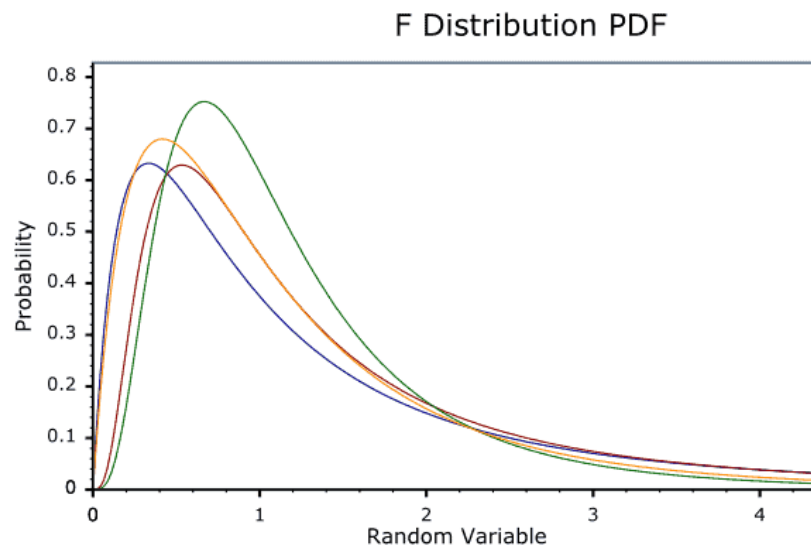
## Denominator of the *F*-statistic:

- Variance in the dependent variable left unexplained by the model
  - Denominator df:
    - $df_{\text{Error}} = n - PA$

# Sampling Distributions of F-statistics

## Sampling Distribution of F-statistics

- The shape varies depending on the numerator (df = PA-PC) and denominator degrees of freedom (df =  $n$ -PA)



- Let's calculate the  $F$ -statistic for our example and get a corresponding  $p$ -value

$$F = \frac{\frac{SSR}{(PA - PC)}}{\frac{SSE(A)}{(n - PA)}} = \frac{\frac{180}{(1 - 0)}}{\frac{192}{(20 - 1)}}$$

$$F = 17.81$$

# ANOVA Summary Table

MS = “Mean Squared”

- A special term for variance
- $MS = \frac{SS}{df}$

Source	SS	df	MS	$F$	PRE	$p$
Reduced	SSR	PA-PC	$MSR = \frac{SSR}{PA - PC}$	$F = \frac{MSR}{MSE}$	$PRE = \frac{SSR}{SSE(C)}$	
Model A (Error)	SSE(A)	n-PA	$MSE = \frac{SSE(A)}{n - PA}$			
Model C (Total)	SSE(C)	n-PC				

# ANOVA Summary Table

Source	SS	df	MS	$F$	PRE	$p$
Reduced	180	1	180	17.81	0.48	?
Model A (Error)	192	19	10.1053			
Model C (Total)	372	20				

To obtain the  $p$ -value corresponding to our  $F$ -statistic, we will run the analysis in R.

# Obtaining the *p*-value

- Import the data
  - Stored it in an object called `temps`
  - Check out the first few rows
- Fit the model
  - The `lm` function in R is used to conduct a regression analysis
    - `lm(DV ~ X1 + X2 + ... + Xp, data = data)`
  - Here, I'm using what's called an "intercept-only model"
  - In `lm`, when you predict a DV from 1, it finds the single numerical value that best reduces SSE in the data

```
temps <- import("freezing_temps.xlsx")
```

```
> head(temps)
  freezing_temp
1            -6
2            -8
3             -3
4             0
5            -5
```

```
model <- lm(freezing_temp ~ 1, data = temps)
```

# Obtaining the *p*-value

- Look at the model output using
  - ``summary(model)``, and
  - ``Anova(model)`` from the car package

```
> summary(model)

Call:
lm(formula = freezing_temp ~ 1, data = temps)

Residuals:
    Min       1Q   Median       3Q      Max
-5.00  -2.25   0.00   3.00   5.00

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.0000     0.7108  -4.22 0.000463 ***
---
```

```
> Anova(model, type = 3) # car package
Anova Table (Type III tests)

Response: freezing_temp
              Sum Sq Df F value    Pr(>F)
(Intercept)    180   1  17.812 0.0004632 ***
Residuals     192  19
---
```



# Obtaining the $p$ -value

- `summary()` output:

- Estimate:
  - The current model only has a single parameter being estimated ( $\beta_0$ )
    - $b_0 = -3.00$

- Std. Error:

- The standard error is  $\sqrt{\frac{MSE}{n}} = \sqrt{\frac{10.1053}{20}} = .7108$

- $t$ -value:

- $t = \frac{b_0}{Std.Error} = \frac{-3}{.7108} = -4.22$

- $Pr(>|t|)$

- $p < .001$ , The probability of obtaining this  $t$ -statistic when the null hypothesis is true is less than .05

```
> summary(model)

Call:
lm(formula = freezing_temp ~ 1, data = temps)

Residuals:
    Min       1Q   Median       3Q      Max
 -5.00  -2.25   0.00   3.00   5.00

Coefficients:
(Intercept)  -3.0000    0.7108   -4.22  0.000463 ***
---

```

```
> Anova(model, type = 3) # car package
Anova Table (Type III tests)

Response: freezing_temp
          Sum Sq Df F value    Pr(>F)
(Intercept)    180  1  17.812 0.0004632 ***
Residuals      192 19
---

```

# Obtaining the $p$ -value

- Anova() output:
  - $SSR = 180$ 
    - $df_{\text{Reduced}} = PA - PC = 1$
  - $SSE(A) = 192$ 
    - $df_{\text{Error}} = n - PA = 19$
  - $F = 17.81$
  - $\text{Pr}( > F )$ 
    - $p < .001$
    - Reject the null hypothesis because  $p < .05$

```
> summary(model)

Call:
lm(formula = freezing_temp ~ 1, data = temps)

Residuals:
    Min       1Q   Median       3Q      Max
-5.00  -2.25   0.00   3.00   5.00

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -3.0000     0.7108  -4.22 0.000463 ***
---

```

```
> Anova(model, type = 3) # car package
Anova Table (Type III tests)

Response: freezing_temp

```

	Sum Sq	Df	F value	Pr(>F)	
(Intercept)	180	1	17.812	0.0004632	***
Residuals	192	19			

```
---
```

# Effect Size

- PRE is a standardized measure of effect size
  - The proportional reduction in error is unitless so can be compared across studies

---

$\eta^2$   
(Cohen, 1977)

---

.02 = small

.13 = medium

.26 = large

---

# 95% Confidence Interval

- Remember that point estimates are unlikely to be exactly equal to the population parameter they are attempting to estimate
- $b_0$  is a point estimate of  $\beta_0$ 
  - Can construct a 95%CI to get a range of values within which we are more confident the true value of  $\beta_0$  lies

$$Y_i = \beta_0 + \varepsilon_i$$



$$Y_i = b_0 + e_i$$

# 95% Confidence Interval around $b_0$

$$95\%CI = b_0 \pm \sqrt{\frac{F_{CV} * MSE}{n}}$$

- $b_0$  is our point estimate
- $F_{CV}$  is the  $F$ -critical value corresponding to 1 and  $n-1$  df and an alpha of .05
- MSE is the mean squared error,  $MSE = \frac{SSE(A)}{n - PA}$
- $n$  is the sample size

$$95\%CI = b_0 \pm (t_{CV} * \sqrt{\frac{MSE}{n}})$$

- This formula should look familiar to the one used to calculate 95%CI around a sample mean
  - $t$  is a special case of  $F$  when  $df_{\text{Numerator}} = PA - PC = 1$ 
    - $t = \sqrt{F}$
- $t_{CV}$  is the  $t$ -critical value corresponding to  $n-1$  df and an alpha of .05

# 95% Confidence Interval

$$95\%CI = b_0 \pm \sqrt{\frac{F_{1,n-1;\alpha}MSE}{n}}$$

- $b_0 = -3.00$
- $F_{CV}(1, 19) = 4.381$
- $MSE = \frac{SSE(A)}{n - PA} = \frac{192}{19} = 10.1053$
- $n = 20$

- Lower bound =  $-3 - \sqrt{\frac{4.381 \cdot 10.1053}{20}} = -4.49$
- Upper bound =  $-3 + \sqrt{\frac{4.381 \cdot 10.1053}{20}} = -1.51$

The `confint()` function can be used to obtain 95% CIs around each of the parameter estimates in one's model:

```
> confint(model)
                2.5 %      97.5 %
(Intercept) -4.487761 -1.512239
```

APA Style: 95%CI[-4.49, -1.51]

# Summary of Findings

- Summary of findings:
  - The average freezing point of the asteroid water ( $M = -3.00$ ,  $SD = 3.18$ ) was significantly lower than zero,  $F(1, 19) = 17.81$ ,  $p < .001$ ,  $PRE = 0.48$ ,  $b_0 = -3.00$ ,  $95\%CI[-4.49, -1.51]$ .



# Using Model Comparisons to Test Hypotheses

- ❑ Develop a **theory** by performing a literature review in your area of study
- ❑ State your **hypotheses**
- ❑ Specify the regression **models** corresponding to each of your hypotheses
- ❑ State your **decision rule**
- ❑ **Fit** both models to the data and then **evaluate** which model fits the real-world observations better
- ❑ **Interpret** the results & consider the **effect size**