



UNIVERSITY *of*  
WEST FLORIDA

## Model Accuracy

Statistics for Data Science II

Recall the linear model,

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon$$

Even if we know the true parameters ( $\beta_0$ ,  $\beta_1$ , etc.), the linear model will not perfectly fit the data.

Thus, when constructing a model, we want to measure the model fit.

We can assess this with two quantities:

Residual standard error (RSE)

$R^2$  statistic

---

Recall the residual,

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

This measures the distance between the observed value ( $y$ ) and the predicted value ( $\hat{y}$ ) for every observation.

We can also calculate the residual sum of squares (RSS),

$$\text{RSS} = \sum_{i=1}^n e_i^2$$

# Residual Standard Error

Then, we can calculate the residual standard error (RSE),

$$\text{RSE} = \sqrt{\frac{1}{n-2} \text{RSS}} = \sqrt{\frac{1}{n-2} \sum_{i=1}^n e_i^2}$$

The RSE estimates the standard deviation of  $\varepsilon$ .

We can actually pull this from the ANOVA table:

$$\text{RSE} = \sqrt{\text{MSE}}$$

## Example:

Recall the models from the previous lecture,

$$M1:\hat{y} = 13.04 + 0.14\text{flipper} - 7.44\text{Chinstrap} - 33.52\text{Gentoo} + 0.09(\text{flipper} \times \text{Chinstrap}) + 0.18(\text{flipper} \times \text{Gentoo})$$

$$M2:\hat{y} = 37.26 + 3.13\text{male} + 9.32\text{Chinstrap} + 8.31\text{Gentoo} + 1.39\text{male} \times \text{Chinstrap} + 0.78\text{male} \times \text{Gentoo}$$

$$M3:\hat{y} = -26.75 + 0.34\text{flipper} + 0.01\text{body mass} - 0.00003(\text{flipper} \times \text{body mass})$$

Let's find the RSE for each model.

We can use either `summary()` or `anova()`

## Example: M1

```
##
## Call:
## lm(formula = bill_length_mm ~ flipper_length_mm + species_Chinstrap +
##     species_Gentoo + species_Chinstrap:flipper_length_mm + species_Gentoo:flipper_length_mm,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.698 -1.703  0.036   1.566 12.439
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      13.0358     6.1840   2.11 0.03579 *
## flipper_length_mm    0.1357     0.0325   4.17 0.000039 ***
## species_Chinstrap   -7.4424    10.5683  -0.70 0.48180
## species_Gentoo     -33.5237     9.9202  -3.38 0.00081 ***
## flipper_length_mm:species_Chinstrap  0.0852     0.0545   1.56 0.11909
## flipper_length_mm:species_Gentoo    0.1776     0.0483   3.68 0.00027 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.55 on 327 degrees of freedom
## Multiple R-squared:  0.785, Adjusted R-squared:  0.782
## F-statistic: 239 on 5 and 327 DF, p-value: <2e-16
```

## Example: M1

```
anova(m1)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: bill_length_mm
```

```
##
```

```
## flipper_length_mm
```

```
## species_Chinstrap
```

```
## species_Gentoo
```

```
## flipper_length_mm:species_Chinstrap
```

```
## flipper_length_mm:species_Gentoo
```

```
## Residuals
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
flipper_length_mm	1	4235	4235	649.67	< 2e-16	***
species_Chinstrap	1	3350	3350	513.88	< 2e-16	***
species_Gentoo	1	124	124	19.05	0.000017	***
flipper_length_mm:species_Chinstrap	1	0	0	0.01	0.92647	
flipper_length_mm:species_Gentoo	1	88	88	13.54	0.00027	***
Residuals	327	2132	7			

Notice that  $RSE = \sqrt{7} = 2.65 \neq 2.55$

## Example: M1

Beware that sometimes R rounds things displayed – we may want to save things as a tibble (or data frame) to see the true values.

```
print.data.frame(round(anova(m1), 4))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
## flipper_length_mm	1	4235.0137	4235.0137	649.6737	0.0000
## species_Chinstrap	1	3349.7977	3349.7977	513.8768	0.0000
## species_Gentoo	1	124.1841	124.1841	19.0505	0.0000
## flipper_length_mm:species_Chinstrap	1	0.0556	0.0556	0.0085	0.9265
## flipper_length_mm:species_Gentoo	1	88.2438	88.2438	13.5371	0.0003
## Residuals	327	2131.6078	6.5187	NA	NA

Thus, we see that the MSE is not actually 7. Instead, it is 6.5187.

Now, we match the output from `summary()`:  $\sqrt{6.5187} = 2.55$ .



## Example: M2

```
##  
## Call:  
## lm(formula = bill_length_mm ~ sex_male + species_Chinstrap +  
##       species_Gentoo + species_Chinstrap:sex_male + species_Gentoo:sex_male,  
##       data = data)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -5.790 -1.374 -0.064   1.210 11.426   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)      37.258     0.271  137.47 < 2e-16 ***  
## sex_male          3.133     0.383   8.17 6.6e-15 ***  
## species_Chinstrap  9.316     0.481  19.38 < 2e-16 ***  
## species_Gentoo     8.306     0.407  20.39 < 2e-16 ***  
## sex_male:species_Chinstrap 1.388     0.680   2.04  0.042 *   
## sex_male:species_Gentoo   0.777     0.572   1.36  0.175   
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 2.32 on 327 degrees of freedom  
## Multiple R-squared:  0.823, Adjusted R-squared:  0.821   
## F-statistic: 305 on 5 and 327 DF, p-value: <2e-16
```

## Example: M3

```
##
## Call:
## lm(formula = bill_length_mm ~ flipper_length_mm + body_mass_g +
##     flipper_length_mm:body_mass_g, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.106 -2.746 -0.786  2.158 18.955
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -2.68e+01   2.11e+01   -1.26   0.2068
## flipper_length_mm    3.39e-01   1.07e-01    3.17   0.0016 **
## body_mass_g      5.86e-03   4.84e-03    1.21   0.2270
## flipper_length_mm:body_mass_g -2.59e-05   2.34e-05   -1.10   0.2701
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.15 on 329 degrees of freedom
## Multiple R-squared:  0.43,    Adjusted R-squared:  0.425
## F-statistic: 82.8 on 3 and 329 DF,  p-value: <2e-16
```

## Example:

Comparing the three models,

Model	Predictors	RSE
1	species, flipper length	2.55
2	species, sex	2.32
3	flipper length, body mass	4.15

In theory, the model with the smallest RSE is the “best fitting” ... but how different is 2.32 and 2.55?

The downside to using the RSE is that it takes on the units of  $Y$ .

This means we cannot define a “good” RSE.

We now turn to a proportion, the  $R^2$  statistic.

$R^2$  is the proportion of variance explained by the model.

Because it is a proportion,  $R^2 \in [0, 1]$  and is independent of the units of  $Y$ .

If  $R^2 \rightarrow 0$ , the model does not fit the data well; if  $R^2 \rightarrow 1$ , the model fits the data well.

Note that if  $R^2 = 1$ , all observations fall on the response surface.

---

To calculate  $R^2$ , we need RSS as well as the total sum of squares (TSS),

$$\text{RSS} = \sum_{i=1}^n e_i^2 \quad \text{and} \quad \text{TSS} = \sum_{i=1}^n (y_i - \bar{y})^2$$

Then,

$$R^2 = \frac{\text{TSS} - \text{RSS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

Note that other places may have different language (using the ANOVA table),

$$R^2 = \frac{\text{SSTot} - \text{SSE}}{\text{SSTot}} = 1 - \frac{\text{SSReg}}{\text{SSTot}}$$

Remember that we are partitioning the variability in  $Y$  (TSS, or SSTot) into two pieces:

The variability explained by the regression model (TSS - RSS, or SSReg) and the variability explained by outside sources (RSS, or SSE).

As predictors are added to the model, we necessarily increase SSReg / decrease SSE.

We want a measure of model fit that is resistant to this fluxuation.

$$R_{\text{adj}}^2 = 1 - \left( \frac{n-1}{n-k-1} \right) (1 - R^2)$$

## Example:

Recall the models from the previous lecture,

$$M1:\hat{y} = 13.04 + 0.14\text{flipper} - 7.44\text{Chinstrap} - 33.52\text{Gentoo} + 0.09(\text{flipper} \times \text{Chinstrap}) + 0.18(\text{flipper} \times \text{Gentoo})$$

$$M2:\hat{y} = 37.26 + 3.13\text{male} + 9.32\text{Chinstrap} + 8.31\text{Gentoo} + 1.39\text{male} \times \text{Chinstrap} + 0.78\text{male} \times \text{Gentoo}$$

$$M3:\hat{y} = -26.75 + 0.34\text{flipper} + 0.01\text{body mass} - 0.00003(\text{flipper} \times \text{body mass})$$

Let's find the  $R^2_{\text{adj}}$  for each model.

Again, we can use either `summary()` or `anova()`

## Example: M1

```
##
## Call:
## lm(formula = bill_length_mm ~ flipper_length_mm + species_Chinstrap +
##     species_Gentoo + species_Chinstrap:flipper_length_mm + species_Gentoo:flipper_length_mm,
##     data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.698 -1.703  0.036  1.566 12.439
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      13.0358      6.1840   2.11 0.03579 *
## flipper_length_mm    0.1357    0.0325   4.17 0.000039 ***
## species_Chinstrap   -7.4424    10.5683  -0.70 0.48180
## species_Gentoo     -33.5237    9.9202  -3.38 0.00081 ***
## flipper_length_mm:species_Chinstrap  0.0852    0.0545   1.56 0.11909
## flipper_length_mm:species_Gentoo    0.1776    0.0483   3.68 0.00027 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.55 on 327 degrees of freedom
## Multiple R-squared:  0.785, Adjusted R-squared:  0.782
## F-statistic: 239 on 5 and 327 DF, p-value: <2e-16
```



## Example: M2

```
##  
## Call:  
## lm(formula = bill_length_mm ~ sex_male + species_Chinstrap +  
##     species_Gentoo + species_Chinstrap:sex_male + species_Gentoo:sex_male,  
##     data = data)  
##  
## Residuals:  
##      Min       1Q   Median       3Q      Max   
## -5.790 -1.374 -0.064   1.210 11.426   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)      37.258      0.271  137.47 < 2e-16 ***  
## sex_male          3.133      0.383   8.17 6.6e-15 ***  
## species_Chinstrap 9.316      0.481  19.38 < 2e-16 ***  
## species_Gentoo    8.306      0.407  20.39 < 2e-16 ***  
## sex_male:species_Chinstrap 1.388      0.680   2.04  0.042 *  
## sex_male:species_Gentoo 0.777      0.572   1.36  0.175      
## ---  
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
##  
## Residual standard error: 2.32 on 327 degrees of freedom  
## Multiple R-squared:  0.823, Adjusted R-squared:  0.821   
## F-statistic: 305 on 5 and 327 DF, p-value: <2e-16
```

## Example: M3

```
##
## Call:
## lm(formula = bill_length_mm ~ flipper_length_mm + body_mass_g +
##     flipper_length_mm:body_mass_g, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.106 -2.746 -0.786  2.158 18.955
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.68e+01   2.11e+01   -1.26   0.2068
## flipper_length_mm    3.39e-01   1.07e-01    3.17   0.0016 **
## body_mass_g       5.86e-03   4.84e-03    1.21   0.2270
## flipper_length_mm:body_mass_g -2.59e-05   2.34e-05   -1.10   0.2701
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.15 on 329 degrees of freedom
## Multiple R-squared:  0.43,    Adjusted R-squared:  0.425
## F-statistic: 82.8 on 3 and 329 DF,  p-value: <2e-16
```

## Example:

Comparing the three models,

Model	Predictors	RSE	$R^2_{\text{adj}}$
1	species, flipper length	2.55	0.782
2	species, sex	2.32	0.821
3	flipper length, body mass	4.15	0.425

Model 2 accounts for the highest amount of variability in bill length at 82.1

Model 3 accounts for the lowest amount of variability in bill length at 42.5

Both RSE and  $R_{\text{adj}}^2$  are tools for us to assess how a model fits the data.

Sometimes, we just aren't going to have a high  $R_{\text{adj}}^2$ ... and we should be cautious about  $R_{\text{adj}}^2$  that are too high.

We should consider the change in  $R_{\text{adj}}^2$  when removing/adding predictors to the model.

How much “better” is the model with/without the term?

If it is going to complicate the analysis (e.g., an interaction term), is it worth the increase in  $R_{\text{adj}}^2$ ?