INTERPRETING R OUTPUT FOR SIMPLE LINEAR REGRESSION

Jonathan M Brown
Updated 2021-12-29

NBA 2014-15 data from Basketball-Reference.com

Model: $Games\ Won = \beta_0 + \beta_1 * Points\ Per\ Game + \varepsilon$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -194.5434    39.4841  -4.927 3.38e-05 ***
PTS.G          2.3550     0.3944   5.971 1.98e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.082 on 28 degrees of freedom
Multiple R-squared:  0.5601,	Adjusted R-squared:  0.5444 
F-statistic: 35.65 on 1 and 28 DF,  p-value: 1.98e-06
```

These are the parameters in your model. You will always have an intercept, unless you tell R to not have an intercept.

Model: $Games\ Won = \beta_0 + \beta_1 * Points\ Per\ Game + \varepsilon$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -194.5434    39.4841  -4.927 3.38e-05 ***
PTS.G          2.3550     0.3944   5.971 1.98e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.082 on 28 degrees of freedom
Multiple R-squared:  0.5601,	Adjusted R-squared:  0.5444 
F-statistic: 35.65 on 1 and 28 DF,  p-value: 1.98e-06
```

These are the coefficients for your parameters.

<u>Intercept estimate:</u> The average value for Y when X = 0.

<u>Slope estimate interpretation when both variables are CONTINUOUS:</u>
Given a one unit increase in X, this is the expected change in Y, on average.
(This interpretation changes for categorical variables and some variable transformations.)

Model: $Games\ Won = \beta_0 + \beta_1 * Points\ Per\ Game + \varepsilon$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -194.5434    39.4841  -4.927 3.38e-05 ***
PTS.G          2.3550     0.3944   5.971 1.98e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.082 on 28 degrees of freedom
Multiple R-squared:  0.5601,	Adjusted R-squared:  0.5444 
F-statistic: 35.65 on 1 and 28 DF,  p-value: 1.98e-06
```

These are the standard errors on each of the coefficients.

The standard error is the estimated variability in a coefficient due to SAMPLING VARIABILITY. I.e., a different sample may result in different coefficients, and the variability of coefficients across samples is estimated by the standard error of the respective coefficient.

Model: $Games\ Won = \beta_0 + \beta_1 * Points\ Per\ Game + \varepsilon$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -194.5434    39.4841  -4.927 3.38e-05 ***
PTS.G          2.3550     0.3944   5.971 1.98e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.082 on 28 degrees of freedom
Multiple R-squared:  0.5601,	Adjusted R-squared:  0.5444 
F-statistic: 35.65 on 1 and 28 DF,  p-value: 1.98e-06
```

These are the t-values for each of the coefficients.

$t\ value = \frac{\hat{\beta}}{SE_{\hat{\beta}}}$ = “Estimate” / “Std. Error”

Model: $Games\ Won = \beta_0 + \beta_1 * Points\ Per\ Game + \varepsilon$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -194.5434    39.4841  -4.927 3.38e-05 ***
PTS.G          2.3550     0.3944   5.971 1.98e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.082 on 28 degrees of freedom
Multiple R-squared:  0.5601,	Adjusted R-squared:  0.5444
F-statistic: 35.65 on 1 and 28 DF,  p-value: 1.98e-06
```

These are the p-values for the t-values on each of the coefficients, given n – 2 degrees of freedom.

Model: $Games\ Won = \beta_0 + \beta_1 * Points\ Per\ Game + \varepsilon$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -194.5434    39.4841  -4.927 3.38e-05 ***
PTS.G          2.3550     0.3944   5.971 1.98e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.082 on 28 degrees of freedom
Multiple R-squared:  0.5601,	Adjusted R-squared:  0.5444 
F-statistic: 35.65 on 1 and 28 DF,  p-value: 1.98e-06
```

These are the significance codes for the p-values on each of the coefficients.

The key for how to interpret the significance codes is shown by the red box.

$$\textbf{Model: } Games\ Won = \beta_0 + \beta_1 * Points\ Per\ Game + \varepsilon$$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -194.5434    39.4841  -4.927 3.38e-05 ***
PTS.G          2.3550     0.3944   5.971 1.98e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.082 on 28 degrees of freedom
Multiple R-squared:  0.5601,  Adjusted R-squared:  0.5444
F-statistic: 35.65 on 1 and 28 DF,  p-value: 1.98e-06
```

This is the “Standard Error of the Model”, “Standard Deviation of the Residuals”, “Regression Standard Error”, “Root Mean Square Error”, “RMSE”, “Typical Error”, etc. In general, it quantifies how well or poorly the model does at predicting Y values in the data, on average. Thus, it’s like an “average error” for the model.

Symbolically, the book lists this as: $\hat{\sigma}_e$

Mathematically: $\hat{\sigma}_e = \sqrt{\frac{\sum(y-\hat{y})^2}{n-2}} = \sqrt{\frac{Sum\ of\ Squared\ Errors}{n-2}} = \sqrt{\frac{Sum\ of\ Squared\ Errors}{Degrees\ of\ Freedom\ Error}}$

Model: $Games\ Won = \beta_0 + \beta_1 * Points\ Per\ Game + \varepsilon$

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -194.5434    39.4841  -4.927 3.38e-05 ***
PTS.G          2.3550     0.3944   5.971 1.98e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.082 on 28 degrees of freedom
Multiple R-squared:  0.5601,	Adjusted R-squared:  0.5444 
F-statistic: 35.65 on 1 and 28 DF,  p-value: 1.98e-06
```

This is the R^2 for the model. It is interpreted as the percentage of variation in the response variable that is explained by variation in the explanatory variable.

It is calculated as the percentage of the total sums of squares (SST) that is composed of sums of squares from the model (SSM).

Mathematically: $R^2 = \frac{SSM}{SST} = 1 - \frac{SSE}{SST}$

Model: $Games\ Won = \beta_0 + \beta_1 * Points\ Per\ Game + \varepsilon$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -194.5434    39.4841  -4.927 3.38e-05 ***
PTS.G          2.3550     0.3944   5.971 1.98e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.082 on 28 degrees of freedom
Multiple R-squared:  0.5601,	Adjusted R-squared:  0.5444
F-statistic: 35.65 on 1 and 28 DF,  p-value: 1.98e-06
```

This is the adjusted R^2 for the model.

It is calculated the same way as R^2 except it is adjusted for the number of variables in the model. What it does is penalize you for adding useless variables to the model.

If you add more variables to a model, mathematically, R^2 will always go up. However, if the variables you're adding are statistically insignificant, then adjusted R^2 takes this into account, and presents a much LOWER value than R^2.

Model: $\boldsymbol{Games\ Won = \beta_0 + \beta_1 * Points\ Per\ Game + \varepsilon}$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -194.5434    39.4841  -4.927 3.38e-05 ***
PTS.G          2.3550     0.3944   5.971 1.98e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.082 on 28 degrees of freedom
Multiple R-squared:  0.5601,	Adjusted R-squared:  0.5444 
F-statistic: 35.65 on 1 and 28 DF,  p-value: 1.98e-06
```

This is the F-value for the model, with two values for degrees of freedom (numerator and denominator, respectively, see below).

The F-value measures the significance of the OVERALL model, and not just one variable. This has more use when there is more than one explanatory variable. For a one variable model, F = t-value^2 (the t-value on the given variable).

Mathematically: $F = \frac{Mean\ Square\ Model}{Mean\ Square\ Error} = \frac{Mean\ Square\ Model}{\hat{\sigma}_e{}^2} = \frac{SSM / Degrees\ of\ Freedom\ Model}{SSE / Degrees\ of\ Freedom\ Error}$

Model: $Games\ Won = \beta_0 + \beta_1 * Points\ Per\ Game + \varepsilon$

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -194.5434    39.4841  -4.927 3.38e-05 ***
PTS.G          2.3550     0.3944   5.971 1.98e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.082 on 28 degrees of freedom
Multiple R-squared:  0.5601,	Adjusted R-squared:  0.5444 
F-statistic: 35.65 on 1 and 28 DF,  p-value: 1.98e-06
```

This is the p-value for the F-value of the model.