Q15

Rahul Atre

2023-10-13

Q15 [Regression Modeling]

We investigate the t-statistic for the null hypothesis $H_0 : \beta = 0$ in simple linear regression without an intercept. To begin, we generate a predictor x and a response y as follows.

```r
set.seed(1)
x=rnorm(100)
y=2*x+rnorm(100)
```

1. Perform a simple linear regression of y onto x, without an intercept. Report the coefficient estimate β, the standard error of this coefficient estimate, and the t-statistic and p-value associated with the null hypothesis $H_0 : \beta = 0$. Comment on these results. (You can perform regression without an intercept using the command lm(x ~ y + 0)).

```r
lin_model = lm(y ~ x + 0)
summary(lin_model)
```

```
##
## Call:
## lm(formula = y ~ x + 0)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9154 -0.6472 -0.1771  0.5056  2.3109
##
## Coefficients:
##   Estimate Std. Error t value Pr(>|t|)
## x   1.9939     0.1065   18.73   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9586 on 99 degrees of freedom
## Multiple R-squared:  0.7798, Adjusted R-squared:  0.7776
## F-statistic: 350.7 on 1 and 99 DF,  p-value: < 2.2e-16
```

The coefficient estimate $\hat{\beta}$ is 1.9939, the standard error for this estimate is 0.1065, the t-statistic is 18.73, and the p-value is less than 2e-16, which itself is substantially less than 0.05. From the extremely small p-value, we can reject the null hypothesis $H_0 : \beta = 0$.

2. Now perform a simple linear regression of x onto y without an intercept, and report the coefficient estimate, its standard error, and the corresponding t-statistic and p-values associated with the null hypothesis $H_0 : \beta = 0$. Comment on these results.

```
lin_model2 = lm(x ~ y + 0)
summary(lin_model2)
```

```
##
## Call:
## lm(formula = x ~ y + 0)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8699 -0.2368  0.1030  0.2858  0.8938
##
## Coefficients:
##   Estimate Std. Error t value Pr(>|t|)
## y  0.39111    0.02089   18.73   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4246 on 99 degrees of freedom
## Multiple R-squared:  0.7798, Adjusted R-squared:  0.7776
## F-statistic: 350.7 on 1 and 99 DF,  p-value: < 2.2e-16
```

From the above function call, we get that the estimated coefficient for β is 0.39111, the standard error is 0.02089, the t-statistic is 18.73, and the p-value is less than 2e-16, which is substantially less than 0.05. This implies that the parameter is statistically significant, and we can reject the null-hypothesis H_0.

3. What is the relationship between the results obtained in 1. and 2.?

From the results in 1 and 2, we can see that the t-statistic and p-value are the same. However, the estimated beta coefficient values are different.

5. Using the results from 4., argue that the t-statistic for the regression of y onto x is the same as the t-statistic for the regression of x onto y.

The t-statistic formula is given as:

$$\frac{(\sqrt{n-1})\sum_{i=1}^{n} x_i y_i}{\sqrt{(\sum_{i=1}^{n} x_i^2)(\sum_{i'}^{n} y_{i'}^2) - (\sum_{i'}^{n} x_{i'} y_{i'})^2}}$$

We can see from the formula that, if we interchange the x and y values, the t-statistic would remain the same. Thus, the t-statistic for regression of y onto x will always be the same as the t-statistic for the regression of x onto y.

6. In R, show that when regression is performed with an intercept, the t-statistic for $H_0 : \beta_1 = 0$ is the same for the regression of y onto x as it is for the regression of x onto y.

We will use the same data, except instead we will allow R to choose the intercept instead of setting it strictly to 0.

```r
lin_model_intercept = lm(y ~ x)
summary(lin_model_intercept)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8768 -0.6138 -0.1395  0.5394  2.3462
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.03769    0.09699  -0.389    0.698
## x            1.99894    0.10773  18.556   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9628 on 98 degrees of freedom
## Multiple R-squared:  0.7784, Adjusted R-squared:  0.7762
## F-statistic: 344.3 on 1 and 98 DF,  p-value: < 2.2e-16
```

```r
lin_model_intercept2 = lm(x ~ y)
summary(lin_model_intercept)
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.8768 -0.6138 -0.1395  0.5394  2.3462
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.03769    0.09699  -0.389    0.698
## x            1.99894    0.10773  18.556   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9628 on 98 degrees of freedom
## Multiple R-squared:  0.7784, Adjusted R-squared:  0.7762
## F-statistic: 344.3 on 1 and 98 DF,  p-value: < 2.2e-16
```

As we can see from the two summary tables, the t-statistics for the beta parameters are given as 18.556 for β_1, which is the same for both models.