Model Inference & Partitioning Regression Error

SDS 291

2/5/2020

Regression Output

```
##
## Call:
## lm(formula = y1 ~ x1, data = anscombe)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.92127 -0.45577 -0.04136  0.70941  1.83882
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0001     1.1247   2.667  0.02573 *
## x1            0.5001     0.1179   4.241  0.00217 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.237 on 9 degrees of freedom
## Multiple R-squared:  0.6665, Adjusted R-squared:  0.6295
## F-statistic: 17.99 on 1 and 9 DF,  p-value: 0.00217
```

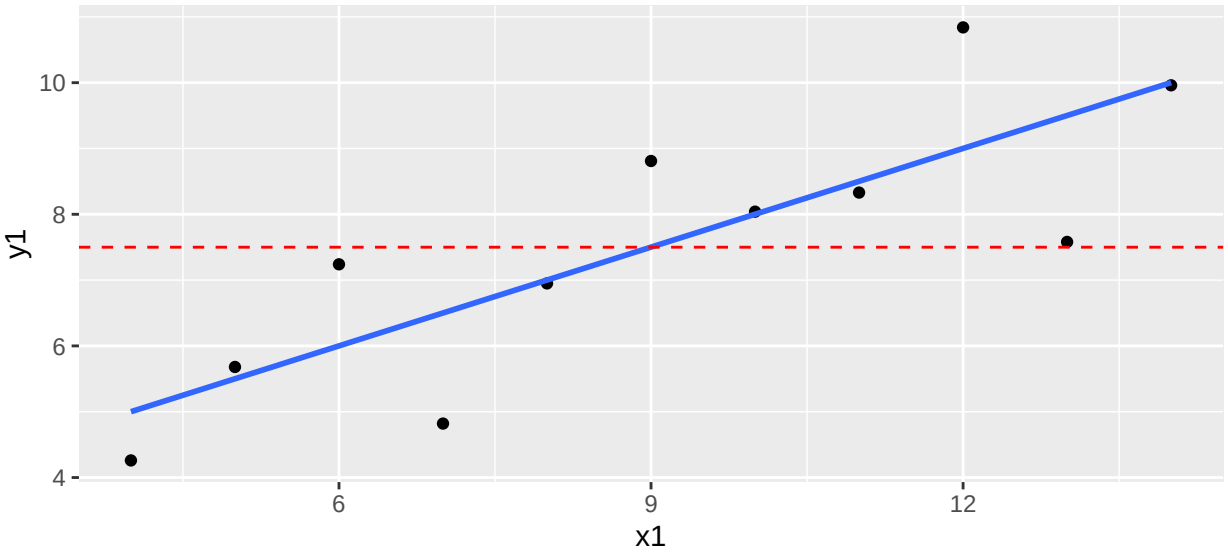
1. Interpret the slope in a sentence.

2. Use the information in the table for x1 to calculate the t-value for the slope.

2. Test the hypothesis that x1 has a linear relationship with y1.

3. Calculate the 95% confidence interval for the slope to 3 decimal places ($t^* = 2.26$) and interpret in a sentence.

Residuals and Model Error



x1	observed (y)	predicted ($\hat{y}$ line)	residual ($y - \hat{y}$)	residual^2	mean y ($\bar{y}$)	model deviation $\hat{y} - \bar{y}$	model^2
4	4.26				7.5		
5	5.68				7.5		
6	7.24				7.5		
7	4.82				7.5		
8	6.95				7.5		
9	8.81				7.5		
10	8.04				7.5		
11	8.33				7.5		
12	10.84				7.5		
13	7.58				7.5		
14	9.96				7.5		
Sum the		*Residuals*			*and Model*		

1. Calculate the residual (observed - predicted, or $y - \bar{y}$) and the model (predicted - mean, or $\hat{y} - \bar{y}$) for each value of x.
2. Square each residual and model term.
3. Calculate the sum of squared residuals (SSR) and the Sum of Squares of the Model (SSM).
4. Calculate the sum of squares total (SST = SSR + SSM)

5. Calculate r^2 ($r^2 = \frac{SSModel}{SSTotal}$ or $r^2 = 1 - \frac{SSError}{SSTotal}$)

ANOVA Table

Source	df	Sum of Squares	Mean Square	F-Statistic
Model	1			
Error	n-2			
—	—	————	————	————
Total	n-1			
—	—	————	————	————

1. Calculate the mean squares for the model ($MSModel = \frac{SSModel}{1}$) and the $MSError = \frac{SSE}{n-2}$)

2. Calculate the F statistic: $F = \frac{MSModel}{MSError}$

3. Look at the F Distribution calculator (at https://gallery.shinyapps.io/dist_calc/) and estimate the p-value for your F statistic with 1 and n-2 degrees of freedom.

Comparing to R output

Regression Output

```
summary(m1)
```

```
##
## Call:
## lm(formula = y1 ~ x1, data = anscombe)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.92127 -0.45577 -0.04136  0.70941  1.83882
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.0001     1.1247   2.667  0.02573 *
## x1            0.5001     0.1179   4.241  0.00217 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.237 on 9 degrees of freedom
## Multiple R-squared:  0.6665, Adjusted R-squared:  0.6295
## F-statistic: 17.99 on 1 and 9 DF,  p-value: 0.00217
```

Confidence Intervals

```
confint(m1)
```

```
##                 2.5 %    97.5 %
## (Intercept) 0.4557369 5.5444449
## x1          0.2333701 0.7668117
```

ANOVA table

```
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: y1
##           Df Sum Sq Mean Sq F value  Pr(>F)
## x1         1 27.510 27.5100   17.99 0.00217 **
## Residuals  9 13.763  1.5292
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1. Identify where in the output above you find the results that you calculated:
    1. t-statistic for the slope
    2. confidence interval for the slope
    3. F statistic
    4. R^2