

MATH 423/533 – SIMPLE LINEAR REGRESSION: ANOVA

For the simple linear regression model we have the sums of squares decomposition

$$SS_T = SS_{Res} + SS_R.$$

The ANOVA F test uses the statistic

$$F = \frac{SS_R/(p-1)}{SS_{Res}/(n-p)}$$

with $p = 2$ to test the hypothesis $H_0 : \beta_1 = 0$: if H_0 is true, then

$$F \sim \text{Fisher}(p-1, n-p).$$

The ANOVA table arranges the requires information in tabular form:

Source	SS	df	MS	F
Regression	SS_R	$p-1$	$SS_R/(p-1)$	F
Residual	SS_{Res}	$n-p$	$SS_{Res}/(n-p)$	
Total	SS_T	$n-1$		

Note that R swaps the positions of columns 1 and 2.

Source	df	SS	MS	F
Regression	$p-1$	SS_R	$SS_R/(p-1)$	F
Residual	$n-p$	SS_{Res}	$SS_{Res}/(n-p)$	
Total	$n-1$	SS_T		

In either case we have summation results in columns 1 and 2

$$SS_R + SS_{Res} = SS_T$$

and

$$(p-1) + (n-p) = (n-1)$$

from the previous theory.

R also adds the p -value from the test of the null hypothesis in a further final column.

The anova function can be used to form the ANOVA table:

Simple Linear Regression: ANOVA

```

1 > x<-RocketProp$Age
2 > y<-RocketProp$Strength
3 > n<-length(x)
4 > summary(fit.RP)
5
6 Call:
7 lm(formula = y ~ x)
8
9 Residuals:
10      Min       1Q   Median       3Q      Max
11 -215.98  -50.68   28.74   66.61  106.76
12
13 Coefficients:
14             Estimate Std. Error t value Pr(>|t|)
15 (Intercept)  2627.822     44.184   59.48  < 2e-16 ***
16 x           -37.154      2.889  -12.86 1.64e-10 ***
17 ---
18 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1  1
19
20 Residual standard error: 96.11 on 18 degrees of freedom
21 Multiple R-squared:  0.9018,    Adjusted R-squared:  0.8964
22 F-statistic: 165.4 on 1 and 18 DF,  p-value: 1.643e-10

```

Line 22 contains the computation of the F statistic (165.4), the degrees of freedom values $p-1 = 1$ and $n-p = 18$, and the p -value in the test of the hypothesis $H_0 : \beta_1 = 0$ (1.643×10^{-10}); the hypothesis is strongly rejected.

```

23 > anova(fit.RP)
24 Analysis of Variance Table
25
26 Response: y
27      Df  Sum Sq Mean Sq F value    Pr(>F)
28 x      1 1527483 1527483  165.38 1.643e-10 ***
29 Residuals 18  166255    9236
30 ---
31 Signif. codes:  0  ***  0.001  **  0.01  *  0.05  .  0.1  1
32 >
33 > anova(fit.RP)[, 'Df' ]
34 [1]  1 18
35 > anova(fit.RP)[, 'Sum Sq' ]
36 [1] 1527482.7 166254.9
37 >
38 > sum(anova(fit.RP)[, 'Sum Sq' ])-(n-1)*var(y)
39 [1] -9.313226e-10

```

The ANOVA table is contained between lines 24 and 31. The 'Regression' entry is listed on line 28 as x. Note that the F statistic and p -values, listed in the columns headed F value and $\text{Pr}(>F)$, are identical to the ones from the `lm` summary (line 22).