**ANOVA approach to hypothesis testing**

**The F Test**

* In this section, we will talk about a different strategy to testing *H*0:*β*1=0

In simple linear regression, this results in the *exact same-value* as the test that uses the t-statistic.

However, this strategy is more applicable to general linear hypotheses that we’ll discuss in multiple linear regression.

 Testing *H*0:*β*1=0

is really a comparison between the two models:

*H*0: *Yi*=*β*0+*ϵi*

*HA*: *Yi*=*β*0+*β*1*Xi*+*ϵi*

 We will call the first model the **reduced** model and the second model the **full** model. This is because the reduced model is a subset of the full model (you get the reduced from the full by setting *β*1=0).

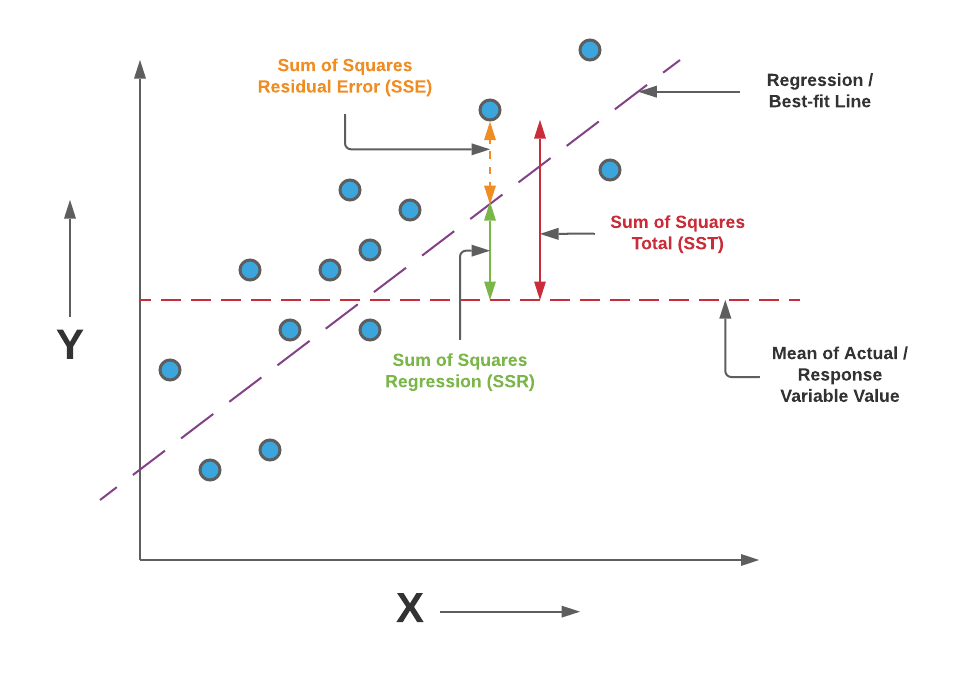
 Our strategy will be to compare the residuals under *H*0 and *HA*. If *HA* were true, we would expect those residuals to be much smaller than the residuals under *H*0 (because the line fits a lot better).

If *H*0 were true, then we would expect the residuals under *HA* to only be a little bit smaller than those under *H*0

 We fit *HA* by the method of least squares, obtaining the OLS estimates and the corresponding residuals.

 We fit *H*0 also by least squares. It turns out that under *H*0, the OLS estimate is just *Y*¯ We measure how small the residuals are by the sum of squared residuals.

PICTORIAL and PLOT representation SSTO, SSE, and SSR



Example:

Chart, scatter chart

Description automatically generated

SSTO : Sum of Squares Total SSR : Sum of Squares due to Regression SSE : Sum of Squares due to Errors Note: Using the R generated ANOVA table, SSE is referred to as the Sum of Squares Residuals

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | Y | Yi(hat)  Predicted  Value | YI - Yi(hat)  Residuals | (YI - Y(bar))2 | (YI - Y(hat))2 | (Yi(hat) - Y(bar))2 |
| 34 | 5 | 4.1505 | .8496 | 25 | .72165 | 34.21665 |
| 108 | 17 | 14.9693 | 2.0307 | 49 | 4.12374 | 24.69394 |
| 64 | 11 | 8.5365 | 2.4635 | 1 | 6.06883 | 2.14183 |
| 88 | 8 | 12.0453 | -4.0453 | 4 | 16.3644 | 4.18325 |
| 99 | 14 | 13.6535 | .3465 | 16 | .12006 | 13.34906 |
| 51 | 5 | 6.6359 | -1.6359 | 25 | 2.6762 | 11.31717 |
| TOTALS |  |  |  | 120 | 30.07488 | 89.9019 |

Y(bar) = 10

The following Regression Model for the data is found by using R coding: lm(y ~ x)

Y(hat) = -0.8203 + 0.1462X

The following ANOVA table is generated by using R coding: anova(lm(y ~ x)

Analysis of Variance Table

Response: y

Df Sum Sq Mean Sq F value Pr(>F)

x 1 89.925 89.925 11.96 0.02586 \*

Residuas 4 30.075 7.519

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Now let us conduct the F test to determine if *H*o: *Yi*=*β*o+*ϵi* should be rejected.

If the Fvalue, is greater than the Fcritical, we reject *H*o and conclude *H*a

If the Fvalue, is less than or equal to Fcritical, we fail to reject *H*a

The Fvalue, is found in the ANOVA table.

The Fcritical value if found in the F distribution table, using the construction

F(1 – alpha; 1, n -2) or by using

For our case, Fvalue, is 11.96 (found in the ANOVA output table).

The Fcritical value can be found as follows;

Let alpha = .05, and our designated degrees of freedom are 1 and 4.

We therefore require F(.95,1, 4), now going to the F table in the back of the textbook ;we get 7.71

We can also use R code as follows to get the Fcritical value.

qf(p=.05, df1=1, df2=4, lower.tail=FALSE)

7.708647

Since Fvalue,(11.96) is greater than Fcritical, ,(7.71), we will reject the null

Hypothesis that B1= 0 and conclude that B1 does not = 0. Moreover, a linear

relationship does exist between Y and X.

Graphical Interpretation: F Distribution Curve

Diagram

Description automatically generated with low confidence

If Fvalue is greater than (to the right of Fcritical you are to reject Ho

If Fvalue is less than (to the left)of Fcritical or equal to Fcritical , you fail to reject Ho

