14.4: Hypothesis Test for Simple Linear Regression

We will now describe a hypothesis test to determine if the regression model is meaningful; in other words, does the value of X in any way help predict the expected value of Y?

Simple Linear Regression ANOVA Hypothesis Test

Model Assumptions

- The residual errors are random and are normally distributed.
- The standard deviation of the residual error does not depend on X
- A linear relationship exists between X and Y
- The samples are randomly selected

Test Hypotheses

H_o: X and Y are not correlated

H_a: X and Y are correlated

H_o: β_1 (slope) = 0

H_a: β_1 (slope) ≠ 0

Test Statistic

$$F = \frac{MS_{\text{Regression}}}{MS_{\text{Error}}}$$

$$df_{\text{num}} = 1$$

$$df_{\text{den}} = n - 2$$

Sum of Squares

$$SS_{\text{Total}} = \sum (Y - \bar{Y})^2$$

$$SS_{\text{Error}} = \sum (Y - \hat{Y})^2$$

$$SS_{\text{Regression}} = SS_{\text{Total}} - SS_{\text{Error}}$$

In simple linear regression, this is equivalent to saying “Are X an Y correlated?”

In reviewing the model, $Y = \beta_0 + \beta_1 X + \varepsilon$, as long as the slope (β_1) has any non-zero value, X will add value in helping predict the expected value of Y. However, if there is no correlation between X and Y, the value of the slope (β_1) will be zero. The model we can use is very similar to One Factor ANOVA.

The Results of the test can be summarized in a special ANOVA table:

Source of Variation	Sum of Squares (SS)	[illegible]
Factor (due to X)	$SS_{\text{Regression}}$	[illegible]
Error (Residual)	SS_{Error}	[illegible]
Total	SS_{Total}	[illegible]

Example: Rainfall and sales of sunglasses

Design: Is there a significant correlation between rainfall and sales of sunglasses?

Research Hypotheses:

H_o: Sales and Rainfall are not correlated H_o: □1 (slope) = 0

H_a: Sales and Rainfall are correlated H_a: □1 (slope) ≠ 0

Type I error would be to reject the Null Hypothesis and t claim that rainfall is correlated with sales of sunglasses, when they are not correlated. The test will be run at a level of significance (α) of 5%.

The test statistic from the table will be $\mathrm{F} = \dfrac{\text{MSRegression}}{\text{MSError}}$. The degrees of freedom for the numerator will be 1, and the degrees of freedom for denominator will be 5-2=3.

Critical Value for F at $\alpha$of 5% with $df_{num} = 1$ and df_{den}=3} is 10.13. Reject \(H_o if $F > 10.13$. We will also run this test using the p-value method with statistical software, such as Minitab.

Data/Results

Source	SS	df	MS	F	p-value
Regression	341.422	1	341.422	26.551	0.0142
Error	38.578	3	12.859		
TOTAL	380.000	4			

$F = 341.422/12.859 = 26.551$, which is more than the critical value of 10.13, so Reject H_o. Also, the p-value = 0.0142 < 0.05 which also supports rejecting H_o.

Conclusion

Sales of Sunglasses and Rainfall are negatively correlated.