Lesson 19: Chapter 10 Section 2

Caleb Moxley

BSC Mathematics

11 November 15

§10.1 More Detail about Simple Linear Regression

In this section, we'll cover two more tests which deal will linear regression — the F test for β_1, i.e. for the slope, and the ρ test.

§10.1 More Detail about Simple Linear Regression

In this section, we'll cover two more tests which deal will linear regression — the F test for β_1, i.e. for the slope, and the ρ test. The F test involves an analysis of variance — a type of test which is used in Chapters 12 & 13.

§10.1 More Detail about Simple Linear Regression

In this section, we'll cover two more tests which deal will linear regression — the F test for β_1, i.e. for the slope, and the ρ test. The F test involves an analysis of variance — a type of test which is used in Chapters 12 & 13. Let's look at the framework for this test first!

§10.1 More Detail about Simple Linear Regression

Every data point in the response variable column can be obtained in the basic fashion:

$$\text{DATA} = \text{FIT} + \text{RESIDUAL}.$$

§10.1 More Detail about Simple Linear Regression

Every data point in the response variable column can be obtained in the basic fashion:

$$\text{DATA} = \text{FIT} + \text{RESIDUAL}.$$

That is,

$$y = (\hat{y}) + (y - \hat{y}).$$

§10.1 More Detail about Simple Linear Regression

Every data point in the response variable column can be obtained in the basic fashion:

$$\text{DATA} = \text{FIT} + \text{RESIDUAL}.$$

That is,

$$y = (\hat{y}) + (y - \hat{y}).$$

If we subtract $\bar{y}$ from both sides, we get

$$y - \bar{y} = (\hat{y} - \bar{y}) + (y - \hat{y})$$

§10.1 More Detail about Simple Linear Regression

Every data point in the response variable column can be obtained in the basic fashion:

$$\text{DATA} = \text{FIT} + \text{RESIDUAL}.$$

That is,

$$y = (\hat{y}) + (y - \hat{y}).$$

If we subtract $\bar{y}$ from both sides, we get

$$y - \bar{y} = (\hat{y} - \bar{y}) + (y - \hat{y})$$

This means that the deviations also follow the pattern

$$\text{DATA} = \text{FIT} + \text{RESIDUAL}.$$

§10.1 More Detail about Simple Linear Regression

If we square each of these deviations and add over all observations, it can be shown that the equation below is also true.

$$\sum(y - \bar{y})^2 = \sum(\hat{y} - \bar{y})^2 + \sum(y - \hat{y})^2$$

§10.1 More Detail about Simple Linear Regression

If we square each of these deviations and add over all observations, it can be shown that the equation below is also true.

$$\sum(y-\bar{y})^2 = \sum(\hat{y}-\bar{y})^2 + \sum(y-\hat{y})^2$$

We call these the total sum of squares (SST), the sum of squares due to the model (SSM), and the sum of squares due to deviations from the model, i.e. due to errors, (SSE), respectively. They have degrees of freedom $n-1$, 1, and $n-2$ respectively. We denote them DFT, DFM, and DFE.

The mean square error (MSE) is $\frac{SSE}{DFE}$, and the mean square for the total and the mean square for the model are refined analogously. We have the following:

§10.1 More Detail about Simple Linear Regression

The mean square error (MSE) is $\frac{SSE}{DFE}$, and the mean square for the total and the mean square for the model are refined analogously. We have the following:

Definition (sum of squares, degrees of freedom, and mean squares)

Sum of squares represent variation present in the responses. They are calculated by summing square deviations. **Analysis of variance** partitions the total variation between two sources.

$$\text{SST} = \text{SSM} + \text{SSE}$$

The **degrees of freedom** are associated with each sum of squares.

$$\text{DFT} = \text{DFM} + \text{DFE}$$

The **mean squares** are $\frac{\text{sum of squares}}{\text{degrees of freedom}}$.

§10.1 More Detail about Simple Linear Regression

It was mentioned before that r^2 — the square of linear correlation coefficient, sometimes called the coefficient of determination — is the ratio of the explained variation to the total variation. This is because:

$$r^2 = \frac{\mathrm{SSM}}{\mathrm{SST}}.$$

Definition (Analysis of Variance F test)

In the simple linear regression model, the hypotheses

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

are tested by the F statistic

$$F = \frac{\text{MSM}}{\text{MSE}}.$$

The P-value is the probability that a random variable having the $F(1, n-2)$ distribution is greater than or equal to the calculated value of the F statistic.

§10.1 More Detail about Simple Linear Regression

Definition (Analysis of Variance F test)

In the simple linear regression model, the hypotheses

$$H_0 : \beta_1 = 0$$

$$H_1 : \beta_1 \neq 0$$

are tested by the F statistic

$$F = \frac{\text{MSM}}{\text{MSE}}.$$

The P-value is the probability that a random variable having the $F(1, n-2)$ distribution is greater than or equal to the calculated value of the F statistic.

Typically, all these calculations are done with the help of technology.

§10.1 More Detail about Simple Linear Regression

We have one final test: the test for zero population linear correlation.

§10.1 More Detail about Simple Linear Regression

We have one final test: the test for zero population linear correlation.

Definition (test for a zero population correlation)

To test the hypothesis $H_0 : \rho = 0$, compute the t test statistic

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}},$$

where n is the sample size and r is the sample correlation. In terms of the $t(n-2)$ random variable, the P-value for the test of H_0 against

$$\begin{array}{lcr} H_a : \rho > 0 & \text{is} & P(T > t) \\ H_a : \rho < 0 & \text{is} & P(T < t) \\ H_a : \rho \neq 0 & \text{is} & 2P(T > |t|) \end{array}$$

§10.1 More Detail about Simple Linear Regression

Note: The test for ρ is **equivalent to the test for** β_1 **from the previous section!**

§10.1 More Detail about Simple Linear Regression

Note: The test for ρ is **equivalent to the test for β_1 from the previous section!** In order to use the test for ρ, the distribution of x (the explanatory variable) must be normal, and the distribution of y for a fixed x must be normal as well.

§10.1 More Detail about Simple Linear Regression

Example (mice's weights)

Use technology to test the claim that the data below is from a population which is not linearly correlated and that β_1 is non-zero.

Example (mice's weights)

Use technology to test the claim that the data below is from a population which is not linearly correlated and that β_1 is non-zero.

Rice (g)	10	12	14	16	18
Weight gain (g)	2.1	3.2	3.3	4.5	5.0

Well, the MSE and MSM are 0.227 and 5.041 respectively. Thus, the $F(1,3)$ stat is 66.62.

§10.1 More Detail about Simple Linear Regression

Example (mice's weights)

Use technology to test the claim that the data below is from a population which is not linearly correlated and that β_1 is non-zero.

Rice (g)	10	12	14	16	18
Weight gain (g)	2.1	3.2	3.3	4.5	5.0

Well, the MSE and MSM are 0.227 and 5.041 respectively. Thus, the $F(1,3)$ stat is 66.62. Thus, the P-value for the β_1 test is 0.004.

§10.1 More Detail about Simple Linear Regression

Example (mice's weights)

Use technology to test the claim that the data below is from a population which is not linearly correlated and that β_1 is non-zero.

Rice (g)	10	12	14	16	18
Weight gain (g)	2.1	3.2	3.3	4.5	5.0

Well, the MSE and MSM are 0.227 and 5.041 respectively. Thus, the $F(1,3)$ stat is 66.62. Thus, the P-value for the β_1 test is 0.004. We reject H_0 and support our claim that β_1 is non-zero. Now, $r = 0.9782$, thus our $t(3)$ test statistic is 8.162 and the P-value is also 0.004. And we reject our null hypothesis and fail to support the claim that our claim that the population is not linearly correlated.

Example (mice's weights)

Use technology to test the claim that the data below is from a population which is not linearly correlated and that β_1 is non-zero.

Rice (g)	10	12	14	16	18
Weight gain (g)	2.1	3.2	3.3	4.5	5.0

Well, the MSE and MSM are 0.227 and 5.041 respectively. Thus, the $F(1,3)$ stat is 66.62. Thus, the P-value for the β_1 test is 0.004. We reject H_0 and support our claim that β_1 is non-zero. Now, $r = 0.9782$, thus our $t(3)$ test statistic is 8.162 and the P-value is also 0.004. And we reject our null hypothesis and fail to support the claim that our claim that the population is not linearly correlated.