Linear Regression Models

Segment 6 – Advanced Topics in Linear Regression

Topic 3 – F-test in Linear Regression

Sudarsan N.S. Acharya (sudarsan.acharya@manipal.edu)

Topics

1. Recalling Hypothesis Tests

2. The F Statistic

3. The F-test Recipe

Recalling Hypothesis Testing

Recalling Hypothesis Testing

- Suppose we want to know if there is a linear relationship between the jth predictor predictor and the response variable.

Recalling Hypothesis Testing

- Suppose we want to know if there is a linear relationship between the jth predictor predictor and the response variable.
- We consider a *null hypothesis* $\beta_j = 0$;

Recalling Hypothesis Testing

- Suppose we want to know if there is a linear relationship between the jth predictor predictor and the response variable.
- We consider a *null hypothesis* $\beta_j = 0$; that is, we hypothesize that there is no relationship between the jth predictor and the response variable.

Recalling Hypothesis Testing

- Suppose we want to know if there is a linear relationship between the jth predictor predictor and the response variable.
- We consider a *null hypothesis* $\beta_j = 0$; that is, we hypothesize that there is no relationship between the jth predictor and the response variable.

- *Assuming that the null hypothesis is true,*

Recalling Hypothesis Testing

- Suppose we want to know if there is a linear relationship between the jth predictor predictor and the response variable.
- We consider a *null hypothesis* $\beta_j = 0$; that is, we hypothesize that there is no relationship between the jth predictor and the response variable.
- *Assuming that the null hypothesis is true,* $T = \frac{\hat{\beta}_j - E[\hat{\beta}_j]}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \underbrace{\beta_j}_{=0}}{SE(\hat{\beta}_j)}$

Recalling Hypothesis Testing

- Suppose we want to know if there is a linear relationship between the jth predictor predictor and the response variable.
- We consider a *null hypothesis* $\beta_j = 0$; that is, we hypothesize that there is no relationship between the jth predictor and the response variable.
- *Assuming that the null hypothesis is true,* $T = \frac{\hat{\beta}_j - E[\hat{\beta}_j]}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \underbrace{\beta_j}_{=0}}{SE(\hat{\beta}_j)}$ follows a t-distribution with $n-2$ degrees of freedom.

Recalling Hypothesis Testing

- Suppose we want to know if there is a linear relationship between the jth predictor predictor and the response variable.
- We consider a *null hypothesis* $\beta_j = 0$; that is, we hypothesize that there is no relationship between the jth predictor and the response variable.
- *Assuming that the null hypothesis is true,* $T = \frac{\hat{\beta}_j - E[\hat{\beta}_j]}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \underbrace{\beta_j}_{=0}}{SE(\hat{\beta}_j)}$ follows a t-distribution with $n-2$ degrees of freedom.
- From a particular dataset, we have the estimate $\hat{\beta}_j$ and the realized value of T denoted as t.

Recalling Hypothesis Testing

- Suppose we want to know if there is a linear relationship between the jth predictor predictor and the response variable.
- We consider a *null hypothesis* $\beta_j = 0$; that is, we hypothesize that there is no relationship between the jth predictor and the response variable.
- *Assuming that the null hypothesis is true,* $T = \frac{\hat{\beta}_j - E[\hat{\beta}_j]}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \underbrace{\beta_j}_{=0}}{SE(\hat{\beta}_j)}$ follows a t-distribution with $n-2$ degrees of freedom.
- From a particular dataset, we have the estimate $\hat{\beta}_j$ and the realized value of T denoted as t.
- We calculate $P(T \geq |t|)$, called the p-value,

Recalling Hypothesis Testing

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- Suppose we want to know if there is a linear relationship between the jth predictor predictor and the response variable.
- We consider a *null hypothesis* $\beta_j = 0$; that is, we hypothesize that there is no relationship between the jth predictor and the response variable.
- *Assuming that the null hypothesis is true,* $T = \frac{\hat{\beta}_j - E[\hat{\beta}_j]}{SE(\hat{\beta}_j)} = \frac{\hat{\beta}_j - \underbrace{\beta_j}_{=0}}{SE(\hat{\beta}_j)}$ follows a t-distribution with $n-2$ degrees of freedom.
- From a particular dataset, we have the estimate $\hat{\beta}_j$ and the realized value of T denoted as t.
- We calculate $P(T \geq |t|)$, called the p-value, and reject the null hypothesis if the p-value is smaller than a threshold, typically 0.05.

The F Statistic

The F Statistic

- The approach so far (called the t-test or the Wald test) lets us test whether the regression coefficient associated with *one* predictor is zero.

The F Statistic

- The approach so far (called the t-test or the Wald test) lets us test whether the regression coefficient associated with *one* predictor is zero. What if we want to test for several regression parameters *simultaneously*?

The F Statistic

- The approach so far (called the t-test or the Wald test) lets us test whether the regression coefficient associated with *one* predictor is zero. What if we want to test for several regression parameters *simultaneously*?
- Suppose we want to test the null hypothesis that the regression coefficients associated with all the predictors in a *particular set* S are equal to zeros.

The F Statistic

- The approach so far (called the t-test or the Wald test) lets us test whether the regression coefficient associated with *one* predictor is zero. What if we want to test for several regression parameters *simultaneously*?
- Suppose we want to test the null hypothesis that the regression coefficients associated with all the predictors in a *particular set* S are equal to zeros.
- For this purpose, we calculate

The F Statistic

- The approach so far (called the t-test or the Wald test) lets us test whether the regression coefficient associated with *one* predictor is zero. What if we want to test for several regression parameters *simultaneously*?
- Suppose we want to test the null hypothesis that the regression coefficients associated with all the predictors in a *particular set* S are equal to zeros.
- For this purpose, we calculate (1) the residual vector $\mathbf{r}$ using all the p predictors

The F Statistic

- The approach so far (called the t-test or the Wald test) lets us test whether the regression coefficient associated with *one* predictor is zero. What if we want to test for several regression parameters *simultaneously*?
- Suppose we want to test the null hypothesis that the regression coefficients associated with all the predictors in a *particular set* S are equal to zeros.
- For this purpose, we calculate (1) the residual vector $\mathbf{r}$ using all the p predictors (2) the residual vector $\mathbf{r}_S$ by *excluding* the predictors in the set S,

The F Statistic

- The approach so far (called the t-test or the Wald test) lets us test whether the regression coefficient associated with *one* predictor is zero. What if we want to test for several regression parameters *simultaneously*?
- Suppose we want to test the null hypothesis that the regression coefficients associated with all the predictors in a *particular set* S are equal to zeros.
- For this purpose, we calculate (1) the residual vector $\mathbf{r}$ using all the p predictors (2) the residual vector $\mathbf{r}_S$ by *excluding* the predictors in the set S, (3) calculate the F statistic

$$\frac{\left(\|\mathbf{r}_S\|^2 - \|\mathbf{r}\|^2\right)/(p - |S|)}{\|\mathbf{r}\|/(n-p)}$$

.

The F-test Recipe

The F-test Recipe

- The null hypothesis is that $\beta_j = 0$ for regression coefficients asscoiated with the predictors in the set S.

The F-test Recipe

- The null hypothesis is that $\beta_j = 0$ for regression coefficients ascoiated with the predictors in the set S.
- If this null hypothesis is true, and if the random error terms are normally distributed, then the F distribution can be used to determine how unlikely our observed value of the F statistic is given that the null hypothesis is true.

The F-test Recipe

- The null hypothesis is that $\beta_j = 0$ for regression coefficients ascoiated with the predictors in the set S.
- If this null hypothesis is true, and if the random error terms are normally distributed, then the F distribution can be used to determine how unlikely our observed value of the F statistic is given that the null hypothesis is true.
- If the F statistic much larger than 1, then we reject the null hypothesis;

The F-test Recipe

- The null hypothesis is that $\beta_j = 0$ for regression coefficients ascoiated with the predictors in the set S.
- If this null hypothesis is true, and if the random error terms are normally distributed, then the F distribution can be used to determine how unlikely our observed value of the F statistic is given that the null hypothesis is true.
- If the F statistic much larger than 1, then we reject the null hypothesis; otherwise, we fail to reject the null hypothesis.

The F-test Recipe

MANIPAL ACADEMY of HIGHER EDUCATION
(Institution of Eminence Deemed to be University)

- The null hypothesis is that $\beta_j = 0$ for regression coefficients ascoiated with the predictors in the set S.
- If this null hypothesis is true, and if the random error terms are normally distributed, then the F distribution can be used to determine how unlikely our observed value of the F statistic is given that the null hypothesis is true.
- If the F statistic much larger than 1, then we reject the null hypothesis;
- The F-test is a simple quantification our level of uncertainty about the linear relationship of multiple predictors at once with the response variable.

Summary

Summary

- F statistic and the F-test in the context of linear regression.

Summary

- F statistic and the F-test in the context of linear regression.
- t-test and F-test for analyzing predictors in linear regression.