

Hypothesis tests

Tiago M. Magalhães

Department of Statistics - ICE-UFJF

Juiz de Fora, September 30, 2026



Outline

- 1 Introduction
- 2 Hypothesis testing
- 3 F statistics
- 4 Chi-square statistics
- 5 Applications
- 6 References



Outline

- 1 Introduction
- 2 Hypothesis testing
- 3 F statistics
- 4 Chi-square statistics
- 5 Applications
- 6 References



Linear regression model

Suppose that $Y_1, Y_2, \dots, Y_n$ satisfy

$$Y_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \quad (1)$$



Linear regression model

Suppose that $Y_1, Y_2, \dots, Y_n$ satisfy

$$Y_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \quad (1)$$

where $\mathbf{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \dots, x_{\ell p})^\top$ is known, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ is a vector of unknown parameters to be estimated, $\varepsilon_\ell \sim \mathcal{D}(0, \sigma^2)$, $\ell = 1, 2, \dots, n$, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are independent random variables with mean zero and common variance σ^2 , also unknown, to be estimated.



Linear regression model

Suppose that $Y_1, Y_2, \dots, Y_n$ satisfy

$$Y_\ell = \mathbf{x}_\ell^\top \boldsymbol{\beta} + \varepsilon_\ell, \quad (1)$$

where $\mathbf{x}_\ell = (x_{\ell 1}, x_{\ell 2}, \dots, x_{\ell p})^\top$ is known, $\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^\top$ is a vector of unknown parameters to be estimated, $\varepsilon_\ell \sim \mathcal{D}(0, \sigma^2)$, $\ell = 1, 2, \dots, n$, $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are independent random variables with mean zero and common variance σ^2 , also unknown, to be estimated.



Matrix form

Equation (1), which we defined as the **linear regression model** (LRM), can be written in matrix form as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$



Matrix form

Equation (1), which we defined as the **linear regression model** (LRM), can be written in matrix form as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$

where $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$, $\boldsymbol{\varepsilon} \sim \mathcal{D}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, where $\mathbf{0}$ is the zero vector of dimension n , $\mathbf{I}_n$ is the identity matrix of order n and $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top$ is the design matrix.



Matrix form

Equation (1), which we defined as the **linear regression model** (LRM), can be written in matrix form as follows:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (2)$$

where $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^\top$, $\boldsymbol{\varepsilon} \sim \mathcal{D}_n(\mathbf{0}, \sigma^2 \mathbf{I}_n)$, where $\mathbf{0}$ is the zero vector of dimension n , $\mathbf{I}_n$ is the identity matrix of order n and $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^\top$ is the design matrix.



Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;



Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;
- The errors

Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;
- The errors
 - have mean zero;

Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;
- The errors
 - have mean zero;
 - have constant variance;



Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;
- The errors
 - have mean zero;
 - have constant variance;
 - are uncorrelated.



Assumptions

In summary, the **assumptions** of a linear regression model are:

- The relationship between the response and predictor variables is linear;
- The errors
 - have mean zero;
 - have constant variance;
 - are uncorrelated.



Distribution of the MLE

Under the normal linear regression model (NLM), the MLE of β has the exact distribution

$$\hat{\beta} \sim \mathcal{N}_p \left(\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \right), \quad (3)$$

and, for the MLE of σ^2 ,



Distribution of the MLE

Under the normal linear regression model (NLM), the MLE of β has the exact distribution

$$\hat{\beta} \sim \mathcal{N}_p \left(\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \right), \quad (3)$$

and, for the MLE of σ^2 , we have

$$\frac{n\hat{\sigma}_{\text{MLE}}^2}{\sigma^2} \sim \chi_{n-p}^2.$$



Distribution of the MLE

Under the normal linear regression model (NLM), the MLE of β has the exact distribution

$$\hat{\beta} \sim \mathcal{N}_p \left(\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \right), \quad (3)$$

and, for the MLE of σ^2 , we have

$$\frac{n\hat{\sigma}_{\text{MLE}}^2}{\sigma^2} \sim \chi_{n-p}^2.$$

Note: $\hat{\sigma}_{\text{MLE}}^2$ can also be written as a function of the mean square error (MSE).



Distribution of the MLE

Under the normal linear regression model (NLM), the MLE of β has the exact distribution

$$\hat{\beta} \sim \mathcal{N}_p \left(\beta, \sigma^2 (\mathbf{X}^\top \mathbf{X})^{-1} \right), \quad (3)$$

and, for the MLE of σ^2 , we have

$$\frac{n\hat{\sigma}_{\text{MLE}}^2}{\sigma^2} \sim \chi_{n-p}^2.$$

Note: $\hat{\sigma}_{\text{MLE}}^2$ can also be written as a function of the mean square error (MSE).



Introduction

Motivation

Result (3) is very important, as it allows the construction of inferential procedures such as confidence intervals and hypothesis tests.

Outline

- 1 Introduction
- 2 Hypothesis testing**
- 3 F statistics
- 4 Chi-square statistics
- 5 Applications
- 6 References



Hypothesis testing

Let $c_{mm'}$ be the (m, m') th element of the matrix $\mathbf{C} = (\mathbf{X}^\top \mathbf{X})^{-1}$, with $m, m' = 1, 2, \dots, p$, i.e., $\text{Cov}(\hat{\beta}_m, \hat{\beta}_{m'}) = \sigma^2 c_{mm'}$. By (3), we have that,

$$\frac{\hat{\beta}_m - \beta_m}{\sqrt{\sigma^2 c_{mm}}} \sim \mathcal{N}(0, 1), \quad (4)$$

with $m = 1, 2, \dots, p$.

Hypothesis testing

Let $c_{mm'}$ be the (m, m') th element of the matrix $\mathbf{C} = (\mathbf{X}^\top \mathbf{X})^{-1}$, with $m, m' = 1, 2, \dots, p$, i.e., $\text{Cov}(\hat{\beta}_m, \hat{\beta}_{m'}) = \sigma^2 c_{mm'}$. By (3), we have that,

$$\frac{\hat{\beta}_m - \beta_m}{\sqrt{\sigma^2 c_{mm}}} \sim \mathcal{N}(0, 1), \quad (4)$$

with $m = 1, 2, \dots, p$.



Hypothesis tests

Since the statistic in (4) depends on the unknown parameter σ^2 , we need to replace it with an unbiased estimator, in this case, by the MSE, and thus we have

$$\frac{\hat{\beta}_m - \beta_m}{\sqrt{\text{MSE } c_{mm}}} \sim t(n - p), \quad (5)$$

with $m = 1, 2, \dots, p$.



Hypothesis tests

Since the statistic in (4) depends on the unknown parameter σ^2 , we need to replace it with an unbiased estimator, in this case, by the MSE, and thus we have

$$\frac{\hat{\beta}_m - \beta_m}{\sqrt{\text{MSE } c_{mm}}} \sim t(n - p), \quad (5)$$

with $m = 1, 2, \dots, p$.



Hypothesis test for β_m

Thus, from (5), we can construct a procedure for testing hypotheses about the parameter β_m . The procedure is as follows:

$$\begin{cases} \mathcal{H}: \beta_m = \beta_m^{(0)} \\ \mathcal{A}: \beta_m \neq \beta_m^{(0)} \end{cases},$$

where $\beta_m^{(0)}$ is a specified value, $m = 1, 2, \dots, p$.

Hypothesis test for β_m

Thus, from (5), we can construct a procedure for testing hypotheses about the parameter β_m . The procedure is as follows:

$$\begin{cases} \mathcal{H}: \beta_m = \beta_m^{(0)} \\ \mathcal{A}: \beta_m \neq \beta_m^{(0)} \end{cases},$$

where $\beta_m^{(0)}$ is a specified value, $m = 1, 2, \dots, p$.



Hypothesis test for β_m

And by (5), we can define the following test statistic as

$$t_c = \frac{\hat{\beta}_m - \beta_m}{\sqrt{\text{MSE } c_{mm}}}.$$



Hypothesis test for β_m

And by (5), we can define the following test statistic as

$$t_c = \frac{\hat{\beta}_m - \beta_m}{\sqrt{\text{MSE } c_{mm}}}.$$

Under $\mathcal{H}$, $t_c \sim t(n - p)$. Thus, the null hypothesis will be rejected, for a given nominal level α , if $|t_c|$ is greater than the $100(1 - \alpha/2)$ th percentile of the t_{n-p} distribution.



Hypothesis test for β_m

And by (5), we can define the following test statistic as

$$t_c = \frac{\hat{\beta}_m - \beta_m}{\sqrt{\text{MSE } c_{mm}}}.$$

Under $\mathcal{H}$, $t_c \sim t(n - p)$. Thus, the null hypothesis will be rejected, for a given nominal level α , if $|t_c|$ is greater than the $100(1 - \alpha/2)$ th percentile of the t_{n-p} distribution.



Hypothesis test for β_m

Notes:

- This is a partial test, since it assesses the contribution of x_m conditional on the other predictors in the model;



Hypothesis test for β_m

Notes:

- This is a partial test, since it assesses the contribution of x_m conditional on the other predictors in the model;
- In practice, the test is performed with $\beta_m^{(0)} = 0$, that is, to verify whether the coefficient β_m is significant, $m = 1, 2, \dots, p$. This is a test of the contribution of x_m given that the other predictors are in the model;



Hypothesis test for β_m

Notes:

- This is a partial test, since it assesses the contribution of x_m conditional on the other predictors in the model;
- In practice, the test is performed with $\beta_m^{(0)} = 0$, that is, to verify whether the coefficient β_m is significant, $m = 1, 2, \dots, p$. This is a test of the contribution of x_m given that the other predictors are in the model;
- One-sided tests can also be performed. It is only necessary to adjust the rejection region of $\mathcal{H}$.



Hypothesis test for β_m

Notes:

- This is a partial test, since it assesses the contribution of x_m conditional on the other predictors in the model;
- In practice, the test is performed with $\beta_m^{(0)} = 0$, that is, to verify whether the coefficient β_m is significant, $m = 1, 2, \dots, p$. This is a test of the contribution of x_m given that the other predictors are in the model;
- One-sided tests can also be performed. It is only necessary to adjust the rejection region of $\mathcal{H}$.



Outline

- 1 Introduction
- 2 Hypothesis testing
- 3 F statistics**
- 4 Chi-square statistics
- 5 Applications
- 6 References



Hypothesis test for a subset of regressors

Suppose that we can make the following partitions in the NLR,

$$\beta = (\beta_1^\top, \beta_2^\top)^\top \text{ and } \mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2],$$



Hypothesis test for a subset of regressors

Suppose that we can make the following partitions in the NLR,

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top \text{ and } \mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2],$$

where $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_q)^\top$ is a q -dimensional vector, $\boldsymbol{\beta}_2 = (\beta_{q+1}, \dots, \beta_p)^\top$ is a $(p-q)$ -dimensional parameter vector, $\mathbf{X}_1, \mathbf{X}_2$ being $n \times q$ and $n \times (p-q)$, respectively.



Hypothesis test for a subset of regressors

Suppose that we can make the following partitions in the NLR,

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1^\top, \boldsymbol{\beta}_2^\top)^\top \text{ and } \mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2],$$

where $\boldsymbol{\beta}_1 = (\beta_1, \dots, \beta_q)^\top$ is a q -dimensional vector, $\boldsymbol{\beta}_2 = (\beta_{q+1}, \dots, \beta_p)^\top$ is a $(p-q)$ -dimensional parameter vector, $\mathbf{X}_1$, $\mathbf{X}_2$ being $n \times q$ and $n \times (p-q)$, respectively.



Hypothesis test for a subset of regressors

The NLR in (2) can be written as follows:

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon. \quad (6)$$

Hypothesis test for a subset of regressors

The NLR in (2) can be written as follows:

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}. \quad (6)$$

We will refer to (6) as the full model. Now, if we assume that $\boldsymbol{\beta}_1 = \mathbf{0}_q$, we have a reduced model,



Hypothesis test for a subset of regressors

The NLR in (2) can be written as follows:

$$\mathbf{Y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon. \quad (6)$$

We will refer to (6) as the full model. Now, if we assume that $\beta_1 = \mathbf{0}_q$, we have a reduced model, i.e.,

$$\mathbf{Y} = \mathbf{X}_2\beta_2 + \varepsilon.$$



Hypothesis test for a subset of regressors

The NLR in (2) can be written as follows:

$$\mathbf{Y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon. \quad (6)$$

We will refer to (6) as the full model. Now, if we assume that $\beta_1 = \mathbf{0}_q$, we have a reduced model, i.e.,

$$\mathbf{Y} = \mathbf{X}_2\beta_2 + \varepsilon.$$



Partial F test

We can use the extra sum of squares to construct a procedure for testing whether $\beta_1 = \mathbf{0}_q$, as follows:

$$\begin{cases} \mathcal{H} : \beta_1 = \mathbf{0}_q \\ \mathcal{A} : \beta_1 \neq \mathbf{0}_q \end{cases} .$$

Partial F test

We can use the extra sum of squares to construct a procedure for testing whether $\beta_1 = \mathbf{0}_q$, as follows:

$$\begin{cases} \mathcal{H} : \beta_1 = \mathbf{0}_q \\ \mathcal{A} : \beta_1 \neq \mathbf{0}_q \end{cases}.$$

The test statistic is given by:

$$F_c = \frac{\text{MSR}(\mathbf{X}_1 | \mathbf{X}_2)}{\text{MSE}}.$$



Partial F test

We can use the extra sum of squares to construct a procedure for testing whether $\beta_1 = \mathbf{0}_q$, as follows:

$$\begin{cases} \mathcal{H} : \beta_1 = \mathbf{0}_q \\ \mathcal{A} : \beta_1 \neq \mathbf{0}_q \end{cases}.$$

The test statistic is given by:

$$F_c = \frac{\text{MSR}(\mathbf{X}_1 | \mathbf{X}_2)}{\text{MSE}}.$$



Partial F test

Under $\mathcal{H}$, $F_c \sim F(q, n-p)$. Thus, we will reject $\mathcal{H}$ if $F_c > F(1-\alpha; q, n-p)$, that is, the null hypothesis will be rejected, for a given nominal level α , if the test statistic is greater than the $100(1-\alpha)\%$ th percentile of the $F(q, n-p)$ distribution.



F test

Suppose that we wish to test

$$\left\{ \begin{array}{l} \mathcal{H} : \mathbf{B}\beta = \beta^{(0)} \\ \mathcal{A} : \mathbf{B}\beta \neq \beta^{(0)} \end{array} \right. ,$$

where $\mathbf{B}$ is an $s \times p$ matrix, with rank equal to $s \leq p$.

F test

Suppose that we wish to test

$$\left\{ \begin{array}{l} \mathcal{H} : \mathbf{B}\beta = \beta^{(0)} \\ \mathcal{A} : \mathbf{B}\beta \neq \beta^{(0)} \end{array} \right. ,$$

where $\mathbf{B}$ is an $s \times p$ matrix, with rank equal to $s \leq p$.

F test

Example 1. If $Y_\ell = \beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \beta_3 x_{\ell 3}$ and we are interested in testing

$$\begin{cases} \mathcal{H}: \beta_1 = \beta_2 \\ \mathcal{A}: \beta_1 \neq \beta_2 \end{cases},$$

F test

Example 1. If $Y_\ell = \beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \beta_3 x_{\ell 3}$ and we are interested in testing

$$\left\{ \begin{array}{l} \mathcal{H} : \beta_1 = \beta_2 \\ \mathcal{A} : \beta_1 \neq \beta_2 \end{array} \right. ,$$

we would have:

F test

Example 1. If $Y_\ell = \beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \beta_3 x_{\ell 3}$ and we are interested in testing

$$\begin{cases} \mathcal{H}: \beta_1 = \beta_2 \\ \mathcal{A}: \beta_1 \neq \beta_2 \end{cases},$$

we would have:

$$B = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}.$$

F test

Example 1. If $Y_\ell = \beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \beta_3 x_{\ell 3}$ and we are interested in testing

$$\begin{cases} \mathcal{H}: \beta_1 = \beta_2 \\ \mathcal{A}: \beta_1 \neq \beta_2 \end{cases},$$

we would have:

$$\mathbf{B} = \begin{pmatrix} 1 & -1 & 0 \end{pmatrix}.$$

F test

Example 2. If $Y_\ell = \beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \beta_3 x_{\ell 3}$ and we are interested in testing

$$\left\{ \begin{array}{l} \mathcal{H} : \beta_1 = \beta_2 \text{ e } \beta_3 = 0 \\ \mathcal{A} : \text{At least one of the equalities does not hold} \end{array} \right. ,$$



F test

Example 2. If $Y_\ell = \beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \beta_3 x_{\ell 3}$ and we are interested in testing

$$\left\{ \begin{array}{l} \mathcal{H} : \beta_1 = \beta_2 \text{ e } \beta_3 = 0 \\ \mathcal{A} : \text{At least one of the equalities does not hold} \end{array} \right. ,$$

we would have:

$$B = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} .$$

F test

Example 2. If $Y_\ell = \beta_1 x_{\ell 1} + \beta_2 x_{\ell 2} + \beta_3 x_{\ell 3}$ and we are interested in testing

$$\begin{cases} \mathcal{H}: \beta_1 = \beta_2 \text{ e } \beta_3 = 0 \\ \mathcal{A}: \text{At least one of the equalities does not hold} \end{cases},$$

we would have:

$$\mathbf{B} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$



F test

The procedure for testing:

$$\begin{cases} \mathcal{H} : \mathbf{B}\beta = \beta^{(0)} \\ \mathcal{A} : \mathbf{B}\beta \neq \beta^{(0)} \end{cases},$$

F test

The procedure for testing:

$$\begin{cases} \mathcal{H} : \mathbf{B}\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)} \\ \mathcal{A} : \mathbf{B}\boldsymbol{\beta} \neq \boldsymbol{\beta}^{(0)} \end{cases},$$

has the test statistic given by:

$$F_c = \frac{1}{s\hat{\sigma}^2} \left(\mathbf{B}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)} \right)^\top \left[\mathbf{B} \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{B}^\top \right]^{-1} \left(\mathbf{B}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)} \right).$$

F test

The procedure for testing:

$$\begin{cases} \mathcal{H} : \mathbf{B}\boldsymbol{\beta} = \boldsymbol{\beta}^{(0)} \\ \mathcal{A} : \mathbf{B}\boldsymbol{\beta} \neq \boldsymbol{\beta}^{(0)} \end{cases},$$

has the test statistic given by:

$$F_c = \frac{1}{s\hat{\sigma}^2} \left(\mathbf{B}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)} \right)^\top \left[\mathbf{B} \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{B}^\top \right]^{-1} \left(\mathbf{B}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^{(0)} \right).$$

F test

Under $\mathcal{H}$, $F_c \sim F(s, n-p)$. Thus, we will reject $\mathcal{H}$ if $F_c > F(1-\alpha; s, n-p)$, that is, the null hypothesis will be rejected, for a given nominal level α , if the test statistic is greater than the $100(1-\alpha)\%$ th percentile of the $F(s, n-p)$ distribution.



Outline

- 1 Introduction
- 2 Hypothesis testing
- 3 F statistics
- 4 Chi-square statistics**
- 5 Applications
- 6 References



Chi-square statistics

Suppose the same partitions discussed in the previous section and that our interest is in testing:

$$\begin{cases} \mathcal{H} : \beta_1 = \beta_1^{(0)} \\ \mathcal{A} : \beta_1 \neq \beta_1^{(0)} \end{cases},$$

Chi-square statistics

Suppose the same partitions discussed in the previous section and that our interest is in testing:

$$\begin{cases} \mathcal{H} : \beta_1 = \beta_1^{(0)} \\ \mathcal{A} : \beta_1 \neq \beta_1^{(0)} \end{cases},$$

where $\beta_1^{(0)}$ is a specified vector. β_1 and β_2 are referred to as the parameter vectors of interest and nuisance parameters, respectively.



Chi-square statistics

Suppose the same partitions discussed in the previous section and that our interest is in testing:

$$\begin{cases} \mathcal{H} : \beta_1 = \beta_1^{(0)} \\ \mathcal{A} : \beta_1 \neq \beta_1^{(0)} \end{cases},$$

where $\beta_1^{(0)}$ is a specified vector. β_1 and β_2 are referred to as the parameter vectors of interest and nuisance parameters, respectively.



Chi-square statistics

Let $\hat{\beta} = (\hat{\beta}_1^\top, \hat{\beta}_2^\top)^\top$ and $\tilde{\beta} = (\beta_1^{(0)\top}, \tilde{\beta}_2^\top)^\top$ be the unrestricted and restricted maximum likelihood estimates of β under $\mathcal{H}$, respectively.

The partition $\beta = (\beta_1^\top, \beta_2^\top)^\top$ leads to the corresponding partitions in the score vector $\mathbf{U}_\beta = (\mathbf{U}_{\beta_1}^\top, \mathbf{U}_{\beta_2}^\top)^\top$, with



Chi-square statistics

Let $\hat{\beta} = (\hat{\beta}_1^\top, \hat{\beta}_2^\top)^\top$ and $\tilde{\beta} = (\beta_1^{(0)\top}, \tilde{\beta}_2^\top)^\top$ be the unrestricted and restricted maximum likelihood estimates of β under $\mathcal{H}$, respectively.

The partition $\beta = (\beta_1^\top, \beta_2^\top)^\top$ leads to the corresponding partitions in the score vector $\mathbf{U}_\beta = (\mathbf{U}_{\beta_1}^\top, \mathbf{U}_{\beta_2}^\top)^\top$, with

$$\mathbf{U}_{\beta_1} = \frac{1}{\sigma^2} \mathbf{X}_1^\top (\mathbf{Y} - \mathbf{X}_1 \beta_1) \text{ and } \mathbf{U}_{\beta_2} = \frac{1}{\sigma^2} \mathbf{X}_2^\top (\mathbf{Y} - \mathbf{X}_2 \beta_2),$$

Chi-square statistics

Let $\hat{\beta} = (\hat{\beta}_1^\top, \hat{\beta}_2^\top)^\top$ and $\tilde{\beta} = (\beta_1^{(0)\top}, \tilde{\beta}_2^\top)^\top$ be the unrestricted and restricted maximum likelihood estimates of β under $\mathcal{H}$, respectively.

The partition $\beta = (\beta_1^\top, \beta_2^\top)^\top$ leads to the corresponding partitions in the score vector $\mathbf{U}_\beta = (\mathbf{U}_{\beta_1}^\top, \mathbf{U}_{\beta_2}^\top)^\top$, with

$$\mathbf{U}_{\beta_1} = \frac{1}{\sigma^2} \mathbf{X}_1^\top (\mathbf{Y} - \mathbf{X}_1 \beta_1) \text{ and } \mathbf{U}_{\beta_2} = \frac{1}{\sigma^2} \mathbf{X}_2^\top (\mathbf{Y} - \mathbf{X}_2 \beta_2),$$

Chi-square statistics

in the Fisher information matrix and its inverse, given by

$$\mathbf{K}_{\beta} = \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{pmatrix} = \frac{1}{\sigma^2} \begin{pmatrix} \mathbf{x}_1^{\top} \mathbf{x}_1 & \mathbf{x}_1^{\top} \mathbf{x}_2 \\ \mathbf{x}_2^{\top} \mathbf{x}_1 & \mathbf{x}_2^{\top} \mathbf{x}_2 \end{pmatrix} \quad (7)$$

$$\text{and } \mathbf{K}_{\beta}^{-1} = \sigma^2 \begin{pmatrix} \mathbf{K}^{11} & \mathbf{K}^{12} \\ \mathbf{K}^{21} & \mathbf{K}^{22} \end{pmatrix}.$$

The inverse matrix $\mathbf{K}_{\beta}^{-1}$ can be obtained using standard results for partitioned matrix inversion (Lu and Shiou, 2002).



Chi-square statistics

in the Fisher information matrix and its inverse, given by

$$\mathbf{K}_{\beta} = \begin{pmatrix} \mathbf{K}_{11} & \mathbf{K}_{12} \\ \mathbf{K}_{21} & \mathbf{K}_{22} \end{pmatrix} = \frac{1}{\sigma^2} \begin{pmatrix} \mathbf{x}_1^{\top} \mathbf{x}_1 & \mathbf{x}_1^{\top} \mathbf{x}_2 \\ \mathbf{x}_2^{\top} \mathbf{x}_1 & \mathbf{x}_2^{\top} \mathbf{x}_2 \end{pmatrix} \quad (7)$$

$$\text{and } \mathbf{K}_{\beta}^{-1} = \sigma^2 \begin{pmatrix} \mathbf{K}^{11} & \mathbf{K}^{12} \\ \mathbf{K}^{21} & \mathbf{K}^{22} \end{pmatrix}.$$

The inverse matrix $\mathbf{K}_{\beta}^{-1}$ can be obtained using standard results for partitioned matrix inversion (Lu and Shiou, 2002).



Chi-square statistics

The likelihood ratio (Wilks, 1938), Wald (Wald, 1943), score (Rao, 1948), and gradient (Terrell, 2002) statistics for testing $\mathcal{H}$ are defined, respectively, as

$$\begin{aligned} S_1 &= 2 \left[\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2) - \ell(\beta_1^{(0)}, \tilde{\beta}_2, \tilde{\sigma}^2) \right], \\ S_2 &= \left(\hat{\beta}_1 - \beta_1^{(0)} \right)^\top \left(\hat{\mathbf{K}}^{11} \right)^{-1} \left(\hat{\beta}_1 - \beta_1^{(0)} \right), \\ S_3 &= \tilde{\mathbf{U}}_{\beta_1}^\top \tilde{\mathbf{K}}^{11} \tilde{\mathbf{U}}_{\beta_1}, \\ S_4 &= \tilde{\mathbf{U}}_{\beta_1}^\top \left(\hat{\beta}_1 - \beta_1^{(0)} \right). \end{aligned} \tag{8}$$



Chi-square statistics

The likelihood ratio (Wilks, 1938), Wald (Wald, 1943), score (Rao, 1948), and gradient (Terrell, 2002) statistics for testing $\mathcal{H}$ are defined, respectively, as

$$\begin{aligned} S_1 &= 2 \left[\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2) - \ell(\beta_1^{(0)}, \tilde{\beta}_2, \tilde{\sigma}^2) \right], \\ S_2 &= \left(\hat{\beta}_1 - \beta_1^{(0)} \right)^\top \left(\hat{\mathbf{K}}^{11} \right)^{-1} \left(\hat{\beta}_1 - \beta_1^{(0)} \right), \\ S_3 &= \tilde{\mathbf{U}}_{\beta_1}^\top \tilde{\mathbf{K}}^{11} \tilde{\mathbf{U}}_{\beta_1}, \\ S_4 &= \tilde{\mathbf{U}}_{\beta_1}^\top \left(\hat{\beta}_1 - \beta_1^{(0)} \right). \end{aligned} \tag{8}$$



Chi-square statistics

Under $\mathcal{H}$, the four statistics have a chi-square distribution with q degrees of freedom. The null hypothesis will be rejected, for a given nominal level α and a specified statistic, if the test statistic is greater than the $100(1-\alpha)\%$ th percentile of the χ_q^2 distribution.



Chi-square statistics

Buse (1982) presents in a didactic way the geometric interpretation of the likelihood ratio, score, and Wald tests for the case of simple hypotheses; more recently, Montoril and Souza (2013) perform the same analysis including the gradient statistic.



Chi-square statistics

Specifically, for the NLM, the four statistics presented in (8) can be written as follows:

$$\begin{aligned} S_1 &= n \log \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \right) + \frac{1}{\tilde{\sigma}^2} \left(\mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\beta}} \right)^\top \left(\mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\beta}} \right) - \frac{1}{\hat{\sigma}^2} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right)^\top \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right), \\ S_2 &= \frac{1}{\hat{\sigma}^2} \left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)} \right)^\top \left(\hat{\mathbf{R}}^\top \hat{\mathbf{R}} \right) \left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)} \right), \\ S_3 &= \frac{1}{\tilde{\sigma}^2} \left(\mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\beta}} \right)^\top \mathbf{X}_1 \left(\tilde{\mathbf{R}}^\top \tilde{\mathbf{R}} \right)^{-1} \mathbf{X}_1^\top \left(\mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\beta}} \right), \\ S_4 &= \frac{1}{\tilde{\sigma}^2} \left(\mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\beta}} \right)^\top \mathbf{X}_1 \left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)} \right), \end{aligned} \tag{9}$$



Chi-square statistics

Specifically, for the NLM, the four statistics presented in (8) can be written as follows:

$$\begin{aligned} S_1 &= n \log \left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2} \right) + \frac{1}{\tilde{\sigma}^2} \left(\mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\beta}} \right)^\top \left(\mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\beta}} \right) - \frac{1}{\hat{\sigma}^2} \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right)^\top \left(\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}} \right), \\ S_2 &= \frac{1}{\hat{\sigma}^2} \left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)} \right)^\top \left(\hat{\mathbf{R}}^\top \hat{\mathbf{R}} \right) \left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)} \right), \\ S_3 &= \frac{1}{\tilde{\sigma}^2} \left(\mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\beta}} \right)^\top \mathbf{X}_1 \left(\tilde{\mathbf{R}}^\top \tilde{\mathbf{R}} \right)^{-1} \mathbf{X}_1^\top \left(\mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\beta}} \right), \\ S_4 &= \frac{1}{\tilde{\sigma}^2} \left(\mathbf{Y} - \mathbf{X} \tilde{\boldsymbol{\beta}} \right)^\top \mathbf{X}_1 \left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1^{(0)} \right), \end{aligned} \tag{9}$$



Chi-square statistics

where $\mathbf{R} = \mathbf{X}_1 - \mathbf{X}_2 \mathbf{C}$, with $\mathbf{C} = (\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{X}_1$. Since there are no parameters involved in $\mathbf{R}$, $\mathbf{R} = \hat{\mathbf{R}} = \tilde{\mathbf{R}}$.

The $q \times (p - q)$ matrix $\mathbf{C}$ can be interpreted as estimated regression coefficients, obtained from an NLM between the columns of $\mathbf{X}_1$ and the design matrix $\mathbf{X}_2$.



Chi-square statistics

where $\mathbf{R} = \mathbf{X}_1 - \mathbf{X}_2 \mathbf{C}$, with $\mathbf{C} = (\mathbf{X}_2^\top \mathbf{X}_2)^{-1} \mathbf{X}_2^\top \mathbf{X}_1$. Since there are no parameters involved in $\mathbf{R}$, $\mathbf{R} = \hat{\mathbf{R}} = \tilde{\mathbf{R}}$.

The $q \times (p - q)$ matrix $\mathbf{C}$ can be interpreted as estimated regression coefficients, obtained from an NLM between the columns of $\mathbf{X}_1$ and the design matrix $\mathbf{X}_2$.



Outline

- 1 Introduction
- 2 Hypothesis testing
- 3 F statistics
- 4 Chi-square statistics
- 5 Applications**
- 6 References



Application

Application 1. *Westwood Company* data set (Neter et al., 1983, p. 36).

After fitting the model,

$$\hat{Y}_\ell = 10 + 2x_{\ell 2},$$

where Y_ℓ : man-hours, $x_{\ell 2}$: lot size, $\ell = 1, 2, \dots, 10$.

Application

Application 1. *Westwood Company* data set (Neter et al., 1983, p. 36).

After fitting the model,

$$\hat{Y}_\ell = 10 + 2x_{\ell 2},$$

where Y_ℓ : man-hours, $x_{\ell 2}$: lot size, $\ell = 1, 2, \dots, 10$.



Application

For the data, using (3) and (5), we have:

Table 1: Parameter estimates.

Parameter	Estimate	SE	t_c
β_1	10	2.503	3.995
β_2	2	0.047	42.583

Application

For the data, using (3) and (5), we have:

Table 1: Parameter estimates.

Parameter	Estimate	SE	t_c
β_1	10	2.503	3.995
β_2	2	0.047	42.583

For $\alpha = 0.05$, the 95% quantile of a $t(8) = 2.31$. Therefore, the null hypotheses $\mathcal{H} : \beta_1 = 0$ and $\mathcal{H} : \beta_2 = 0$ are rejected at the 5% significance level.



Application

For the data, using (3) and (5), we have:

Table 1: Parameter estimates.

Parameter	Estimate	SE	t_c
β_1	10	2.503	3.995
β_2	2	0.047	42.583

For $\alpha = 0.05$, the 95% quantile of a $t(8) = 2.31$. Therefore, the null hypotheses $\mathcal{H} : \beta_1 = 0$ and $\mathcal{H} : \beta_2 = 0$ are rejected at the 5% significance level.



Application

Application 2. (Neter et al., 1983, p. 247) The *Zarthan Company* sells a special skin cream through fashion stores exclusively. It operates in 15 marketing districts and is interested in predicting district sales.

Sales (gross of jars; 1 gross = 12 dozen) are to be treated as the dependent variable Y , and target population (thousands of persons) and per capita discretionary income (dollars) as independent variables x_2 and x_3 , respectively.



Application

Application 2. (Neter et al., 1983, p. 247) The *Zarthan Company* sells a special skin cream through fashion stores exclusively. It operates in 15 marketing districts and is interested in predicting district sales.

Sales (gross of jars; 1 gross = 12 dozen) are to be treated as the dependent variable Y , and target population (thousands of persons) and per capita discretionary income (dollars) as independent variables x_2 and x_3 , respectively.



Application

After fitting the model, we have the following estimated model,

$$\hat{Y}_\ell = 3.453 + 0.496x_{\ell 2} + 0.009x_{\ell 3},$$

where Y_ℓ : sales (gross of jars), $x_{\ell 2}$: target population (thousands of persons), $x_{\ell 3}$: per capita discretionary income (dollars), $\ell = 1, 2, \dots, 15$.



Application

After fitting the model, we have the following estimated model,

$$\hat{Y}_\ell = 3.453 + 0.496x_{\ell 2} + 0.009x_{\ell 3},$$

where Y_ℓ : sales (gross of jars), $x_{\ell 2}$: target population (thousands of persons), $x_{\ell 3}$: per capita discretionary income (dollars), $\ell = 1, 2, \dots, 15$.



Application

For the data, using (3) and (5), we have:

Table 2: Parameter estimates.

Parameter	Estimate	SE	t_c
β_1	3.453	2.431	1.420
β_2	0.496	0.006	81.924
β_3	0.009	0.001	9.502

Application

For the data, using (3) and (5), we have:

Table 2: Parameter estimates.

Parameter	Estimate	SE	t_c
β_1	3.453	2.431	1.420
β_2	0.496	0.006	81.924
β_3	0.009	0.001	9.502

For $\alpha = 0.05$, the 95% quantile of a $t(12) = 2.179$. Therefore, the null hypotheses $\mathcal{H} : \beta_2 = 0$ and $\mathcal{H} : \beta_3 = 0$ are rejected at the 5% significance level, whereas $\mathcal{H} : \beta_1 = 0$ is not.



Application

For the data, using (3) and (5), we have:

Table 2: Parameter estimates.

Parameter	Estimate	SE	t_c
β_1	3.453	2.431	1.420
β_2	0.496	0.006	81.924
β_3	0.009	0.001	9.502

For $\alpha = 0.05$, the 95% quantile of a $t(12) = 2.179$. Therefore, the null hypotheses $\mathcal{H} : \beta_2 = 0$ and $\mathcal{H} : \beta_3 = 0$ are rejected at the 5% significance level, whereas $\mathcal{H} : \beta_1 = 0$ is not.



Outline

- 1 Introduction
- 2 Hypothesis testing
- 3 F statistics
- 4 Chi-square statistics
- 5 Applications
- 6 References**



References

- Buse, A. (1982). The likelihood ratio, wald and lagrange multiplier tests: an expository note. *The American Statistician*, 36(3):153–157.
- Lu, T.-T. and Shiou, S.-H. (2002). Inverses of 2×2 block matrices. *Computers & Mathematics with Applications*, 43(1):119–129.
- Montoril, M. H. and Souza, E. A. (2013). Estatística gradiente: propriedades e aplicações. *Revista Brasileira de Biometria*, 31(1):43–60.
- Neter, J., Wasserman, W., and Kutner, M. H. (1983). *Applied linear regression models*. Richard D. Irwin Inc, Homewood, Illinois.



References

- Rao, C. R. (1948). Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Proceedings of the Cambridge Philosophical Society*, 44(1):50–57.
- Terrell, G. R. (2002). The gradient statistic. *Computing Science and Statistics*, 34:206–215.
- Wald, A. (1943). Test of statistical hypotheses concerning several parameter when the number of observations is large. *Transactions of the American Mathematical Society*, 54(3):426–482.



References

Wilks, S. S. (1938). The large-sample distribution of likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, 9(1):60–62.



Thank you!

✉ tiago.magalhaes@ufjf.br

📄 ufjf.br/tiago_magalhaes

🌐 Department of Statistics, Room 319

