

Confidence intervals

Tiago M. Magalhães

Department of Statistics - ICE-UFJF

Juiz de Fora, October 02, 2026



Outline

- 1 Introduction
- 2 Confidence interval for β_m
- 3 Confidence intervals for μ_h and σ^2
- 4 Confidence intervals based on test statistics
- 5 Applications
- 6 References



Outline

- 1 Introduction
- 2 Confidence interval for β_m
- 3 Confidence intervals for μ_h and σ^2
- 4 Confidence intervals based on test 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:

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

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 Confidence interval for β_m
- 3 Confidence intervals for μ_h and σ^2
- 4 Confidence intervals based on test statistics
- 5 Applications
- 6 References



Confidence interval

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

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

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

Confidence interval

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

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

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

Confidence interval

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

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



Confidence interval

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

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

where $t(n - p)$ is the Student's t distribution (Gosset “Student”, 1908), with $n - p$ degrees of freedom, $m = 1, 2, \dots, p$.



Confidence interval

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

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

where $t(n - p)$ is the Student's t distribution (Gosset “Student”, 1908), with $n - p$ degrees of freedom, $m = 1, 2, \dots, p$.



Confidence interval for β_m

From (5), the confidence interval for β_m , with confidence level $(1 - \alpha)$, is given by

$$CI(\beta_m; 1 - \alpha) = \hat{\beta}_m \mp t(1 - \alpha/2; n - p) \sqrt{\text{MSE } c_{mm}},$$

Confidence interval for β_m

From (5), the confidence interval for β_m , with confidence level $(1 - \alpha)$, is given by

$$\text{CI}(\beta_m; 1 - \alpha) = \hat{\beta}_m \mp t(1 - \alpha/2; n - p) \sqrt{\text{MSE } c_{mm}},$$

where $t(1 - \alpha/2; n - p)$ is the $100(1 - \alpha/2)\%$ quantile of the Student's t distribution, with $n - p$ degrees of freedom, $m = 1, 2, \dots, p$.



Confidence interval for β_m

From (5), the confidence interval for β_m , with confidence level $(1 - \alpha)$, is given by

$$\text{CI}(\beta_m; 1 - \alpha) = \hat{\beta}_m \mp t(1 - \alpha/2; n - p) \sqrt{\text{MSE } c_{mm}},$$

where $t(1 - \alpha/2; n - p)$ is the $100(1 - \alpha/2)\%$ quantile of the Student's t distribution, with $n - p$ degrees of freedom, $m = 1, 2, \dots, p$.



Outline

- 1 Introduction
- 2 Confidence interval for β_m
- 3 Confidence intervals for μ_h and σ^2
- 4 Confidence intervals based on test statistics
- 5 Applications
- 6 References



Inference on μ_h

We may also be interested in estimating μ_h , the mean of Y_h , for a vector of observations $\mathbf{x}_h = (x_{h1}, x_{h2}, \dots, x_{hp})^\top$, which, by the invariance property of maximum likelihood estimators, has its estimator given by

$$\hat{\mu}_h = \mathbf{x}_h^\top \hat{\boldsymbol{\beta}}. \quad (6)$$



Inference on μ_h

We may also be interested in estimating μ_h , the mean of Y_h , for a vector of observations $\mathbf{x}_h = (x_{h1}, x_{h2}, \dots, x_{hp})^\top$, which, by the invariance property of maximum likelihood estimators, has its estimator given by

$$\hat{\mu}_h = \mathbf{x}_h^\top \hat{\boldsymbol{\beta}}. \quad (6)$$



Inference on μ_h

From (6), we have that

$$\mathbb{E}(\hat{\mu}_h) = \mathbb{E}(\mathbf{x}_h^\top \hat{\boldsymbol{\beta}}) = \mathbf{x}_h^\top \mathbb{E}(\hat{\boldsymbol{\beta}}) = \mathbf{x}_h^\top \boldsymbol{\beta} = \mu_h$$

Inference on μ_h

From (6), we have that

$$\mathbb{E}(\hat{\mu}_h) = \mathbb{E}(\mathbf{x}_h^\top \hat{\boldsymbol{\beta}}) = \mathbf{x}_h^\top \mathbb{E}(\hat{\boldsymbol{\beta}}) = \mathbf{x}_h^\top \boldsymbol{\beta} = \mu_h$$

and

$$\text{Var}(\hat{\mu}_h) = \text{Var}(\mathbf{x}_h^\top \hat{\boldsymbol{\beta}}) = \mathbf{x}_h^\top \text{Var}(\hat{\boldsymbol{\beta}}) \mathbf{x}_h = \sigma^2 \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h.$$



Inference on μ_h

From (6), we have that

$$\mathbb{E}(\hat{\mu}_h) = \mathbb{E}(\mathbf{x}_h^\top \hat{\boldsymbol{\beta}}) = \mathbf{x}_h^\top \mathbb{E}(\hat{\boldsymbol{\beta}}) = \mathbf{x}_h^\top \boldsymbol{\beta} = \mu_h$$

and

$$\text{Var}(\hat{\mu}_h) = \text{Var}(\mathbf{x}_h^\top \hat{\boldsymbol{\beta}}) = \mathbf{x}_h^\top \text{Var}(\hat{\boldsymbol{\beta}}) \mathbf{x}_h = \sigma^2 \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h.$$



Inference on μ_h

From (6), it follows that the distribution of $\hat{\mu}_h$ is given by:

$$\hat{\mu}_h \sim \mathcal{N} \left\{ \mu_h, \sigma^2 \mathbf{x}_h^\top \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{x}_h \right\}.$$

Inference on μ_h

From (6), it follows that the distribution of $\hat{\mu}_h$ is given by:

$$\hat{\mu}_h \sim \mathcal{N} \left\{ \mu_h, \sigma^2 \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right\}.$$

The estimator of the variance of $\hat{\mu}_h$ is given by:

$$\hat{\text{Var}}(\hat{\mu}_h) = \text{MSE} \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h.$$



Inference on μ_h

From (6), it follows that the distribution of $\hat{\mu}_h$ is given by:

$$\hat{\mu}_h \sim \mathcal{N} \left\{ \mu_h, \sigma^2 \mathbf{x}_h^\top \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{x}_h \right\}.$$

The estimator of the variance of $\hat{\mu}_h$ is given by:

$$\hat{\text{Var}}(\hat{\mu}_h) = \text{MSE} \mathbf{x}_h^\top \left(\mathbf{X}^\top \mathbf{X} \right)^{-1} \mathbf{x}_h.$$



Confidence interval for μ_h

The confidence interval for μ_h :

$$\text{CI}(\mu_h; 1 - \alpha) = \hat{\mu}_h \mp t(1 - \alpha/2; n - p) \sqrt{\text{MSE } \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h},$$



Confidence interval for μ_h

The confidence interval for μ_h :

$$\text{CI}(\mu_h; 1 - \alpha) = \hat{\mu}_h \mp t(1 - \alpha/2; n - p) \sqrt{\text{MSE } \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h},$$

where $t(1 - \alpha/2; n - p)$ is the $100(1 - \alpha/2)\%$ quantile of the Student's t distribution, with $n - p$ degrees of freedom.



Confidence interval for μ_h

The confidence interval for μ_h :

$$\text{CI}(\mu_h; 1 - \alpha) = \hat{\mu}_h \mp t(1 - \alpha/2; n - p) \sqrt{\text{MSE } \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h},$$

where $t(1 - \alpha/2; n - p)$ is the $100(1 - \alpha/2)\%$ quantile of the Student's t distribution, with $n - p$ degrees of freedom.



Confidence band for μ_h

As an alternative to the previous CI, the Working-Hotelling method (Working and Hotelling, 1929) is used to construct a confidence band (CB) for μ_h :

$$\text{CB}(\mu_h; 1 - \alpha) = \hat{\mu}_h \mp W \sqrt{\text{MSE } \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h},$$



Confidence band for μ_h

As an alternative to the previous CI, the Working-Hotelling method (Working and Hotelling, 1929) is used to construct a confidence band (CB) for μ_h :

$$\text{CB}(\mu_h; 1 - \alpha) = \hat{\mu}_h \mp W \sqrt{\text{MSE } \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h},$$

where $W^2 = pF(1 - \alpha; p, n - p)$ and $F(1 - \alpha; p, n - p)$ is the $100(1 - \alpha)\%$ quantile of the F distribution, with p and $n - p$ degrees of freedom.



Confidence band for μ_h

As an alternative to the previous CI, the Working-Hotelling method (Working and Hotelling, 1929) is used to construct a confidence band (CB) for μ_h :

$$\text{CB}(\mu_h; 1 - \alpha) = \hat{\mu}_h \mp W \sqrt{\text{MSE } \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h},$$

where $W^2 = pF(1 - \alpha; p, n - p)$ and $F(1 - \alpha; p, n - p)$ is the $100(1 - \alpha)\%$ quantile of the F distribution, with p and $n - p$ degrees of freedom.



Confidence interval for μ_h

Remark

The method is also known as the Working-Hotelling-Scheffé confidence band, due to the similarity between the Working-Hotelling method and the method of Scheffé (1959) in the analysis of contrasts in an analysis of variance.

Confidence interval for σ^2

The confidence interval for σ^2 :

$$CI(\sigma^2; 1 - \alpha) = \left[\frac{(n - p) \text{MSE}}{\chi^2(1 - \alpha/2; n - p)}; \frac{(n - p) \text{MSE}}{\chi^2(\alpha/2; n - p)} \right].$$

Confidence interval for σ^2

The confidence interval for σ^2 :

$$CI(\sigma^2; 1 - \alpha) = \left[\frac{(n - p) \text{MSE}}{\chi^2(1 - \alpha/2; n - p)}; \frac{(n - p) \text{MSE}}{\chi^2(\alpha/2; n - p)} \right].$$

where $\chi^2(\alpha/2; n - p)$ and $\chi^2(1 - \alpha/2; n - p)$ are, respectively, the $100\alpha/2\%$ and $100(1 - \alpha/2)\%$ quantiles of the chi-square distribution, with $n - p$ degrees of freedom.



Confidence interval for σ^2

The confidence interval for σ^2 :

$$CI(\sigma^2; 1 - \alpha) = \left[\frac{(n - p) \text{MSE}}{\chi^2(1 - \alpha/2; n - p)}; \frac{(n - p) \text{MSE}}{\chi^2(\alpha/2; n - p)} \right].$$

where $\chi^2(\alpha/2; n - p)$ and $\chi^2(1 - \alpha/2; n - p)$ are, respectively, the $100\alpha/2\%$ and $100(1 - \alpha/2)\%$ quantiles of the chi-square distribution, with $n - p$ degrees of freedom.



Outline

- 1 Introduction
- 2 Confidence interval for β_m
- 3 Confidence intervals for μ_h and σ^2
- 4 Confidence intervals based on test statistics
- 5 Applications
- 6 References



Confidence intervals based on test statistics

Let S_1 to S_4 be, respectively, the test statistics of the likelihood ratio (Wilks, 1938), Wald (Wald, 1943), score (Rao, 1948), and gradient (Terrell, 2002) tests. We have seen that, when $n \rightarrow \infty$,

$$S_i \sim \chi_q^2, \quad i = 1, 2, 3, 4,$$



Confidence intervals based on test statistics

Let S_1 to S_4 be, respectively, the test statistics of the likelihood ratio (Wilks, 1938), Wald (Wald, 1943), score (Rao, 1948), and gradient (Terrell, 2002) tests. We have seen that, when $n \rightarrow \infty$,

$$S_i \sim \chi_q^2, \quad i = 1, 2, 3, 4,$$

where q is the number of parameters in the null hypothesis, the dimension of the vector $\beta_1^{(0)}$.



Confidence intervals based on test statistics

Let S_1 to S_4 be, respectively, the test statistics of the likelihood ratio (Wilks, 1938), Wald (Wald, 1943), score (Rao, 1948), and gradient (Terrell, 2002) tests. We have seen that, when $n \rightarrow \infty$,

$$S_i \sim \chi_q^2, \quad i = 1, 2, 3, 4,$$

where q is the number of parameters in the null hypothesis, the dimension of the vector $\beta_1^{(0)}$.



Confidence intervals based on test statistics

Confidence intervals based on the statistics S_i are constructed by considering the set of parameter values satisfying

$$S_i \leq \chi^2(1 - \alpha; q),$$

Confidence intervals based on test statistics

Confidence intervals based on the statistics S_i are constructed by considering the set of parameter values satisfying

$$S_i \leq \chi^2(1 - \alpha; q),$$

where $\chi^2(1 - \alpha; q)$ is the $100(1 - \alpha)\%$ quantile of a chi-square distribution with q degrees of freedom, $i = 1, 2, 3, 4$.



Confidence intervals based on test statistics

Confidence intervals based on the statistics S_i are constructed by considering the set of parameter values satisfying

$$S_i \leq \chi^2(1 - \alpha; q),$$

where $\chi^2(1 - \alpha; q)$ is the $100(1 - \alpha)\%$ quantile of a chi-square distribution with q degrees of freedom, $i = 1, 2, 3, 4$.



Confidence intervals based on test statistics

Suppose that our interest is in testing,

$$\begin{cases} \mathcal{H}: \beta_1 = \beta^* \\ \mathcal{A}: \beta_1 \neq \beta^* \end{cases}, \quad (7)$$

Confidence intervals based on test statistics

Suppose that our interest is in testing,

$$\begin{cases} \mathcal{H}: \beta_1 = \beta^* \\ \mathcal{A}: \beta_1 \neq \beta^* \end{cases}, \quad (7)$$

the likelihood ratio (LR) statistic can be written as

$$S_1 = 2 \left[\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2) - \ell(\beta^*, \tilde{\beta}_2, \tilde{\sigma}^2) \right].$$



Confidence intervals based on test statistics

Suppose that our interest is in testing,

$$\begin{cases} \mathcal{H}: \beta_1 = \beta^* \\ \mathcal{A}: \beta_1 \neq \beta^* \end{cases}, \quad (7)$$

the likelihood ratio (LR) statistic can be written as

$$S_1 = 2 \left[\ell(\hat{\beta}_1, \hat{\beta}_2, \hat{\sigma}^2) - \ell(\beta^*, \tilde{\beta}_2, \tilde{\sigma}^2) \right].$$



CI based on the LR statistic

The CI based on the LR statistic is obtained as follows, (algorithm adapted from Hosmer et al., 2013, p. 19):

- 1 define the set of values of β^* such that $\mathcal{H}$, in (7), is not rejected;



CI based on the LR statistic

The CI based on the LR statistic is obtained as follows, (algorithm adapted from Hosmer et al., 2013, p. 19):

- ① define the set of values of β^* such that $\mathcal{H}$, in (7), is not rejected;
- ② the lower and upper limits of the interval will be the 2.5th and 97.5th percentiles of the resulting set of β^* values.



CI based on the LR statistic

The CI based on the LR statistic is obtained as follows, (algorithm adapted from Hosmer et al., 2013, p. 19):

- ① define the set of values of β^* such that $\mathcal{H}$, in (7), is not rejected;
- ② the lower and upper limits of the interval will be the 2.5th and 97.5th percentiles of the resulting set of β^* values.



CI based on the LR statistic

Remarks:

- 1 The procedure is valid for all β_m , $m = 1, 2, \dots, p$;

CI based on the LR statistic

Remarks:

- ① The procedure is valid for all β_m , $m = 1, 2, \dots, p$;
- ② To construct confidence regions, it is only necessary to change the restrictions in $\mathcal{H}$. For example, $\mathcal{H} : \beta_m = \beta^*, \beta_{m'} = \beta^{**}$;



CI based on the LR statistic

Remarks:

- ① The procedure is valid for all β_m , $m = 1, 2, \dots, p$;
- ② To construct confidence regions, it is only necessary to change the restrictions in $\mathcal{H}$. For example, $\mathcal{H} : \beta_m = \beta^*, \beta_{m'} = \beta^{**}$;
- ③ As mentioned, CIs based on the Wald, score, and gradient statistics can also be constructed. In some cases, they may have a closed-form expression.



CI based on the LR statistic

Remarks:

- ① The procedure is valid for all β_m , $m = 1, 2, \dots, p$;
- ② To construct confidence regions, it is only necessary to change the restrictions in $\mathcal{H}$. For example, $\mathcal{H} : \beta_m = \beta^*, \beta_{m'} = \beta^{**}$;
- ③ As mentioned, CIs based on the Wald, score, and gradient statistics can also be constructed. In some cases, they may have a closed-form expression.



Outline

- 1 Introduction
- 2 Confidence interval for β_m
- 3 Confidence intervals for μ_h and σ^2
- 4 Confidence intervals based on test statistics
- 5 Applications**
- 6 References



Application

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

After fitting the model,

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

Application

Application 1. *Westwood Company* dataset (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, $\ell = 1, 2, \dots, 10$, $\text{MSE} = 7.5$.



Application

Application 1. *Westwood Company* dataset (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, $\ell = 1, 2, \dots, 10$, $\text{MSE} = 7.5$.



Application

We also 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 $\alpha = 0.05$, $t(0.975; 8) = 2.31$. Thus, the t -based confidence interval is given by:

$$CI(\beta_1; 0.95) = [4.23; 15.77] ,$$

$$CI(\beta_2; 0.95) = [1.89; 2.11] .$$



Application

Application 2. (Neter et al., 1983, p. 247) *Zarthan Company* dataset.

After fitting the model,

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

Application

Application 2. (Neter et al., 1983, p. 247) *Zarthan Company* dataset.

After fitting the 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), $\ell = 1, 2, \dots, 15$, and $\text{MSE} = 4.74$.



Application

Application 2. (Neter et al., 1983, p. 247) *Zarthan Company* dataset.

After fitting the 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), $\ell = 1, 2, \dots, 15$, and $\text{MSE} = 4.74$.



Application

We also 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 $\alpha = 0.05$, $t(0.975; 12) = 2.179$. Thus, the t -based confidence interval is given by:

$$CI(\beta_1; 0.95) = [-1.843; 8.749],$$

$$CI(\beta_2; 0.95) = [0.483; 0.509],$$

$$CI(\beta_3; 0.95) = [0.007; 0.011].$$



Application

For $\alpha = 0.05$, $t(0.975; 12) = 2.179$. Thus, the t -based confidence interval is given by:

$$CI(\beta_1; 0.95) = [-1.843; 8.749],$$

$$CI(\beta_2; 0.95) = [0.483; 0.509],$$

$$CI(\beta_3; 0.95) = [0.007; 0.011].$$

As we can see, the CI for β_1 contains zero, that is, the intercept is not statistically significant.



Application

For $\alpha = 0.05$, $t(0.975; 12) = 2.179$. Thus, the t -based confidence interval is given by:

$$CI(\beta_1; 0.95) = [-1.843; 8.749],$$

$$CI(\beta_2; 0.95) = [0.483; 0.509],$$

$$CI(\beta_3; 0.95) = [0.007; 0.011].$$

As we can see, the CI for β_1 contains zero, that is, the intercept is not statistically significant.



Application

Table 3 presents the fit considering a model without an intercept:

Table 3: Parameter estimates.

Parameter	Estimate	SE	t_c
β_2	0.495	0.006	78.99
β_3	0.010	0.001	18.83

MSE = 5.111.



Application

For $\alpha = 0.05$, $t(0.975; 12) = 2.18$. Thus, the t -based confidence interval is given by:

$$CI(\beta_2; 0.95) = [0.482; 0.509] ,$$

$$CI(\beta_3; 0.95) = [0.009; 0.012] .$$



Application

For $\alpha = 0.05$, $t(0.975; 12) = 2.18$. Thus, the t -based confidence interval is given by:

$$CI(\beta_2; 0.95) = [0.482; 0.509],$$

$$CI(\beta_3; 0.95) = [0.009; 0.012].$$



Outline

- 1 Introduction
- 2 Confidence interval for β_m
- 3 Confidence intervals for μ_h and σ^2
- 4 Confidence intervals based on test statistics
- 5 Applications
- 6 References



References

Dorfman, R. (1938). A note on the δ -method for finding variance formulae. *The Biometric Bulletin*, 1:129–137.

Gosset “Student”, W. S. (1908). The probable error of a mean. *Biometrika*, 6(1):1–25.

Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied logistic regression*. Wiley, New Jersey, 3rd edition.

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.
- Scheffé, H. (1959). *The Analysis of Variance*. Wiley, New York.
- 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.
- Working, H. and Hotelling, H. (1929). Applications of the theory of error to the interpretation of trends. *Journal of the American Statistical Association*, 24(165A):73–85.



Thank you!

✉ tiago.magalhaes@ufjf.br

📄 ufjf.br/tiago_magalhaes

🌐 Department of Statistics, Room 319

