

Prediction interval

Tiago M. Magalhães

Department of Statistics - ICE-UFJF

Juiz de Fora, October 07, 2026



Outline

- 1 Introduction
- 2 Prediction of new observations
- 3 Applications
- 4 References



Outline

- 1 Introduction
- 2 Prediction of new observations
- 3 Applications
- 4 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).



Outline

- 1 Introduction
- 2 Prediction of new observations
- 3 Applications
- 4 References



Prediction of new observations

An important application of regression models is the prediction of a new observation $Y_{h(\text{new})}$ corresponding to a given vector $\mathbf{x}_h$. Then

$$\hat{Y}_{h(\text{new})} = \mathbf{x}_h^\top \hat{\boldsymbol{\beta}},$$



Prediction of new observations

An important application of regression models is the prediction of a new observation $Y_{h(\text{new})}$ corresponding to a given vector $\mathbf{x}_h$. Then

$$\hat{Y}_{h(\text{new})} = \mathbf{x}_h^\top \hat{\boldsymbol{\beta}},$$

is the point prediction for a new value of the response $Y_{h(\text{new})}$.



Prediction of new observations

An important application of regression models is the prediction of a new observation $Y_{h(\text{new})}$ corresponding to a given vector $\mathbf{x}_h$. Then

$$\hat{Y}_{h(\text{new})} = \mathbf{x}_h^\top \hat{\boldsymbol{\beta}},$$

is the point prediction for a new value of the response $Y_{h(\text{new})}$.



Prediction of new observations

Consider constructing an interval estimate for the future observation $Y_{h(\text{new})}$. The confidence interval (CI) for the mean response at $\mathbf{x}_h$ is not appropriate for this problem.

This is because it provides an interval for the mean response, rather than an interval that accounts for the variability of a future observation.



Prediction of new observations

Consider constructing an interval estimate for the future observation $Y_{h(\text{new})}$. The confidence interval (CI) for the mean response at $\mathbf{x}_h$ is not appropriate for this problem.

This is because it provides an interval for the mean response, rather than an interval that accounts for the variability of a future observation.



Prediction of new observations

Define the random variable (RV) as:

$$\begin{aligned}\psi &= Y_h - \hat{Y}_h \\ &= Y_h - \mathbf{x}_h^\top \hat{\boldsymbol{\beta}}.\end{aligned}$$

Prediction of new observations

Define the random variable (RV) as:

$$\begin{aligned}\psi &= Y_h - \hat{Y}_h \\ &= Y_h - \mathbf{x}_h^\top \hat{\boldsymbol{\beta}}.\end{aligned}$$

Prediction of new observations

We have that,

$$\mathbb{E}(\psi) = \mathbb{E}(Y_h - \mathbf{x}_h^\top \hat{\beta}) = \mathbb{E}(Y_h) - \mathbf{x}_h^\top \mathbb{E}(\hat{\beta}) = \mathbf{x}_h^\top \beta - \mathbf{x}_h^\top \beta = 0,$$

$$\begin{aligned}\text{Var}(\psi) &= \text{Var}(Y_h - \mathbf{x}_h^\top \hat{\beta}) \stackrel{*}{=} \text{Var}(Y_h) + \mathbf{x}_h^\top \text{Var}(\hat{\beta}) \mathbf{x}_h \\ &= \sigma^2 + \sigma^2 \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \\ &= \sigma^2 \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right].\end{aligned}$$



Prediction of new observations

We have that,

$$\mathbb{E}(\psi) = \mathbb{E}(Y_h - \mathbf{x}_h^\top \hat{\beta}) = \mathbb{E}(Y_h) - \mathbf{x}_h^\top \mathbb{E}(\hat{\beta}) = \mathbf{x}_h^\top \beta - \mathbf{x}_h^\top \beta = 0,$$

$$\begin{aligned}\text{Var}(\psi) &= \text{Var}(Y_h - \mathbf{x}_h^\top \hat{\beta}) \stackrel{*}{=} \text{Var}(Y_h) + \mathbf{x}_h^\top \text{Var}(\hat{\beta}) \mathbf{x}_h \\ &= \sigma^2 + \sigma^2 \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \\ &= \sigma^2 \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right].\end{aligned}$$

Prediction of new observations

∗: Under the regression model assumptions, the future observation Y_h is independent of $\hat{Y}_h$. Indeed, Y_h represents a future, unobserved random variable, whereas $\hat{Y}_h$ is a function of the observed data through $\hat{\beta}$.

Since Y_h and $\hat{Y}_h$ are normally distributed,

$$Y_h - \hat{Y}_h \sim \mathcal{N} \left\{ 0, \sigma^2 \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right] \right\}. \quad (4)$$



Prediction of new observations

∗: Under the regression model assumptions, the future observation Y_h is independent of $\hat{Y}_h$. Indeed, Y_h represents a future, unobserved random variable, whereas $\hat{Y}_h$ is a function of the observed data through $\hat{\beta}$.

Since Y_h and $\hat{Y}_h$ are normally distributed,

$$Y_h - \hat{Y}_h \sim \mathcal{N} \left\{ 0, \sigma^2 \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right] \right\}. \quad (4)$$



Prediction of new observations

From (4),

$$\frac{Y_h - \hat{Y}_h}{\sqrt{\sigma^2 \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right]}} \sim \mathcal{N}(0, 1), \quad (5)$$

Prediction of new observations

From (4),

$$\frac{Y_h - \hat{Y}_h}{\sqrt{\sigma^2 \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right]}} \sim \mathcal{N}(0, 1), \quad (5)$$

Prediction of new observations

Since the statistic in (5) depends on the unknown parameter σ^2 , we replace it with the unbiased estimator MSE, and thus, we have that

$$\frac{Y_h - \hat{Y}_h}{\sqrt{\text{MSE} \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right]}} \sim t(n - p), \quad (6)$$

Prediction of new observations

Since the statistic in (5) depends on the unknown parameter σ^2 , we replace it with the unbiased estimator MSE, and thus, we have that

$$\frac{Y_h - \hat{Y}_h}{\sqrt{\text{MSE} \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right]}} \sim t(n - p), \quad (6)$$

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



Prediction of new observations

Since the statistic in (5) depends on the unknown parameter σ^2 , we replace it with the unbiased estimator MSE, and thus, we have that

$$\frac{Y_h - \hat{Y}_h}{\sqrt{\text{MSE} \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right]}} \sim t(n - p), \quad (6)$$

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



Prediction interval for a new observation

From (6), the prediction interval (PI) for $Y_{h(\text{new})}$, with confidence level $(1 - \alpha)$, is given by

$$\text{PI}(Y_{h(\text{new})}; 1 - \alpha) = \hat{Y}_h \mp t(1 - \alpha/2; n - p) \sqrt{\text{MSE} \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right]},$$



Prediction interval for a new observation

From (6), the prediction interval (PI) for $Y_{h(\text{new})}$, with confidence level $(1 - \alpha)$, is given by

$$\text{PI}(Y_{h(\text{new})}; 1 - \alpha) = \hat{Y}_h \mp t(1 - \alpha/2; n - p) \sqrt{\text{MSE} \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right]},$$

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



Prediction interval for a new observation

From (6), the prediction interval (PI) for $Y_{h(\text{new})}$, with confidence level $(1 - \alpha)$, is given by

$$\text{PI}(Y_{h(\text{new})}; 1 - \alpha) = \hat{Y}_h \mp t(1 - \alpha/2; n - p) \sqrt{\text{MSE} \left[1 + \mathbf{x}_h^\top (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{x}_h \right]},$$

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



Considerations

In a regression analysis, prediction is often overlooked when the focus is on inference, i.e., when the important aspect is the interpretation of the regression coefficients.

There is also a limitation of the procedure: reliable prediction of the response variable depends on the new covariate values being within the range of the observed data; see Section 3.9 of Montgomery et al. (2021).



Considerations

In a regression analysis, prediction is often overlooked when the focus is on inference, i.e., when the important aspect is the interpretation of the regression coefficients.

There is also a limitation of the procedure: reliable prediction of the response variable depends on the new covariate values being within the range of the observed data; see Section 3.9 of Montgomery et al. (2021).



Considerations

If the main focus is prediction, additional procedures need to be used. The first step is to split the dataset into **training** and **validation** sets. The parameter estimates are obtained using the training set.



Considerations

The partitioning of the dataset, together with the subsequent prediction procedures, is commonly used in **supervised learning**, which is part of the broader fields of Statistical Learning and Machine Learning.



Outline

- 1 Introduction
- 2 Prediction of new observations
- 3 Applications**
- 4 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, $\ell = 1, 2, \dots, 10$, $\text{MSE} = 7.5$.



Example

We also have that:

Table 1: Parameter estimates.

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

Critical region for $\alpha = 5\%$: $|t_c| > 2.306$, with $MSE = 7.5$.



Example

We also have that:

Table 1: Parameter estimates.

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

Critical region for $\alpha = 5\%$: $|t_c| > 2.306$, with $MSE = 7.5$.

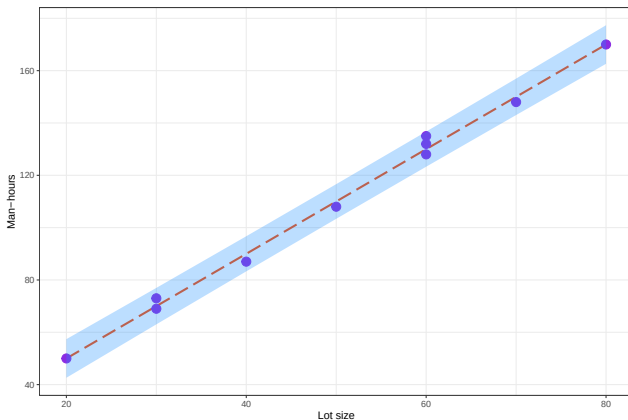


Figure 1: Scatterplot, fitted line, and PI for the *Westwood Company* data.

Application

Application 2. A dataset from 1920, available in Ezekiel (1930), relating car speed (predictor variable) to stopping distance (response variable). After fitting the model, we have the following estimated model,

$$\hat{Y}_\ell = -17.58 + 3.93x_{\ell 2},$$



Application

Application 2. A dataset from 1920, available in Ezekiel (1930), relating car speed (predictor variable) to stopping distance (response variable). After fitting the model, we have the following estimated model,

$$\hat{Y}_\ell = -17.58 + 3.93x_{\ell 2},$$

where Y_ℓ : stopping distance (feet), $x_{\ell 2}$: speed (miles per hour), $\ell = 1, 2, \dots, 50$.



Application

Application 2. A dataset from 1920, available in Ezekiel (1930), relating car speed (predictor variable) to stopping distance (response variable). After fitting the model, we have the following estimated model,

$$\hat{Y}_\ell = -17.58 + 3.93x_{\ell 2},$$

where Y_ℓ : stopping distance (feet), $x_{\ell 2}$: speed (miles per hour), $\ell = 1, 2, \dots, 50$.



Example

We also have that:

Table 2: Parameter estimates.

Parameter	Estimate	SE	t_c
β_1	-17.58	6.76	-2.60
β_2	3.93	0.42	9.46

Critical region for $\alpha = 5\%$: $|t_c| > 2.011$, with $MSE = 236.53$.



Example

We also have that:

Table 2: Parameter estimates.

Parameter	Estimate	SE	t_c
β_1	-17.58	6.76	-2.60
β_2	3.93	0.42	9.46

Critical region for $\alpha = 5\%$: $|t_c| > 2.011$, with $MSE = 236.53$.



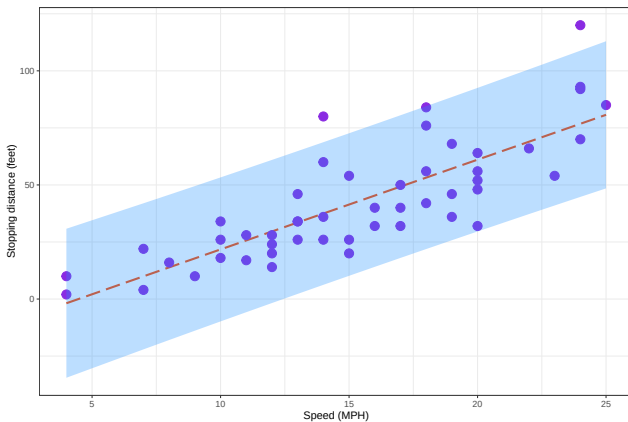


Figure 2: Scatterplot, fitted line, and PI for the data.

Outline

- 1 Introduction
- 2 Prediction of new observations
- 3 Applications
- 4 References



References

Ezekiel, M. (1930). *Methods of Correlation Analysis*. Wiley, New York.

Gosset “Student”, W. S. (1908). The probable error of a mean.

Biometrika, 6(1):1–25.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021). *Introduction to linear regression analysis*. Wiley, New York, 6th edition.

Neter, J., Wasserman, W., and Kutner, M. H. (1983). *Applied linear regression models*. Richard D. Irwin Inc, Homewood, Illinois.



Thank you!

✉ tiago.magalhaes@ufjf.br

📄 ufjf.br/tiago_magalhaes

🌐 Department of Statistics, Room 319

