# What is Prediction Error in Statistics? (Definition & Examples)

Authored by

**stats writer**

July 1, 2024

Prediction error in statistics refers to the discrepancy between the actual and predicted values of a variable or outcome. It is a measure of how accurately a statistical model or method can predict future observations based on past data. In other words, it is the difference between what is expected and what actually happens. Prediction error is a crucial aspect of statistical analysis as it helps in evaluating the performance and accuracy of a model. A high prediction error indicates that the model is not able to effectively capture the underlying patterns and relationships in the data. Examples of prediction error include the difference between the actual and predicted sales figures, the deviation between the forecasted and actual stock prices, or the disparity between the expected and observed weather conditions. By understanding and minimizing prediction error, statisticians can improve the reliability and validity of their models and make more accurate predictions.

## What is Prediction Error in Statistics? (Definition & Examples)

## In statistics, prediction error refers to the difference between the predicted values made by some model and the actual values.

## Prediction error is often used in two settings:

## 1. Linear regression: Used to predict the value of some continuous response variable.

## We typically measure the prediction error of a linear regression model with a metric known as , which stands for root mean squared error.

## It is calculated as:

**RMSE = √Σ(?i - yi)2 / n**

**where:**

**Σ is a symbol that means "sum"?i is the predicted value for the ith observationyi is the observed value for the ith observationn is the sample size**

**2. Logistic Regression: Used to predict the value of some binary response variable.**

**One common way to measure the prediction error of a logistic regression model is with a metric known as the total misclassification rate.**

**It is calculated as:**

**Total misclassification rate = (# incorrect predictions / # total predictions)**

**The lower the value for the misclassification rate, the better the model is able to predict the outcomes of the response variable.**

**The following examples show how to calculate prediction error for both a linear regression model and**

## a logistic regression model in practice.

### Example 1: Calculating Prediction Error in Linear Regression

**Suppose we use a regression model to predict the number of points that 10 players will score in a basketball game.**

**The following table shows the predicted points from the model vs. the actual points the players scored:**

| Predicted Points ( $\hat{y}_i$ ) | Actual points ( $y_i$ ) |
|---|---|
| 14 | 12 |
| 15 | 15 |
| 18 | 20 |
| 19 | 16 |
| 25 | 20 |
| 18 | 19 |
| 12 | 16 |
| 12 | 20 |
| 15 | 16 |
| 22 | 16 |



**We would calculate the root mean squared error (RMSE) as:**

**RMSE = √Σ(?i - yi)2 / nRMSE = √(((14-12)2+(15-15)2+(18-20)2+(19-16)2+(25-20)2+(18-19)**

**2+(12-16)2+(12-20)2+(15-16)2+(22-16)2) / 10)RMSE = 4**

**The root mean squared error is 4. This tells us that the average deviation between the predicted points scored and the actual points scored is 4.**

## Related:

### Example 2: Calculating Prediction Error in Logistic Regression

**Suppose we use a logistic regression model to predict whether or not 10 college basketball players will get drafted into the NBA.**

**The following table shows the predicted outcome for each player vs. the actual outcome (1 = Drafted, 0 = Not Drafted):**

| Prediction | Actual |
|---|---|
| 1 | 0 |
| 1 | 1 |
| 0 | 0 |
| 1 | 1 |
| 1 | 1 |
| 0 | 0 |
| 0 | 1 |
| 1 | 1 |
| 1 | 0 |
| 0 | 1 |

**We would calculate the total misclassification rate as:**

**Total misclassification rate = (# incorrect predictions / # total predictions)Total misclassification rate = 4/10Total misclassification rate = 40%**

**The total misclassification rate is 40%.**

**This value is quite high, which indicates that the model doesn't do a very good job of predicting whether or not a player will get drafted.**

### Additional Resources

**The following tutorials provide an introduction to different types of regression methods:**