How can we predict values in R using a multiple regression model?

Authored by

stats writer

June 29, 2024

Multiple regression is a statistical method used to predict values by analyzing the relationship between a dependent variable and two or more independent variables. In R, this can be achieved by creating a multiple regression model which takes into account the effect of each independent variable on the dependent variable. This model can then be used to make predictions by plugging in values for the independent variables. By utilizing this method, we can accurately predict values in R and gain insights into the relationship between variables.

Predict Values in R Using Multiple Regression Model

You can use the following basic syntax to predict values in R using a fitted multiple linear regression model:

```
#define new observation
new <- data.frame(x1=c(5), x2=c(10), x3=c(12.5))

#use fitted model to predict the response value for the new observation
predict(model, newdata=new)
```

The following example shows how to use this function in practice.

Example: Predict Values Using Fitted Multiple Linear Regression Model

Suppose we have the following dataset in R that contains information about basketball players:

```
#create data frame
df <- data.frame(rating=c(67, 75, 79, 85, 90, 96, 97),
points=c(8, 12, 16, 15, 22, 28, 24),
assists=c(4, 6, 6, 5, 3, 8, 7),
rebounds=c(1, 4, 3, 3, 2, 6, 7))

#view data frame
df

rating points assists rebounds
1 67 8 4 1
2 75 12 6 4
3 79 16 6 3
4 85 15 5 3
5 90 22 3 2
6 96 28 8 6
7 97 24 7 7
```

Now suppose we fit a multiple linear regression model using points, assists, and rebounds as predictor variables and rating as the :

```
#fit multiple linear regression model
model <- lm(rating ~ points + assists + rebounds,
data=df)
```

```
#view model summary
summary(model)

Call:
lm(formula = rating ~ points + assists + rebounds, data
= df)

Residuals:
1 2 3 4 5 6 7
-1.5902 -1.7181 0.2413 4.8597 -1.0201 -0.6082 -0.1644

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 66.4355 6.6932 9.926 0.00218 **
points 1.2152 0.2788 4.359 0.02232 *
assists -2.5968 1.6263 -1.597 0.20860
rebounds 2.8202 1.6118 1.750 0.17847
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.193 on 3 degrees of freedom
Multiple R-squared: 0.9589, Adjusted R-squared: 0.9179
F-statistic: 23.35 on 3 and 3 DF, p-value: 0.01396
```

From the values in the Estimate column, we can write

the fitted regression model:

Rating = 66.4355 + 1.2151(points) - 2.5968(assists) + 2.8202(rebounds)

We can use the following code to predict the rating of a new player who has 20 points, 5 assists, and 2 rebounds:

```
#define new player
new <- data.frame(points=c(20), assists=c(5), rebounds=c(2))

#use the fitted model to predict the rating for the new player
predict(model, newdata=new)

1
83.39607
```

The model predicts that this new player will have a rating of 83.39607.

We can confirm this is correct by plugging in the values for the new player into the fitted regression equation:

Rating = 66.4355 + 1.2151(points) - 2.5968(assists) + 2.8202(rebounds)Rating = 66.4355 + 1.2151(20) - 2.5968(5) + 2.8202(2)Rating = 83.39

This matches the value we calculated using the predict() function in R.

Additional Resources