

How can the Method of Least Squares be utilized in R?

Authored by
stats writer

June 26, 2024

RECOMMENDED CITATION

stats writer (2024). *How can the Method of Least Squares be utilized in R?*.

PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=153134>

The Method of Least Squares is a statistical technique used to find the best fitting line or curve for a given set of data points. In R, this method can be utilized through the use of the "lm" function, which performs linear regression analysis and calculates the coefficients of the line of best fit. This method is commonly used to analyze and model relationships between variables in various fields such as economics, finance, and engineering. By utilizing the Method of Least Squares in R, users can effectively analyze and interpret data, make predictions, and draw conclusions based on the fitted model. This makes it a valuable tool for researchers, analysts, and data scientists. Overall, the Method of Least Squares in R provides a powerful and efficient way to analyze and understand data patterns.

Use Method of Least Squares in R

The method of least squares is a method we can use to find the regression line that best fits a given dataset.

The following video provides a brief explanation of this method:

To use the method of least squares to fit a regression line in R, we can use the lm() function.

This function uses the following basic syntax:

```
model <- lm(response ~ predictor, data=df)
```

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

Example: Method of Least Squares in R

Suppose we have the following data frame in R that shows the number of hours studied and the corresponding exam score for 15 students in some class:

```
#create data frame
```

```
df <- data.frame(hours=c(1, 2, 4, 5, 5, 6, 6, 7, 8, 10, 11,  
11, 12, 12, 14),  
score=c(64, 66, 76, 73, 74, 81, 83, 82, 80, 88, 84, 82, 91,  
93, 89))
```

```
#view first six rows of data frame
```

```
head(df)
```

```
hours score
```

```
1 1 64
```

```
2 2 66
```

```
3 4 76
```

```
4 5 73
```

```
5 5 74
```

```
6 6 81
```

We can use the `lm()` function to use the method of least

squares to fit a regression line to this data:

#use method of least squares to fit regression line

model <- lm(score ~ hours, data=df)

#view regression model summary

summary(model)

Call:

lm(formula = score ~ hours, data = df)

Residuals:

Min 1Q Median 3Q Max

-5.140 -3.219 -1.193 2.816 5.772

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 65.334 2.106 31.023 1.41e-13 ***

hours 1.982 0.248 7.995 2.25e-06 ***

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Residual standard error: 3.641 on 13 degrees of freedom

Multiple R-squared: 0.831, Adjusted R-squared: 0.818

F-statistic: 63.91 on 1 and 13 DF, p-value: 2.253e-06

From the values in the Estimate column of the output, we can write the following fitted regression line:

$$\text{Exam Score} = 65.334 + 1.982(\text{Hours})$$

Here's how to interpret each coefficient in the model:

Intercept: For a student who studies 0 hours, the expected exam score is 65.334.
hours: For each additional hour studied, the expected exam score increases by 1.982.

We can use this equation to estimate the exam score a student will receive based on their hours studied.

$$\text{Exam Score} = 65.334 + 1.982(5) = 75.244$$

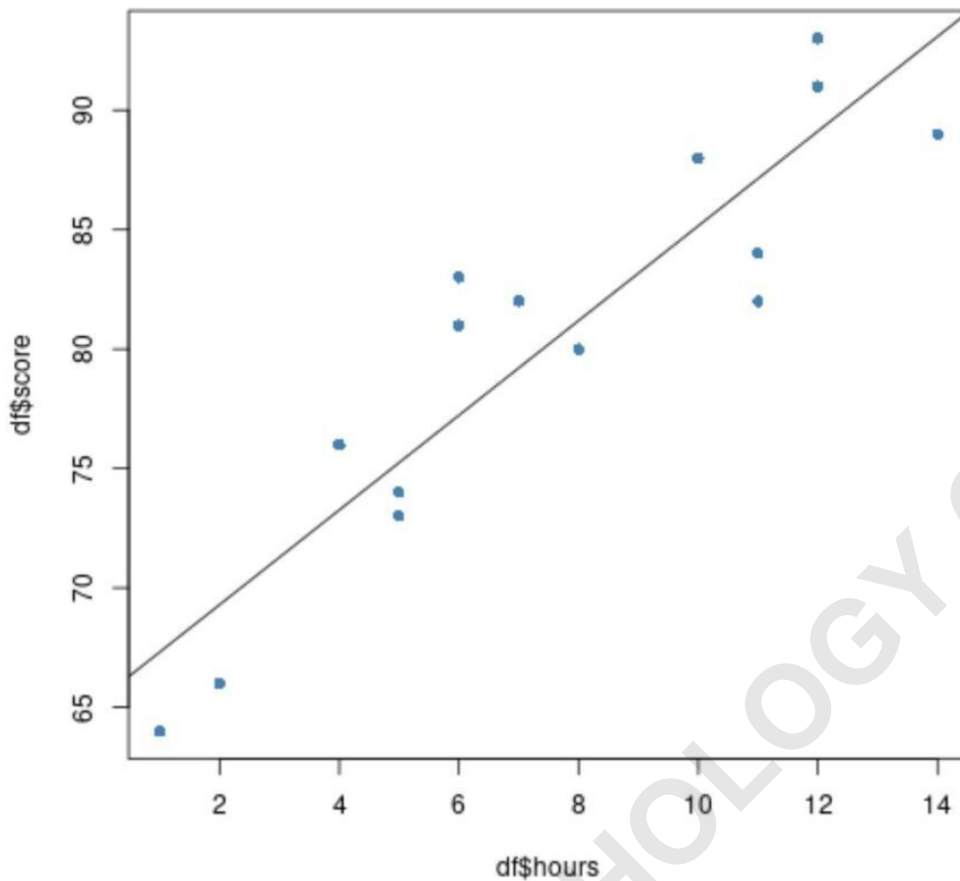
Lastly, we can create a scatter plot of the original data with the fitted regression line overlaid on the plot:

```
#create scatter plot of data
```

```
plot(df$hours, df$score, pch=16, col='steelblue')
```

```
#add fitted regression line to scatter plot
```

```
abline(model)
```



The blue circles represent the data and the black line represents the fitted regression line.