Find The Least Squares Straight Line Y = Mx B To Fit The Data Points: (1, 9),(2, 7),(3, 3),(4, 2). Compute

Find The Least Squares Straight Line Y = Mx + B To Fit The Data Points: (1, 9), (2, 7), (3, 3), (4, 2). Compute

In statistical analysis and data modeling, fitting a straight line to a set of data points is a fundamental task. The least squares method provides an optimal way to determine the best-fit line by minimizing the sum of the squared differences (residuals) between observed and predicted values. This approach is widely used in regression analysis, trend estimation, and predictive modeling. In this article, we will explore how to find the least squares straight line, expressed as $y = Mx + B$, that best fits the data points: (1, 9), (2, 7), (3, 3), and (4, 2). We will also walk through the step-by-step calculations to compute the slope M and intercept B.

Understanding the Least Squares Method

What Is the Least Squares Method?

The least squares method aims to find the line $y = Mx + B$ that minimizes the sum of the squared residuals:

$$
S = \sum_{i=1}^n (y_i - (Mx_i + B))^2
$$

where:
- (x_i, y_i) are the data points,
- M is the slope,
- B is the y-intercept,
- n is the number of data points.

Why Use Least Squares?

- It provides the "best-fit" line in the sense of least total squared error.
- It is mathematically straightforward and computationally efficient.
- It minimizes the impact of large deviations by squaring residuals.

Step-by-Step Calculation for Data Points

Given data points:
- $(x_1, y_1) = (1, 9)$
- $(x_2, y_2) = (2, 7)$
- $(x_3, y_3) = (3, 3)$
- $(x_4, y_4) = (4, 2)$

Our goal is to compute the slope M and intercept B for the line $y = Mx + B$.

Step 1: Calculate the Necessary Summations

Compute the following sums:

1. Sum of all x values: $\sum x_i$

2. Sum of all y values: $\sum y_i$

3. Sum of all x^2: $\sum x_i^2$

4. Sum of all $x y$: $\sum x_i y_i$

Calculations:
- $\sum x_i = 1 + 2 + 3 + 4 = 10$
- $\sum y_i = 9 + 7 + 3 + 2 = 21$
- $\sum x_i^2 = 1^2 + 2^2 + 3^2 + 4^2 = 1 + 4 + 9 + 16 = 30$
- $\sum x_i y_i = (1)(9) + (2)(7) + (3)(3) + (4)(2) = 9 + 14 + 9 + 8 = 40$

Step 2: Compute the Slope M

The formula for the slope in least squares regression is:
$$
M = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}
$$
Where $n = 4$.

Plugging in the values:
$$

M = \frac{4 \times 40 - 10 \times 21}{4 \times 30 - 10^2} = \frac{160 - 210}{120 - 100} = \frac{-50}{20} = -2.5
\]

Step 3: Compute the Intercept \(B \)

The intercept \(B \) is given by:
\[
B = \frac{\sum y_i - M \sum x_i}{n}
\]
Substituting the known values:
\[
B = \frac{21 - (-2.5) \times 10}{4} = \frac{21 + 25}{4} = \frac{46}{4} = 11.5
\]

Final Equation of the Least Squares Line

Based on the calculations:
\[
\boxed{
\text{Least Squares Line: } y = -2.5x + 11.5
}
\]

This line represents the best linear fit for the given data points according to the least squares criterion.

Interpreting the Results

Understanding the Slope \(M = -2.5 \)

- The negative slope indicates an inverse relationship between \(x \) and \(y \).
- For every increase of 1 unit in \(x \), the predicted \(y \) decreases by 2.5 units.

Understanding the Intercept $B = 11.5$

- When $x = 0$, the model predicts $y = 11.5$.
- Although $x = 0$ is outside the data range, it provides a baseline for the linear relationship.

Assessing the Fit of the Line

Residuals and Error

Residuals are the differences between observed and predicted y values:

$$
r_i = y_i - (M x_i + B)
$$

Calculations:
- For $(1, 9)$: $r = 9 - (-2.5 \times 1 + 11.5) = 9 - (-2.5 + 11.5) = 9 - 9 = 0$
- For $(2, 7)$: $r = 7 - (-2.5 \times 2 + 11.5) = 7 - (-5 + 11.5) = 7 - 6.5 = 0.5$
- For $(3, 3)$: $r = 3 - (-2.5 \times 3 + 11.5) = 3 - (-7.5 + 11.5) = 3 - 4 = -1$
- For $(4, 2)$: $r = 2 - (-2.5 \times 4 + 11.5) = 2 - (-10 + 11.5) = 2 - 1.5 = 0.5$

Sum of squared residuals:

$$
SSR = 0^2 + 0.5^2 + (-1)^2 + 0.5^2 = 0 + 0.25 + 1 + 0.25 = 1.5
$$

The relatively small residual sum of squares indicates a decent fit, although further statistical measures such as R^2 could provide a more comprehensive assessment.

Conclusion

In this article, we demonstrated the process of fitting a least squares straight line to a set of data points. By carefully computing the necessary summations and applying the formulas for slope and intercept, we derived the line:

$$
\boxed{
y = -2.5x + 11.5
}
$$

This line provides the best linear approximation to the data in the least squares sense. Such regression lines are invaluable tools in data analysis, enabling predictions, trend analysis, and insights into the underlying relationships between variables.

Additional Tips for Least Squares Regression

- Always verify the range of data to ensure the line's relevance.
- Consider plotting the data points along with the regression line for visual assessment.
- Use statistical software or graphing calculators for larger datasets to automate calculations.
- Explore other metrics like the coefficient of determination (\(R^2 \)) for goodness-of-fit evaluation.

By mastering the calculation of least squares lines, analysts and researchers can enhance their data modeling skills and derive meaningful interpretations from their data sets.

Frequently Asked Questions

How do I find the least squares straight line y = Mx + B for the data points (1, 9), (2, 7), (3, 3), (4, 2)?

To find the least squares line y = Mx + B, you calculate the slope (M) and intercept (B) using the formulas based on the data points' sums: $M = [n(\Sigma xy) - (\Sigma x)(\Sigma y)] / [n(\Sigma x^2) - (\Sigma x)^2]$, and $B = [\Sigma y - M(\Sigma x)] / n$, where n is the number of points.

What are the step-by-step calculations to determine the slope and intercept for these data points?

First, find Σx, Σy, Σxy, and Σx^2: $\Sigma x=10$, $\Sigma y=21$, $\Sigma xy=49$, $\Sigma x^2=30$. Then, with n=4, compute $M = [449 - 1021] / [430 - 10^2] = (196 - 210)/ (120 - 100) = -14/20 = -0.7$. Next, compute $B = (21 - (-0.7)10)/4 = (21 + 7)/4 = 28/4 = 7$.

What is the least squares regression line y = Mx + B for the given data points?

The least squares regression line is $y = -0.7x + 7$.

How well does the line y = -0.7x + 7 fit the data points? How do I measure this?

You can measure the fit using the coefficient of determination (R^2) or by calculating the residuals. The closer the residuals are to zero, the better the fit. For these points, the line captures the trend reasonably well.

Can I use this line to predict the y-value for a new x-value, say x=5?

Yes, plug x=5 into the regression line: y = -0.75 + 7 = -3.5 + 7 = 3.5. So, the predicted y-value is 3.5.

Are there any assumptions or limitations to using the least squares line for this data?

Yes, the least squares method assumes a linear relationship between x and y, and that the residuals are normally distributed with constant variance. Outliers or non-linear patterns can affect the accuracy of the fit.

How can I verify the accuracy of the least squares line I computed?

You can verify by calculating residuals (differences between actual and predicted y-values), plotting the data with the regression line, and computing R^2 to assess how well the line explains the variability in the data.

Additional Resources

Find The Least Squares Straight Line Y = Mx + B To Fit The Data Points: (1, 9), (2, 7), (3, 3), (4, 2). Compute

In the realm of data analysis and statistical modeling, the quest to understand the underlying relationship between variables often leads us to linear regression—a method that fits a straight line to a set of data points. When given a set of observations, such as (1, 9), (2, 7), (3, 3), and (4, 2), determining the best-fitting line involves calculating the slope (M) and the intercept (B) that minimize the overall discrepancy between the line and the data points. This process employs the least squares criterion, a fundamental principle in regression analysis. In this article, we will explore the step-by-step method to find the least squares straight line, understand the mathematical underpinnings, and perform the computations to derive the best-fit line for the given data set.

Understanding the Least Squares Method

Before diving into calculations, it's essential to grasp the core concept behind the least squares method.

What Is Least Squares Regression?

Least squares regression aims to find the line $y = Mx + B$ that minimizes the sum of the squared differences between the observed values y_i and the predicted values $\hat{y}_i$. These differences are called residuals:

$$
\text{Residual}_i = y_i - \hat{y}_i
$$

The total error (or residual sum of squares, RSS) is:

$$
RSS = \sum_{i=1}^n (y_i - (Mx_i + B))^2
$$

The goal is to identify the values of M and B that minimize this sum.

Why Minimize the Sum of Squares?

Squaring the residuals emphasizes larger errors, providing a penalty for significant deviations, and ensures all errors contribute positively to the total. Minimizing this sum yields the line that best represents the data in a least-squares sense.

Step 1: Organize the Data

Our data points are:

x_i	y_i
1	9
2	7
3	3
4	2

For computational purposes, it's helpful to find the following sums:

- Sum of all x_i: $\sum x_i$
- Sum of all y_i: $\sum y_i$

- Sum of all $x_i y_i$: $\sum x_i y_i$
- Sum of all x_i^2: $\sum x_i^2$

Calculating these:

$$
\sum x_i = 1 + 2 + 3 + 4 = 10
$$

$$
\sum y_i = 9 + 7 + 3 + 2 = 21
$$

$$
\sum x_i y_i = (1)(9) + (2)(7) + (3)(3) + (4)(2) = 9 + 14 + 9 + 8 = 40
$$

$$
\sum x_i^2 = 1^2 + 2^2 + 3^2 + 4^2 = 1 + 4 + 9 + 16 = 30
$$

The number of data points, $n = 4$.

Step 2: Derive the Formulas for M and B

Using the least squares method, the formulas for the slope and intercept are:

$$
M = \frac{ n \sum x_i y_i - \sum x_i \sum y_i }{ n \sum x_i^2 - (\sum x_i)^2 }
$$

$$
B = \frac{ \sum y_i - M \sum x_i }{ n }
$$

These formulas stem from setting the partial derivatives of the RSS with respect to M and B to zero, resulting in the normal equations.

Step 3: Compute the Slope M

Plugging in the computed sums:

$$
M = \frac{ 4 \times 40 - 10 \times 21 }{ 4 \times 30 - 10^2 } = \frac{ 160 - 210 }{ 120 - 100 } = \frac{ -50 }{ 20 } = -2.5
$$

The negative slope indicates an inverse relationship: as x increases, y tends to decrease.

Step 4: Compute the Intercept B

Using the value of M:

$$
B = \frac{ 21 - (-2.5) \times 10 }{ 4 } = \frac{ 21 + 25 }{ 4 } = \frac{ 46 }{ 4 } = 11.5
$$

Thus, the least squares regression line is:

$$
Y = -2.5x + 11.5
$$

Interpreting the Results

The derived line $y = -2.5x + 11.5$ offers a mathematical approximation of the relationship between x and y in the data set. The negative slope indicates that with each unit increase in x, the predicted y decreases by 2.5 units, on average. The intercept of 11.5 suggests that when $x = 0$, the predicted y would be 11.5, although this may not have practical relevance depending on the context of the data.

Visual Representation and Fit Quality

Plotting the data points alongside the regression line provides visual confirmation of the fit. The line should ideally minimize the vertical distances (residuals) from each data point.

Assessing the Fit: Residuals and Error Measures

To evaluate how well the line fits the data, residuals for each point are calculated:

x_i	y_i	Predicted $\hat{y}_i = -2.5x_i + 11.5$	Residual $y_i - \hat{y}_i$
1	9	$-2.5(1) + 11.5 = 9.0$	0
2	7	$-2.5(2) + 11.5 = 6.5$	0.5
3	3	$-2.5(3) + 11.5 = 4.0$	-1.0
4	2	$-2.5(4) + 11.5 = 1.5$	0.5

The residuals are relatively small, indicating a good fit. To quantify this, the residual sum of squares (RSS) can be computed:

$$
RSS = 0^2 + 0.5^2 + (-1)^2 + 0.5^2 = 0 + 0.25 + 1 + 0.25 = 1.5
$$

A smaller RSS suggests a closer fit of the line to the data.

Applications and Limitations

This regression approach is fundamental in fields like economics, engineering, and social sciences, where understanding relationships between variables is crucial. However, the least squares method assumes a linear relationship and can be sensitive to outliers. If data points deviate significantly from linearity or contain anomalies, alternative models or robust regression techniques may be preferable.

Conclusion

Finding the least squares straight line to fit a set of data points is a systematic process rooted in minimizing the total squared error between observed and predicted values. In the case of the data points (1, 9), (2, 7), (3, 3), and (4, 2), the computed regression line:

$$
\boxed{
Y = -2.5x + 11.5
}
$$

provides a clear mathematical relationship that captures the trend in the data. This line not only aids in understanding the underlying pattern but also serves as a predictive tool for estimating y values given new \(x \) inputs. The methodology exemplifies the power of statistical techniques in translating raw data into meaningful insights, empowering decision-makers across disciplines.

Find The Least Squares Straight Line Y Mx B To Fit The Data Points 1 9 2 7 3 3 4 2 Compute

Find The Least Squares Straight Line Y Mx B To Fit The Data Points 1 9 2 7 3 3 4 2 Compute

Related Articles

- Helium Is A "special" Element In Group 18 Because It Does Not Have 8 Valence Electrons. Why Doesn't Helium
- Read The Following Passage: "Yawning Is Both One Of Our Most Familiar And Mysterious Behaviors. All Humans
- Question: Your Network Of Friends Can Strongly Influence Your Health Habitsboth Positively And Negatively.

Back to Home