|  |  |  |
| --- | --- | --- |
| Related image | **KONERU LAKSHMAIAH EDUCATION FOUNDATION**  (Deemed to be University estd, u/s, 3 of the UGC Act, 1956) (NAAC Accredited “A++” Grade University)  Green Fields, Guntur District, A.P., India – 522502  **Department of Computer Science and Engineering**  (DST - FIST Sponsored Department) |  |

**B.Tech. II CSE(H) PROGRAM**

**A.Y. 2023-24 ODD, Semester-II**

**Course Code: 22MT2005**

**PROBABILITY, STATISTICS AND QUEUING THEORY**

**Course Outcome-2**

**Session 19:** **Least Squares Method for Regression**

1. **Course Description (Description about the subject)**

Regression analysis is a statistical method used to examine the relationship between one or more independent variables (also known as predictor or input variables) and a dependent variable (the outcome or response variable). It is widely employed in various fields, including economics, social sciences, finance, medicine, and data science, to understand the effect of different variables on the outcome of interest and make predictions based on the observed data. Here we study the Least Square Method for regression analysis.

1. **Aim**

To explain identify the linear relationship between two variables using least squares method regression.

1. **Instructional** **Objectives (Course Objectives)**

To Calculate the linear relationship between two variables using different measures of regression

1. **Learning** **Outcomes (Course Outcome)**

**CO2**: Students will be able to Apply continuous probability distributions to the real-world problems also predict the relationship between variables

1. **Module** **Description** **(CO-2 Description)**

Regression for two variables using least square method

1. **Session** **Introduction**

The main objective of many statistical investigations is to make predictions, preferably on the basis of mathematical equations. Here we will study the least square method for two vaiables.

1. **Session description**

The least square method is a statistical procedure to find the best fit for a set of data points.

The method works by minimizing the sum of the squares of the offsets or residuals of points from the plotted curve. The least squares method is used to predict the behavior of dependent variables. The least squares method provides the overall rationale for the placement of the line of best fit among the data points being studied. The least squares method is a widely used statistical technique that is associated with regression analysis.

In other words, the least square method is a way of fitting a curve or line to a set of data points so that the sum of the squared distances between the points and the curve is as small as possible. This method is often used in regression analysis, which is a statistical technique for estimating the relationship between two or more variables.

**Fitting of a quadratic function by the method of least squares**

If there is no clear indication about the functional form the regression of Y on x, we fit to our data a polynomial regression, that is, the mean of Y at x has the form:

Y=β0+ β1x+ β2x2+.........+ βpxp

where the degree is determined by inspection of the data.

In particular, we consider the case of second degree polynomial regression, that is , the mean of Y at x has the form β0+ β1x+ β2x2

Given a set of data containing n points (xi, yi), we estimate the coefficients β0, β1 and β2 by minimizing,

Taking the partial derivatives with respect to β0, β1 and β2  equating these partial derivatives to zero, rearranging some of the terms, and letting bi be the estimate of βi, we obtain the following normal equations

Solving this system of linear equations in the variables b0, b1 and b2 their values can be obtained.

**Tabular Method (two variables x and y)**

|  |  |  |  |
| --- | --- | --- | --- |
| **x** | **y** | **x2** | **xy** |
| **…** | **…** | **…** | **…** |
| **…** | **…** | **…** | **…** |
| **∑x** | **∑y** | **∑x2** | **∑xy** |

Straight line equation is y = a + bx.

The normal equations are

∑y = an + b∑x ∑xy = a∑x + b∑x2

Solve for and b to get the equation.

1. **Activities/ Case studies/related to the session.**

**NA**

1. **Examples & contemporary extracts of articles/ practices to convey the idea of the Session**

**Example 1:** The following are data on the drying time of a certain varnish and the amount of an additive that is intended to reduce the drying time.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Amount of Varnish additive (grams) x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Drying time (hours) y | 12.0 | 10.5 | 10.0 | 8.0 | 7.0 | 8.0 | 7.5 | 8.5 | 9.0 |

a) Draw a scatter plot to verify that it is reasonable to assume that the relationship is parabolic

b) Fit a second-degree polynomial by the method of least squares

c) Use the result of (b) to predict the drying time of the varnish when 6.5 grams of the additive is used

**Solution:** a) It can be seen from the following figure overall pattern suggests fitting a second degree polynomial having one relative minimum.

Normal equations are: 80.5=9b0+36b1+204b2

299=36b0+204b1+1296b2

1697=204b0+1296b1+8772b2

b0=12.2, b1=-.185 and b2=0.183, we find that the equation of the least squares polynomial is

**c)** Substituting x=6.5 into this equation, we get

That is, predicted drying time or 7.9 hours.

**Example 2:** Consider the time series data given below:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| xi | 8 | 3 | 2 | 10 | 11 | 3 | 6 | 5 | 6 | 8 |
| yi | 4 | 12 | 1 | 12 | 9 | 4 | 9 | 6 | 1 | 14 |

Use the least square method to determine the equation of line of best fit for the data. Then plot the line.

**Solution**:

Mean of xi values = (8 + 3 + 2 + 10 + 11 + 3 + 6 + 5 + 6 + 8)/10 = 62/10 = 6.2

Mean of yi values = (4 + 12 + 1 + 12 + 9 + 4 + 9 + 6 + 1 + 14)/10 = 72/10 = 7.2

Straight line equation is y = a + bx.

The normal equations are

∑y = an + b∑x ∑xy = a∑x + b∑x2

|  |  |  |  |
| --- | --- | --- | --- |
| **x** | **y** | **x2** | **xy** |
| 8 | 4 | 64 | 32 |
| 3 | 12 | 9 | 36 |
| 2 | 1 | 4 | 2 |
| 10 | 12 | 100 | 120 |
| 11 | 9 | 121 | 99 |
| 3 | 4 | 9 | 12 |
| 6 | 9 | 36 | 54 |
| 5 | 6 | 25 | 30 |
| 6 | 1 | 36 | 6 |
| 8 | 14 | 64 | 112 |
| ∑x = 62 | ∑y = 72 | ∑x2 = 468 | ∑xy = 503 |

Substituting these values in the normal equations,

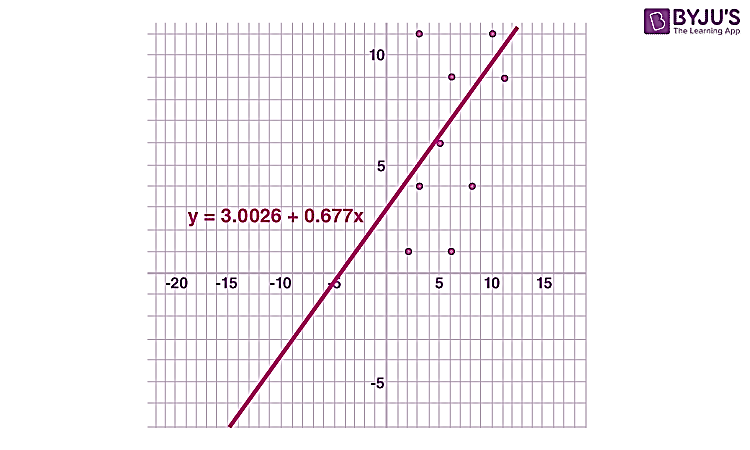
10a + 62b = 72 (1)

62a + 468b = 503 (2)

After solving for a and b, a= 3.0026 and b = 0.677.

The equation is y = 3.0026 + 0.677 x.

The graph of the equation is



1. **SAQ's-Self Assessment Questions**
2. The least square method is a statistical procedure that minimizes the sum of the squares of the \_\_\_\_\_\_\_\_\_\_\_ between the observed data points and the fitted line.

(a) residuals (b) errors (c) deviations (d) distances

The answer is (a).

1. The normal equations for the least square method are \_\_\_\_\_\_\_\_\_\_\_\_.

(a) Σy = aΣx + nb (b) Σxy = aΣx2 + bΣx (c) Σy2 = aΣx2 + bΣx (d) Σx2 = aΣx2 + bΣx

The answer is (a) and (b).

1. The least square method can be used to fit a \_\_\_\_\_\_\_\_\_\_\_ to a set of data points.

(a) line (b) curve (c) both line and curve (d) none of the above

The answer is (c).

1. The least square method is a \_\_\_\_\_\_\_\_\_\_\_ method.

(a) deterministic

(b) probabilistic

(c) both deterministic and probabilistic

(d) none of the above

The answer is (b).

1. The least square method is a powerful tool for \_\_\_\_\_\_\_\_\_\_\_\_.

(a) data analysis

(b) curve fitting

(c) both data analysis and curve fitting

(d) none of the above

The answer is (c).

1. **Summary**

The students will understand the types of Regression and its properties and also the use of Simple Linear Regression and Multiple Linear Regressions to fit a hyperplane to the data.

1. **Terminal Questions**
2. In the accompanying table, x is the tensile force applied to a steel specimen in thousands of pounds, and y is the resulting elongation in thousandths of an inch:

X: 1 2 3 4 5 6

Y: 14 33 40 63 76 85

a) Graph the data to verify that it is reasonable to assume that the regression of Y on x is linear.

b) Find the equation of the least squares line, and use it to predict the elongation when the tensile force is 3.5 thousand pounds.

1. A professor in the school of business in a university polled a dozen colleagues about the number of professional meetings professors attended in the past five years (x) and the number of papers submitted by those to refereed journals (y) during the same period. The summary data are given as follows:

n=12,

Fit a straight line to the given data.

1. The following are measurements of the air velocity and evaporation coefficient of burning fuel droplets in an impulse engine:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Air velocity (cm/s) x | 20 | 60 | 100 | 140 | 180 | 220 | 260 | 300 | 340 | 380 |
| Evaporation coefficient (mm2)y | 0.18 | 0.37 | 0.35 | 0.78 | 0.56 | 0.75 | 1.18 | 1.36 | 1.17 | 1.65 |

Fit a straight line to these data by the method of least squares, and use it to estimate the evaporation coefficient of a droplet when the air velocity is 190 cm/s.

1. Fit a straight line to these data by the method of least squares, and use it to estimate the blood pressure when the age is 50 years.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age (*x*) | 36 | 38 | 42 | 42 | 47 | 49 | 55 | 56 | 60 | 63 | 68 | 72 |
| Blood Pressure (*y*) | 118 | 115 | 125 | 140 | 128 | 145 | 150 | 147 | 155 | 149 | 152 | 160 |

1. Number of man-hours and the corresponding productivity (in units) are furnished below. Fit a simple linear regression equation Y = a + bx applying the method of least squares.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Man-hour | 3.6 | 4.8 | 7.2 | 6.9 | 10.7 | 6.1 | 7.9 | 9.5 | 5.4 |
| Productivity  (in Units) | 9.3 | 10.2 | 11.5 | 12 | 18.6 | 13.2 | 10.8 | 22.7 | 12.7 |

1. **Case Studies (CO Wise)**

**NA**

1. **Answer Key**

**NA**

1. **Glossary**

**NA**

1. **References of books, sites, links Textbooks:**

**Textbooks:**

1. Probability and Statistics Rukmangad Achari E. and E. Keshava Reddy
2. Probability and Statistics for Engineers and Scientists” Ronald E. Walpole, Sharon L. Myers and Keying Ye 8th Edition Pearson pub
3. Probability & Statistics for Engineers Dr. J. Ravichandran first Edition Wiley-India

**Reference books:**

1. Hossein Pishro-Nik, Introduction to Probability, Statistics, and Random Processes, 2014, by Kappa Research LLC; ISBN-13: 978-0990637202

**Web Resources**

1. https://ncert.nic.in/textbook.php?kemh1=0- 16
2. https://ncert.nic.in/textbook.php?jemh1=ps-15
3. **Keywords**

Regression analysis, least square methods.