**LINEAR REGRESSION**

***Regression:***

Regression is a statistical approach used to model the connection between one or more independent variables and a dependent variable. It seeks to identify a mathematical equation that best fits the data and may be used to forecast the values of the dependent variable based on the independent variables.

***Linear regression:***

Linear regression is a type of regression analysis in which the relationship between the dependent and independent variables is assumed to be straight. It's looking for a linear equation that explains the relationship between the variables.

Linear regression is a prominent tool for forecasting outcomes, comprehending variable correlations, and estimating the amount and direction of such interactions.

*Mathematical Overview*

Fitting a straight line to a set of paired observations is the most basic example of a least-squares approximation:

The mathematical expression for the straight line is:

* y - the dependent variable (the variable being predicted)
* x - the independent variable (the variable used to make predictions)
* - y-intercept (the value of y when x is 0)
* - the slope (the change in y corresponding to a one-unit change in x)
* e - is the error term (residuals or unexplained variation) between the model and the observations

*The goal of linear regression is to determine the regression coefficient values (a0 and a1) that minimize the sum of squared residuals, which represent the* ***disparities between predicted*** *and* ***actual values of the dependent variable****.*

Methods such as ordinary least squares (OLS) or maximum likelihood estimation (MLE) are frequently used in this process.

***Criteria for the “Best” Fit***

1. *Minimize the sum of the residual errors for all available data*

n = total number of points

→ *Inadequate criterion!*

1. *Minimize the sum of the absolute values of the errors*

→ *Inadequate criterion!*

*Figure Examples of* ***conditions unsuitable****for establishing theregression line:*

A screenshot of a computer

Description automatically generated*(a) minimize the sum of*

*Errors*

*(b) minimize the sum of*

*the absolute values of*

*the errors*

*(c) minimization of the*

*smallest distance*

*between a point and an*

*approximation right*

1. *The* ***best technique*** *is to minimize the sum of the SQUARES OF THE RESIDUALS between the* ***measured y*** *and the* ***linear model y****.*

A screenshot of a computer

Description automatically generated

*This function will produce a distinct line for a given collection of data.*

***Least-Squares fit of a straight line***

To determine values of and , the previous equation is differentiated with respect to each coefficient:

Setting these derivatives equal to zero will result in a minimum

Now, realizing that **∑** , we can express the equations as a set of two simultaneous linear equations with two unknowns (and ):

They can be solved simultaneously

where A screenshot of a computer

Description automatically generated and A screenshot of a computer

Description automatically generated are the means of **y** and **x** , respectively.

**Real-World Engineering Application**

An investigator reported the results of an experiment to analyze the growth rate of bacteria k (per d) as a function of oxygen concentration c (mg/L). Such information is well understood to be represented by the following equation:

and are parameters

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| c | 0.5 | 0.8 | 1.5 | 2.5 | 4 |
| k | 1.1 | 2.4 | 5.3 | 7.6 | 8.9 |

1. Use a transformation to linearize this equation.
2. Use linear regression to estimate and

***Solution:  
  
a) Inverting the equation gives a linearized equation***

Consequently, a plot of versus should yield a straight line with an intercept of and a slope of

***, ,***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **c(mg/L)** | **k(/d)** | **1/c²** | **1/k** | **1/c² \*1/k** | **(1/c²)²** |
| 0,30 | 1,20 | 11,11 | 0,83 | 9,26 | 123,46 |
| 0,70 | 2,20 | 2,04 | 0,45 | 0,93 | 4,16 |
| 1,40 | 5,10 | 0,51 | 0,20 | 0,10 | 0,26 |
| 2,30 | 7,40 | 0,19 | 0,14 | 0,03 | 0,04 |
| 4,00 | 8,60 | 0,06 | 0,12 | 0,01 | 0,00 |
|  |  |  |  |  |  |
|  | **SUM** | **13,91** | **1,74** | **10,32** | **127,92** |

***b)***

The slope and the intercept can be computed as:

n = 5

***The fit is:***

This equation can be plotted together with the data:

*A graph with a line drawn on it

Description automatically generated*

***Linearized equation***

***A graph with a line going up

Description automatically generated***

Bibliography

<http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm>

<https://nm.mathforcollege.com/mws/gen/06reg/mws_gen_reg_ppt_linear.pdf>

<https://www.academia.edu/32252075/April_2016_Least_Square_Regression_Numerical_Methods>