e-ISSN 2320 –7876 www.ijfans.org
Vol.11, Iss.9, Dec 2022

Research Paper

METHODS OF FINDING COEFFICIENT OF CORRELATION AND REGRESSION

Ishwariya R, [2] Silambarasan D
Department of Mathematics
Faculty of Arts and Science
Bharath Institute of Higher Education and Research (BIHER)
Chennai 600 073
[1] ishwariya.maths@bharathuniv.ac.in, [2] simbungm88@gmail.com

Address for Correspondence

Ishwariya R, [2] Silambarasan D
Department of Mathematics
Faculty of Arts and Science
Bharath Institute of Higher Education and Research (BIHER)
Chennai 600 073
[1] ishwariya.maths@bharathuniv.ac.in, [2] simbungm88@gmail.com

Abstract:

In this article we discuss about the topic under Mathematical Statistics such as Correlation and regression. Statistics may be used to study the vast amounts of data we have about these systems and look for patterns. Mathematical and Statistical models can be used to understand and make predictions about such diverse things as glacial movement, seismic events and tsunamis. Mathematical Statistics are used in Actuarial Science, Astrostatistics, Biostatistics, Business analytics, Chemo metrics, Demography, Econometrics, and Geostatistics, Machine learning, Population ecology, etc. The goal of a correlation analysis is to see whether two measurement variables co vary, and to quantify the strength of the relationship between the variables, whereas regression expresses the relationship in the form of an equation. Correlation and linear regression are the most commonly used techniques for investigating and relationship between two quantitative variables.

Keywords: Correlation, Regression, Mathematical statistics.

1.1 Introduction:

In this chapter we are going to discuss about Mathematical statistics and its uses in various fields. Also we discuss some definition and also the methods of collecting data with some examples. Mathematical Statistics is the application of Probability theory, a branch of mathematics to statistics, as opposed to techniques for collecting statistical data. Specific

mathematical techniques which are used for this include Mathematical analysis, linear algebra, stochastic analysis, differential equations and measure theory. Statistics is the study of the assortment, examination, understanding, organization and presentation of data. This can also collect and summarize the data which we investigated.

According to statistician Sir Arthur Lyon Bowley, Statistics is defined as "Numerical Statements of facts in any department of inquiry placed in relation to each other".

According to Merriam-Webster dictionary, statistics is defined as "Classified facts representing the conditions of a people in a state-especially the facts that can be stated in numbers or any other tabular or classified arrangement".

1.2 Scope of Statistics

Mathematical Statistics can be used in numerous fields like brain research, geography, human science, climate gauging like hood and significantly more. The objective of Mathematical statistics is to acquire recognition from that information. It direct around the application, and thus it also treated as one of the mathematical science.

1.3 Methods:

The technique include gathering summing up, examine and deciphering the variables of data in numerical. Here a portion of the techniques are given as follows

- Collection of Data
- Summarization of Data
- Analysis in Statistical

1.4. Data Collection:

The collection of realities like numerical value, alphabetic, estimation, perception is known to be as the collection of data.

Data types:

1. **Data on Qualitative measure-** it is illustrative information.

   Example: He is good, she work hard.

2. **Quantitative data** – the data given here in number form.

   Example: A hen has two legs.

1.5. Analysis of Statistics:

Analysis on statistics is the method of science on the data collecting, pattern uncovering and

the vogue. It is truly simply one more method of explaining about the term "Statistics". Subsequently the data collecting process also explain about it. This process is one of the most important processes in statistics.

2. Correlation

2.1. Definition:

The meaning of Correlation referred as the analysis is a statistical value which evaluate as well as investigates the degree (or) extend to where the two factors change regarding one another. The correlation communicates about a relationship (or) independence on a pair of set of factors upon one another. One factor might be known as the other relative (dependent) as well as the subject (independent)

Types of Correlation:

Correlation is characterized as numerous kinds.

1. Correlation on Positive as well as negative
2. Simple correlation and multiple correlation
3. Partial as well as total
4. Linear as well as non-linear

2.2. Problem: Correlation for Three data's using rank method

There are three person depute the rank with eight entries in a quiz competition

Person X	1	2	4	3	7	6	5	8
Person Y	3	2	1	5	4	7	6	8
Person Z	1	2	3	4	5	7	8	6

Which Person can give the approach to very close in a drawing competition?

Solution:

The Coefficient of correlation using rank is ρ_{xy}, ρ_{yz} and ρ_{zx}.

x	y	z	$x-y$	$(x-y)^2$	$y-z$	$(y-z)^2$	$z-x$	$(z-x)^2$
1	3	1	-2	4	2	4	0	0
2	2	2	0	0	0	0	0	0
4	1	3	3	9	-2	4	-1	1
3	5	4	-2	4	1	1	1	1
7	4	5	3	9	-1	1	-2	4

6	7	7	-1	1	0	0	1	1
5	6	8	-1	1	-2	4	3	9
8	8	6	0	0	2	4	-2	4
Total			-	28	-	18	-	20

To find ρ_{xy} (correlation between x and y)

$$\rho_{xy} = 1 - \frac{6\sum(x-y)^2}{n(n^2-1)}$$

$$= 1 - \frac{6(28)}{8(8^2-1)}$$

$$= 1 - \frac{168}{8(63)}$$

$$= 1 - \frac{168}{504}$$

$$= 1 - 0.333$$

$$\rho_{xy} = 0.667$$

To find ρ_{yz} (correlation between y and z)

$$\rho_{yz} = 1 - \frac{6\sum(y-z)^2}{n(n^2-1)}$$

$$= 1 - \frac{6(18)}{8(8^2-1)}$$

$$= 1 - \frac{108}{504}$$

$$\rho_{yz} = 0.786$$

To find ρ_{zx}(Correlation between z and x)

$$\rho_{zx} = 1 - \frac{6\sum(z-x)^2}{n(n^2-1)}$$

$$= 1 - \frac{6(20)}{8(63)}$$

$$= 1 - \frac{120}{504}$$

$$= 1 - 0.238$$

$$\rho_{zx} = 0.762$$

Since ρ_{yz} is greater than ρ_{xy} and ρ_{xz} the judges Person Y and Person Z has the approach in drawing competition is very close.

3. Regression Equation:

3.1. Definition

Regression is the investigation of the relation among the set of variables. If we consider the variable Y as dependent and the variable X as independent then the linear relation recommended among the set of variables is known as the equation of regression of Y on X. In place the parameters are resolved by utilizing the method of least square. This kind of regression equations are utilized to assess the Y value comparing to a X value, which we known.

3.2. Problem:

Find the equation of line of regression for the data given below

X	25	28	35	32	36	36	29	38	34	32
Y	43	46	49	41	36	32	31	30	33	39

Solution:

X	y	dx	dy	$(dx)^2$	$(dy)^2$	dxdy
25	43	-7	5	49	25	-35
28	46	-4	8	16	64	-32
35	49	3	11	9	121	33
32	41	0	3	0	9	0
36	36	4	-2	16	4	-8
36	32	4	-6	16	36	-24
29	31	-3	-7	9	49	21
38	30	6	-8	36	64	-48
34	33	2	-5	4	25	-10
32	39	0	1	0	1	0
325	380	5	0	155	398	-103

Let $dx = x - 32$ $dy = y - 38$

$$b_{yx} = \frac{N\sum dxdy - \sum dx \sum dy}{N\sum dx^2 - (\sum dx)^2}$$

$$= \frac{10(-103) - 5 \times 0}{10(155) - 5^2}$$

$$= \frac{-1030}{1550 - 25}$$

$$= \frac{-1030}{1525}$$

$$= -0.6754$$

$$b_{yx} = \frac{N\sum dxdy - \sum dx \sum dy}{N\sum dy^2 - (\sum dy)^2}$$

$$= \frac{10(-103) - 5 \times 0}{10(398) - 0}$$

$$= \frac{-1030}{3980}$$

$$= -0.2588$$

The equation of the regression line of y on x is

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

$$y - 38 = -0.6754(x - 32.5)$$

$$y - 38 = -0.6754x + 21.9505$$

$$y = -0.6754x + 21.9505 + 38$$

$$y = -0.6754x + 59.9505$$

The equation of the regression line of x on y is

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

$$x - 32.5 = -0.2588(y - 38)$$

$$x - 32.5 = -0.2588y + 9.8344$$

$$x = -0.2588y + 9.8344 + 32.5$$

$$x = -0.2588y + 42.3344$$

4. Reference

[1] Casella George, and Berger Roger L.,(2001). *Statistical Inference,* 2nd ed., Thomson Learning

[2] Wackerly Dennis, Mendenhall Williams, Schaffer RichardL.,(2001). *Mathematical Statistics with Application*, Duxbury Press; 6th edition.

[3] Kemp, G.C.and Santos Silva, J.(2012). *Regression towards the mode*. Journal of Econometrics, 170(1):92-101.

[4] Koenker, R. and Bassett, G. (1978). *Regression quantiles*. Econometrica: Journal of the Econometric Society, 46(1):33-50.

[5] Moore, D.S.,& McCabe,G.P.(2003). *Introduction to the practice of Statistics* (4th ed.). Newyork, NY:W.H.Freeman.

[6] Whitley E, Ball J. (2002). *Statistics review 1: Presenting and summarizing data*. Crit care. 6:66.

[7] Muddapur, M.V. (1988). *A simple test for correlation coefficient in a bivariate normal population*. Sanky: Indian Journal of Statistics. Ser.B.50, 60-68.

[8] Neter, J., Wasserman, W., & Kutner, M.H. (1985). Applied Linear Statistical Models (2nd Ed.). Homewood, IL: Richard D.Irwin.

[9] Cohen, J., Chone P., West, S.G., & Aiken, L.S. (2002). *Applied multiple regression/ correlation analysis for the behavioral sciences(3rd ed.)*

[10] Kendall, M.G. (1955). "Rank Correlation Methods", Charles Griffin & co.