Study of Test for Significance of Pearson's Correlation Coefficient

P. L. Suresh[1], Kalidindi Narayana Raju[2]

[1]Lecturer in Mathematics, Department of Mathematics & Humanities, B V Raju Degree College, Bhimavaram, West Godavari Dt. Andhra Pradesh, India
p.arunasuresh[at]gmail.com

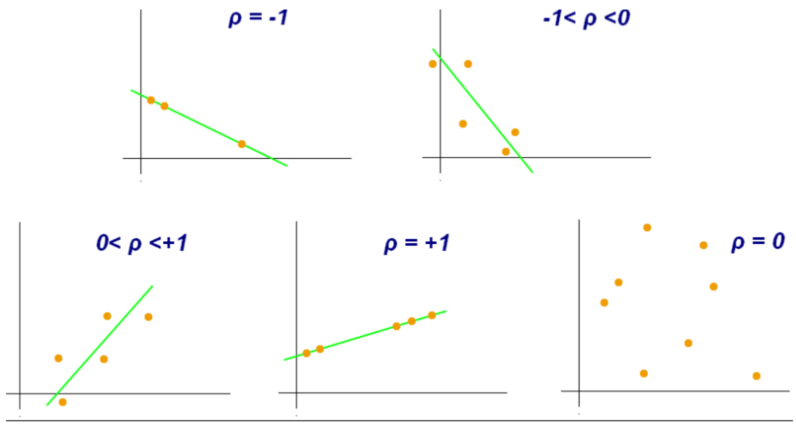
[2]Lecturer in Statistics, Department of Mathematics& Humanities, B V Raju Degree College, Bhimavaram, West Godavari Dt. Andhra Pradesh, India
knraju20[at]gmail.com

Abstract: ***This paper investigated the experiment of significance of Pearson"s correlation coefficient. It provided an in-depth judgment of dissimilar methods of testing for the significance of Pearson's correlation coefficient, as it is frequently called. The t-distribution, Fisher"s z-transformation, and the Statistical Package for Social Sciences (SPSS) were engaged. It was accomplished that each of the methods provided superior enough test for significance of correlation coefficients, which brings to rest the contrasting views that the SPSS does not provide a test for significance of correlation coefficient. The SPSS was recommended ahead of the t-distribution and z transformation due to its easy, robust, and wide applications. Researchers and academics were charged to expose their mentees to this great scientific discovery.***

Keywords: Pearson's correlation coefficient, t-distribution

1. Introduction

In statistics the Pearson's correlation coefficient referred to as Pearson's r, the Pearson product-moment correlation coefficient (PPMCC), or the bivariate correlation is a measure of linear correlation between two sets of data. It is the covariance of two variables, divided by the product of their standard deviation; thus it is fundamentally a normalized measurement of the covariance, such that the result always has a value between −1 and 1. As with covariance itself, the measure can only reflect a linear correlation of variables, and ignores many other types of relationship or correlation. As a simple example, one would expect the age and height of a sample of teenagers from a college to have a Pearson correlation coefficient significantly greater than 0, but less than 1 (as 1 would represent an unrealistically perfect correlation).



Examples of scatter diagrams with different values of correlation coefficient (ρ)

Pearson's Product Moment Correlation (r)

The Pearson"s product Moment Correlation coefficient is a measure of the strength and direction of association that exists between two variables measured on at least an interval scale. A Pearson"s correlation attempts to draw a line of best fit through the data of two variables, and the Pearson"s correlation coefficient, r, indicates how far away all these data points are from this line of best fit.

When the Pearson"s correlation is to be used, one must make necessary checks to ensure that the Pearson"s correlation is the appropriate statistic. The way to do this is to ensure the following four assumptions are passed:

1. The two variables must be measured at the interval or ratio scale.
2. There is a linear relationship between the two variables.
3. There should be no significant outliers. Outliers are single data points within your data that do not follow the usual pattern.
4. The data should be approximately normally distributed.

Computing the Pearson's

Given the bivariate set $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, the Pearson‟s Product Moment Correlation Coefficient (r) is defined as

$$r = \frac{S_{xy}}{S_x S_y}$$

Where
r =Pearson's Product Moment Correlation Coefficient
Sxy = Covariance of x and y values
and Sx and Sy = Standard deviations of x and y values respectively
Given the above relationship, the Pearson‟s Product Moment Correlation Coefficient (r) can be written as

$$r = \frac{\frac{1}{n}\sum xy - \bar{x}\,\bar{y}}{\sqrt{\frac{1}{n}\sum x^2 - \bar{x}^2}\sqrt{\frac{1}{n}\sum y^2 - \bar{y}^2}}$$

where r = Pearson's Product Moment Correlation Coefficient
N = Number of pairs of values or scores
$\sum xy$= Sum of the products of x and y
$\bar{x}$= Mean of the x values
$\bar{y}$= Mean of the y values
$\overline{xy}$= Product of the mean values of x and y
$\sum x^2$= Sum of squares of x values
$\sum y^2$= Sum of squares of y values

With the above equations of the Pearson‟s Product Moment Correlation Coefficient (r) a better computational equation can be written thus:

$$r = \frac{N\sum xy - (\sum x)(\sum y)}{\sqrt{(N\sum x^2 - (\sum x)^2)(N\sum y^2 - (\sum y)^2}}$$

Where
r = peaeson's product Moment Correlation Coefficient
N = Mumber of pairs of values
$\sum xy =$ Sum of the product of x and y
$\sum x =$ sum of the values of x
$\sum y =$ sum of the values of y
$\sum x^2 =$ Sum of the squares of x values
$\sum y^2 =$ Sum of the squares of y values
$(\sum x)^2 =$ Squares of sum of x values
$(\sum y)^2 =$ Squares of sum of y values

Using the distribution of Economics and Commerce scores of B.Com students of Chamber's College

(i) Compute the Pearson‟s Product Moment Correlation Coefficient (r),
(ii) Interpret your result, and

Statistics and Mathematics scores of B.Sc students of B V Raju Degree College

Statistics	58	84	94	76	65	87	69	82	86	72
Mathematics	59	79	78	83	89	84	85	88	78	73

(i)

Statistics (x)	Mathematics(y)	xy	x^2	y^2
58	59	3422	3364	3481
84	79	6636	7056	6241
94	78	7332	8836	6084
76	83	6308	5776	6889
65	89	5785	4225	7921
87	84	7308	7569	7056
69	58	4002	4761	3364
82	88	7216	6724	7744
86	78	6708	7396	6084
72	73	5256	5184	5329
$\sum x =$ 773	$\sum y =$769	$\sum xy =$59 973	$\sum x^2 =$34 640	$\sum y^2 =$60 193

Pearson‟s Product Moment Correlation Coefficient

$$(r)= \frac{10(59973)-(773)(769)}{\sqrt{(10(60891)-(773)^2)(10(60193)-(769)^2)}}$$

$$= \frac{5293}{7211.08792} = 0.734$$

(ii) A positive correlation coefficient of 0.734 implies that Statistics and Mathematics are positively related. In other words, a student who scores highly in Statistics is likely to score highly in Mathematics and vice versa.

Purpose of the Paper

The purpose of this position paper is to show that test of significance of Pearson‟s correlation coefficient is not only possible with the t-distribution and z-transformation, but that the Statistical Package for Social Sciences (SPSS) perfectly does this with utmost ease and accuracy.

2. Results

Test for the significance of relationships between two continuous variables

Pearson's correlation dealings the potency of a relationship between two variables. Even though in research any relationship should be assessed for the significance in addition to its strength. The strength of a relationship is indicated by the correlation coefficient $\boldsymbol{r}$**,** but in reality measured by the coefficient of determination $\boldsymbol{r^2}$**.** The significance of the relationship is expressed in probability levels **p**. The values of **p** tells how improbable a given correlation $\boldsymbol{r}$ will arise given that no relationship exists in the population. It must be noted that larger the correlation, the stronger the relationship, where a smaller **p**-level indicates more significant relationship.

In testing for the significance of correlation coefficient, some assumptions are essential**.** First, let us assume that $\boldsymbol{r}$ is the correlation between two variables x and y in a given sample and that $\boldsymbol{\rho}$ is the correlation between the same two variables x and y in the population. In correlation analysis, this means the null hypothesis that there is no significance relationship between x and y in the sample.

The Test of significance of correlation coefficient $\boldsymbol{r}$ employed one method in this paper i.e. t-distribution.

The t-distribution formula for calculating the approximate t-value to test the significance of correlation coefficient $\boldsymbol{r}$ is given by

$$\boldsymbol{t} = \boldsymbol{r}\sqrt{\frac{\boldsymbol{n}-\mathbf{2}}{\mathbf{1}-\boldsymbol{r}^{\mathbf{2}}}}$$

Where **t** = **t**-value required for the test of significance of the correlation coefficient **r**

n = sample size

r = the computed correlation coefficient being tested for significance.

Let us employ the data in below table heights (in metres) and weights (in kilogramme) of some Foot ball Players to discuss the test of significance of correlation coefficient using the t-distribution, The null hypothesis tested was that "**there is no significant relationship between heights and weights of Foot ball Players**".

Players	Heights(x)	Weights(y)
1	1.64	106
2	1.52	102
3	1.40	100
4	1.56	112
5	1.37	106
6	1.62	120
7	1.38	122
8	1.23	130

Using t-distribution

First, Pearson"s Product Moment Correlation coefficient r was computed using the formula

$$r = \frac{N\sum xy-(\sum x)(\sum y)}{\sqrt{(N\sum x^2-(\sum x)^2)(N\sum y^2-(\sum y)^2}} = 0.23$$

Second, the computed 0.23 was transformed to the t-distribution using the formula

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0.23\sqrt{\frac{8-2}{1-0.0529}} = 0.579$$

Third, the degrees of freedom (df) was found: at 0.95%, df=n-2=6

Fourth, the critical value table for t at 8: df = 1.943

t-value is less than the table value. So we accept the null hypothesis. In other words

"there is no significant relationship between heights and weights of Foot ball Players"

3. Conclusion

This paper has shown that the test of significance of correlation coefficients is very export in research because the degree of relationship alone is not sufficient to bring to a close that a computed correlation coefficient is adequate. It further exposed that one of the method of test of significance i.e. t-distribution provided good sufficient test for significance of correlation coefficients

.

References

[1] Barnette, J.J., & McClean, J.E. (1999, November). Empirically based criteria for determining meaningful effect size. Paper presented at the annual meeting of the Mid-South Educational Research Association, Point Clear, Alabama.

[2] Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from indefinitely large population. Biometrika, 10(4), 507 – 521.

[3] Goldsman, D. (2010). Hypothesis testing. Atlanta: Georgia Institute of Technology Press.

[4] Kpolovie, P. J. (2011). Statistical techniques for advanced research. New Owerri: Springfield Publishers Ltd.

[5] Levin, J. R. (1998). What if there were no more bickering about statistical significance tests? Research in the Schools, 5, 43-54.

[6] McLean, J. E., & Ernest, J.M. (1998). The role of statistical significance testing in educational research. Research in the Schools, 5, 15-22.

[7] Nix, T. W., & Barnette, J. (1998a). The data analysis dilemma: Ban or abandon. A review of null hypothesis significance testing. Research in the Schools, 5, 3-14.

[8] Student (1908). The probable error of a mean. Biometrika, 6(1), 1 – 25.

[9] Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. Educational Researcher, 25(2), 26-30.

[10] Thompson, B. (1998b). Statistical testing and effect size reporting: Portrait