1. Introduction

Welcome to the series of E-learning modules on testing for significance of correlation coefficient. In this module we are going cover the basic concept of test of significance for correlation, procedure for the test, types of tests and the respective test statistic.

By the end of this session, you will be able to:

- Explain the test of significance for Correlation
- Explain the steps involved in the tests of significance for Correlation
- Describe the types of tests of significance for correlation

Very often researchers are interested in more than just one variable that can be measured during their investigation.
For example: An auto insurance company might be interested in the number of vehicles owned by a policy holder as well as the number of drivers in the household.
An economist might need to measure the amount spent per week on groceries in a household and also the number of people in that household.

Not only are both variables important when studied separately, but we also may want to explore whether the significant relationship exists between the two variables.
In statistics, hypotheses about the value of the population correlation coefficient **rho,** between variables X and Y can be tested using either the T test or **Fisher's** transformation applied to the sample correlation coefficient R.

Testing the Significance of a Correlation
As soon as we obtain a value for the correlation coefficient, we can determine the probability that the observed correlation occurred by chance.
That is, we can conduct a significance test.

Most often we are interested in determining the probability that the correlation is a real one and not a chance occurrence. In this case, we are testing the mutually exclusive hypotheses: H naught: Rho is equal to zero, against the alternative hypothesis H one: Rho is not equal to zero or Rho is less than zero or Rho is greater than zero.

As in all hypotheses testing, you need to first determine the significance level. The generally used a most common significance level of alpha is equal to zero point zero five. This means that we are conducting a test where the odds that the correlation is a chance occurrence are no more than five out of one hundred. Before we refer a table for the critical value we have to compute the degrees of freedom or df.

The degree of freedom is simply equal to (n minus two) when the sample size is small and also when we have to test the hypothesis of the type H naught: Rho is equal to zero.
Otherwise we have to refer a Standard Normal table. Finally, one has to decide whether you are conducting a one-tailed or two-tailed test.

In the examples, where we have no strong prior theory to suggest whether the relationship

between two variables would be positive or negative, we will opt for the two tailed test.
With these three pieces of information - the significance level (alpha is equal to zero point zero five), degrees of freedom and type of test, we can now test the significance of the correlation which we have found.

When we look up for the critical value in the respective tables, suppose we find that the critical value which is less than our correlation (in a two-tailed test) we can conclude that the odds are less than five out of one hundred, that, this is a chance occurrence.

Suppose our correlation is actually quite a bit higher, we conclude that it is not a chance finding and that the correlation is "statistically significant" (given the parameters of the test).
We can then reject the null hypothesis and accept the alternative.

2. Steps in Hypothesis Testing for Correlation

Steps in hypothesis testing for correlation
We will formally go through the steps followed to test the significance of a correlation using the logical reasoning and creativity data.

Step 1: Null hypotheses
H naught : rho is equal to zero
H one : rho is not equal to zero
Notice that the hypotheses are stated in terms of population parameters. The null hypothesis specifies an exact value which implies no correlation. The alternative is a two tailed. We set the alpha level at zero point zero five.

Step 2: Assumptions
Howell describes the assumptions associated with testing the significance of correlation. These refer to normality and homogeneity of variance in the array of possible values of one variable at each value of the other variable. These assumptions are difficult to test. A close approximation is if the normality of both variables is approximately normal. Without evidence to the contrary we will assume these assumptions are accepted.

Step 3: Calculate the test statistic
The sampling distribution of R is approximately normal (but bounded at minus one and plus one) when ‘n’ is large and distributes‘t’ when ‘n’ is small.
The simplest formula for computing the appropriate ‘t’ value to test significance of a correlation coefficient employs the ‘t’ distribution:
‘t’ is equal to ‘r’ into square root of ‘n’ minus two divided by (one minus ‘r’ square).

The degrees of freedom for entering the ‘t’ distribution is (n minus two)
Otherwise we go for Fisher’s Z transformation

Step 4: Evaluate the statistic
First we determine the correlation of collected set of observations. Suppose it is significant at the zero point zero zero one level, we infer that the null hypothesis is too unlikely to be correct and we accept the alternative as a more likely explanation of the finding.

Step 5: Interpret the Result
First we can say that there was a significant positive correlation between (say) scores on the logical reasoning test and scores on the creativity test. The positive correlation implies that higher scores on creativity tend to go with higher scores on creativity and lower scores on logical reasoning go with lower scores on creativity.

Second, we might like to speculate about how logical reasoning might be related to creativity.
We cannot say from the above that logical reasoning causes creativity but we might be able to speculate that both cognitive abilities share commonalities. Both might be caused by higher

intelligence.
People with higher intelligence are better at both logical reasoning and creativity because they can entertain a larger range of possible options when it comes to creative tasks.

Notice how the first interpretation stays close to the specific variables that were measured but the second moves away from the specific variables that were measured to discuss the psychological constructs that were operationalised in the first place.
Obviously this is the most controversial part of the whole process but, the part we are mostly interested in.

3. Test One

Test one:
To test a null hypothesis H naught: Rho is equal to zero against an alternative hypothesis H one: Rho is not equal to zero for small as well as large sample size ‘n’

For a sample of small size n:
When the sample size ‘n’ is small we use‘t’- distribution and the critical region is given by C is equal to set of all X such that modulus of‘t’ greater than or equal to‘t’ alpha with (n minus two) degrees of freedom where‘t’ is equal to ‘r’ into square root of ‘n’ minus two by (one minus ‘r’ square). Where ‘r’ is the sample correlation coefficient and ‘n’ is the sample size.

One sided tests
To test a null hypothesis H naught: Rho is equal to zero against an alternative hypothesis H one: Rho is greater than zero for a small sample size ‘n’
C is equal to set of all X such that ‘t’ greater than or equal to ‘t’ two alpha with (n minus two) degrees of freedom where ‘t’ is equal to ‘r’ into square root of ‘n’ minus two by (one minus ‘r’ square). Where ‘r’ is the sample correlation and ‘n’ is the sample size.

One sided tests
To test a null hypothesis H naught: Rho is equal to zero against an alternative hypothesis H one: Rho is greater than zero for a small sample size ‘n’
C is equal to set of all X such that ‘t’ greater than or equal to ‘t’ two alpha with (n minus two) degrees of freedom where ‘t’ is equal to ‘r’ into square root of ‘n’ minus two by (one minus ‘r’ square). Where ‘r’ is the sample correlation and ‘n’ is the sample size.

For large n:
If the sample size is large, then R.A Fisher suggested the following transformation , that is Z is equal to one by two into Ln of (one plus ‘r’ by one minus ‘r’) which is equal to tan ‘h’ inverse of ‘r’ and he also proved that with even for small values of ‘n’. The distribution of Z distribution of Z is approximately Normal with mean zhi is equal to one by two into Ln of (one plus Rho by one minus Rho) which is equal to tan ‘h’ inverse of Rho and variance one by (n minus three)

To test the above hypothesis the test statistic and the procedure is as follows
Compute Z is equal to one by two into Ln of (one plus r by one minus r) and
Zhi is equal to one by two into Ln of (one plus Rho by one minus Rho).
The P value is then computed by treating Z minus zhi is equal to rho by two into (n minus one) is distributed as a normal variable with mean zero and variance one by (n minus three).

Hence obtain Z dash is equal to Z minus one by two into Ln of (one plus Rho by one minus Rho) divided by square root of one by (n minus three)
A null hypothesis is rejected if modulus of Z dash is greater than or equal to Z alpha by two

One sided tests
To test a null hypothesis H naught : Rho is equal to zero against an alternative hypothesis H one: Rho is greater than zero for a large sample size ‘n’

The test procedure is to reject the null hypothesis if the computed value of Z dash is greater than Z alpha.

To test a null hypothesis H naught: Rho is equal to zero against an alternative hypothesis H one: Rho is less than zero for a large sample size ‘n’
The test procedure is to reject the null hypothesis if the computed value of Z dash is less than minus Z alpha

4. Test Two

Test 2:
To test a null hypothesis H naught: Rho is equal to rho naught against an alternative hypothesis H one: Rho is not equal to rho naught for a sample size ‘n’
That is to test if a population correlation rho from a sample of ‘n’ observations with sample correlation ‘r’ is equal to a given rho naught, first apply the Z transformation to ‘r’ and Compute Z is equal to one by two into ‘n’ of (one plus ‘r’ by one minus ‘r’) and zhi not is equal to one by two into Ln of (one plus Rho naught by one minus Rho naught)

The P value is then computed by treating Z minus zhi naught is equal to rho naught by two into (n minus one) is distributed as a normal variable with mean zero and variance one by (n minus three).
Hence obtain Z dash is equal to Z minus one by two into Ln of (one plus Rho naught by one minus Rho naught) divided by square root of one by (n minus three)
A null hypothesis is rejected if modulus of Z dash is greater than or equal to Z alpha by two

One sided tests
To test a null hypothesis H naught: Rho is equal to rho naught against an alternative hypothesis H one: Rho is greater than rho naught for a sample size ‘n’ the test procedure is as follows:
Reject the null hypothesis if the computed value of Z dash is greater than Z alpha.

To test a null hypothesis H naught: Rho is equal to rho naught against an alternative hypothesis H one: Rho is less than rho naught for a sample size ‘n’ the test procedure is as follows:
Reject the null hypothesis if the computed value of Z dash is less than minus Z alpha.

5. Test Three

Test three:

Correlation Between Two Populations

To test a null hypothesis H naught: Rho one is equal to Rho two is equal to Rho (say) against an alternative hypothesis H one: Rho one is not equal to Rho two. That is, assume that sample correlations R one and R two are computed from two independent samples of sizes 'n' one and 'n' two observations, respectively.

To test whether the two corresponding population correlations, rho one and rho two, are equal, first apply the Z transformation to the two sample correlations:
Z one is equal to one by two into Ln of (one plus r one by one minus r one) and
Z two is equal to one by two into Ln of (one plus r two by one minus r two).

We have that Z one is equal to one by two Ln of (one plus 'r' one divided by one minus 'r' one) follows normal distribution with mean (one plus rho by one minus rho) and variance one by (n one minus three).
Similarly Z two is equal to one by two into Ln of (one plus r two by one minus r two) follows Normal distribution with mean one by two into ln of (one plus Rho by one minus Rho) and variance one by (n two minus three)

The P value is derived under the null hypothesis of equal correlation. That is, the difference Z one minus Z two is distributed as a normal random variable with mean zero and variance (one by (n one minus three) plus one by ('n' two minus three))

Therefore the test statistic is Z three is equal to Z one minus Z two divided by square root of one by (n one minus three) plus one by (n two minus three).
A test procedure is to reject a null hypothesis if modulus of Z three is greater than or equal to Z alpha by two.

One sided tests
To test a null hypothesis H naught: Rho one is equal to rho two is equal to Rho (say) against an alternative hypothesis H one: Rho one is greater than Rho two for a sample size 'n' the test procedure is as follows:
Reject the null hypothesis if the computed value of Z three is greater than
Z alpha.

To test a null hypothesis H naught: Rho one is equal to rho two equal to Rho (say) against an alternative hypothesis H one: Rho one is less than Rho two for a sample size 'n' the test procedure is as follows:
Reject the null hypothesis if the computed value of Z three is less than minus
Z alpha

Using the Fisher 'r' to 'z' transformation, the calculated value of Z can be applied to assess the significance of the difference between two correlation coefficients, 'r' one and 'r' two, found in two independent samples. If 'r' one is greater than 'r' two, the resulting value of Z will

have a positive sign
If ‘r’ one is smaller than ‘r’ two, the sign of Z will be negative.

Points to be noted:
Note that a relationship can be strong and yet not significant.
Conversely, a relationship can be weak but significant
The key factor is the size of the sample.

For small samples, it is easy to produce a strong correlation by chance and one must pay attention to significance to keep from jumping to conclusions. That is, rejecting a true null hypothesis, this means making a Type one error.
For large samples, it is easy to achieve significance, and one must pay attention to the strength of the correlation to determine if the relationship explains very much.

Hence to test the significance of an observed sample correlation coefficient from an uncorrelated Bivariate Normal population, either‘t’ tests or Z tests are used. However in random sample of size ‘n’ from a Bivariate Normal population in which a population correlation coefficient is not equal to zero, Prof. R.A Fisher proved that the distribution of a sample correlation coefficient is by no means Normal and in the neighbourhood of Rho is equal to plus or minus one, its probability curve is extremely skewed even for large ‘n’

Hence if Rho is not equal to zero, we have to go for Fisher’s Z transformation and make use of Z test to test the significance of the correlation.
Since Fisher has proved that even for small samples the distribution of Z is approximately Normal with mean Zhi is equal to one by two into Ln of (one plus Rho by one minus Rho) which is equal to tan ‘h’ inverse of Rho and variance one by (n minus three) and for large ‘n’ say (greater than thirty), the approximation is fairly good. Hence Fisher’s Z transformation can be used for small as well as large samples.

Here’s a summary of our learning in this session where we have understood the following:
- Tests of significance for correlation
- Steps involved in the test procedure
- Types of tests of significance for correlation