Frequently Asked Questions

1. Explain correlation with examples?
   Answer:
   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 an associated relationship between the two variables. Hence the property of associated relationship between the variables is known as correlation

2. What do you mean by Fisher's Z transformation?
   Answer:
   In a random sample of size n from a Bivariate Normal population in which a population correlation coefficient is not equal to 0 , Prof. R.A Fisher proved that the distribution of a sample correlation coefficient is by no means Normal and in the neighbourhood of ρ = plus or minus 1, its probability curve is extremely skewed even for large n . Hence when ρ is not equal to 0 he suggested a transformation which is known as Fisher's Z transformation
   In statistics, hypotheses about the value of the population correlation coefficient ρ between variables *X* and *Y* can be tested using the Fisher transformation applied to the sample correlation coefficient *r*.

   3. Name the two attributes of Fisher's Z statistic?
   Answer:
   The two attributes of the distribution of the z' statistic are

- It is normal and
- It has a known standard error of $\sigma_{Z'} = 1/\sqrt{(N-3)}$

4. Explain the test of significance of correlation?
   Answer:
   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_0: against the alternative hypothesis H_1
   As in all hypotheses testing, you need to first determine the significance level. The generally used a most common significance level of alpha = 0.05. This means that I am conducting a test where the odds that the correlation is a chance occurrence are no more than 5 out of 100. Before we refer a table for the critical value we have to compute the degrees of freedom or df.
   The df is simply equal to (n-2) when the sample size is small and also when we have to test the hypothesis of the type H0: Rho=0 Otherwise a Standard Normal table. Finally, one has to decide whether you are doing 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 = .05), 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 5 out of 100 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 reject the null hypothesis and accept the alternative.

5. What are the different types of tests of significance?

Answer:

- Testing whether there exists a significant correlation between the variables
- Testing whether a population correlation is equal to a given value
- Testing for equality of two population correlations

6. Briefly explain the Steps in hypothesis testing for correlation.

Answer:

The steps followed to test the significance of a correlation using the logical reasoning and creativity data

Step 1: Null hypotheses

$H_o: \rho = 0.0$

$H_1: \rho \neq 0$

Notice the hypotheses are stated in terms of population parameters. The null hypothesis specifies an exact value which implies no correlation. The alternative is two-tailed. We set the alpha level at .05.

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 -1.0 and +1.0) 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:

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

The degrees of freedom for entering the t-distribution is (n – 2)

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 0.001 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 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.

7. What is the test procedure for a sample of small size, to test a null hypothesis H0: ρ =0 against an alternative hypothesis H_1: ρ ≠ 0 ?
Answer:
When the sample size n is small we use t- distribution and the critical region is given by

$$C = \{x : |t| \geq t_\alpha (n-2)]$$ where $$t = r\sqrt{\frac{n-2}{1-r^2}}$$

where r is the sample correlation coefficient and n is the sample size
Hence we reject the null hypothesis if $|t| \geq t_\alpha (n-2)$

8. What is the test procedure for a sample of large size, to test a null hypothesis H0: ρ =0 against an alternative hypothesis H_1: ρ ≠ 0 ?
Answer:
If the sample size is large then R.A Fisher suggested the following transformation , that is

$$Z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) = \tanh^{-1} r$$

and he also proved that even for small values of n . The distribution of Z is approximately Normal with mean

$$\xi = \frac{1}{2}\ln\left(\frac{1+\varsigma}{1-\varsigma}\right) = \tanh^{-1} \varsigma$$

and variance 1 / (n-3)
To test the above hypothesis the test statistic the procedure is as follows

Compute $$Z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$$ and $$\xi = \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)$$

The P-value is then computed by treating

$$Z - \xi = \frac{\rho}{2(n-1)}$$

is distributed as a normal random variable with mean zero and variance 1/ (n—3)
a normal random variable with mean zero and variance 1/ (n-3)

Hence obtain $$Z' = \frac{Z - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)}{\sqrt{\frac{1}{n-3}}}$$

A null hypothesis is rejected if $|Z| \geq Z_{\alpha/2}$

9. Write a note on one tailed test procedure to test H0: ρ =0 for sample of small size as well as large size
Answer:
To test a null hypothesis H0: ρ =0 against an alternative hypothesis H1: ρ >0
For a sample of small size n:

A test procedure is to reject the null hypothesis if $t > t_{2\alpha}(n-2)$
Where

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

where r is the sample correlation coefficient and n is the sample size
For large n:
A null hypothesis is rejected if

$$Z' = \frac{Z - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)}{\sqrt{\frac{1}{n-3}}}$$

$$Z' > Z_\alpha$$

To test a null hypothesis H0: ρ=0 against an alternative hypothesis H1: ρ<0
For a sample of small size n:
A test procedure is to reject the null hypothesis if $t < -t_{2\alpha}(n-2)$
Where

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

where r is the sample correlation coefficient and n is the sample size

For large n:

A null hypothesis is rejected if

$$Z' < -Z_\alpha$$

Where

$$Z' = \frac{Z - \frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right)}{\sqrt{\frac{1}{n-3}}}$$

10. State the test procedure to test H0: $\rho = \rho_0$ against an alternative hypothesis H1: $\rho \neq \rho_0$ for a sample size n
Answer:
To test if a population correlation ρ from a sample of n observations with sample correlation r is equal to a given ρ_0, first apply the Z transformation to r and ρ_0:

Compute $Z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$ and $\xi_0 = \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)$

The p-value is then computed by treating $Z - \xi_0 = \frac{\rho_0}{2(n-1)}$ is distributed as a normal random variable with mean zero and variance 1/ (n-3)

Hence obtain

$$Z' = \frac{Z - \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\sqrt{\frac{1}{n-3}}}$$

A null hypothesis is rejected if $|Z'| \geq Z_{\alpha/2}$

11. What is the test procedure to test H0: $\rho = \rho_0$ against an alternative hypothesis H_1: $\rho > \rho_0$ for a sample size n?

Answer:

To test the above hypothesis the test procedure is as follows.

Reject the stated null hypothesis if $Z' > Z_\alpha$

Where

$$Z' = \frac{Z - \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\sqrt{\frac{1}{n-3}}}$$

12. What is the test procedure to test H0: $\rho = \rho_0$ against an alternative hypothesis H1: $\rho < \rho_0$ for a sample size n?

Answer:

To test the above hypothesis the test procedure is as follows. Reject the stated null hypothesis if $Z' < -Z_\alpha$ where

$$Z' = \frac{Z - \frac{1}{2}\ln\left(\frac{1+\rho_0}{1-\rho_0}\right)}{\sqrt{\frac{1}{n-3}}}$$

13. Give the test procedure to test a null hypothesis H_0: $\rho_1 = \rho_2 = \rho$(say) against an alternative hypothesis H_1: $\rho_1 \neq \rho_2$

Answer:

Assume that sample correlations r_1 and r_2 are computed from two independent samples of sizes n_1 and n_2 observations, respectively.

To test whether the two corresponding population correlations, ρ_1 and ρ_2, are equal, first apply the Z transformation to the two sample correlations:

$Z_1 = \frac{1}{2}\ln\left(\frac{1+r_1}{1-r_1}\right)$ and $Z_2 = \frac{1}{2}\ln\left(\frac{1+r_2}{1-r_2}\right)$.

We have that

$$Z_1 = \frac{1}{2}\ln\left(\frac{1+r_1}{1-r_1}\right) \sim N\left(\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{n_1-3}\right)$$

$$Z_2 = \frac{1}{2}\ln\left(\frac{1+r_2}{1-r_2}\right) \sim N\left(\frac{1}{2}\ln\left(\frac{1+\rho}{1-\rho}\right), \frac{1}{n_2-3}\right)$$

Similarly

The p-value is derived under the null hypothesis of equal correlation. That is, the difference Z_1-Z_2 is distributed as a normal random variable with mean zero and variance $\frac{1}{n_1-3} + \frac{1}{n_2-3}$

Therefore the test statistic is

$$Z_3 = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

A test procedure is to reject a null hypothesis if $|Z_3| \geq Z_{\alpha/2}$

14. State the test procedure to test a null hypothesis H_0: $\rho_1 = \rho_2 = \rho$(say) against an alternative hypothesis H_1: $\rho_1 > \rho_2$

Answer:

Assume that sample correlations r_1 and r_2 are computed from two independent samples of sizes n_1 and n_2 observations, respectively. By Applying Fisher's Z transformation we get the test procedure as follows.

If $Z_1 = \frac{1}{2}\ln\left(\frac{1+r_1}{1-r_1}\right)$ and $Z_2 = \frac{1}{2}\ln\left(\frac{1+r_2}{1-r_2}\right)$.

Reject the given null hypothesis if $Z_3 > Z_\alpha$

where

$$Z_3 = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

15. State the test procedure to test a null hypothesis H_0: $\rho_1 = \rho_2 = \rho$(say) against an alternative hypothesis H_1: $\rho_1 < \rho_2$

Answer:

Assume that sample correlations r_1 and r_2 are computed from two independent samples of sizes n_1 and n_2 observations, respectively. By Applying Fisher's Z transformation we get the test procedure as follows. Reject the given null hypothesis if $Z_3 < -Z_\alpha$ where

$$Z_3 = \frac{Z_1 - Z_2}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

$Z_1 = \frac{1}{2}\ln\left(\frac{1+r_1}{1-r_1}\right)$ and $Z_2 = \frac{1}{2}\ln\left(\frac{1+r_2}{1-r_2}\right)$.