PABNA UNIVERSITY OF SCIENCE AND TECHNOLOGY

INFORMATION AND COMMUNICATION ENGINEERING

Presentation On: Distribution of sample correlation coefficient in the null case

Course Name : Engineering Statistics

Course Code : STAT-2201

Submitted By :

Name : Sanjida Israt Suchona

Roll : 220611

session : 2021-22

Department Of Information And Communication Engineering , Pust

Submitted To

Md. Sarwar Hosain

Associate Professor

Department Of Information And Communication Engineering, Pust

Presentation On: Distribution of sample correlation coefficient in the null case

Name: Sanjida Israt Suchona

Roll:220611

Introduction:

The sample correlation coefficient, often denoted as r , is a statistical measure used to quantify the strength and direction of a linear relationship between two continuous variables. When the population correlation coefficient $\rho=0$, it implies that there is no linear relationship between the variables in the population. However, due to sampling variability, the sample correlation coefficient r will rarely be exactly zero even if $\rho=0$. Understanding the distribution of r under these conditions is crucial for hypothesis testing and confidence interval construction.

Distribution of Sample Correlation Coefficient:

Bivariate Normal Distribution:

In the context of bivariate normal distributions, the exact distribution of the sample correlation coefficient r when $\rho=0$ is related to the Student's t-distribution. Specifically, the test statistic:

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

follows a Student's t-distribution with $n-2$ degrees of freedom. This relationship allows researchers to perform hypothesis tests about the population correlation coefficient.

Fisher's Z Transformation:

To simplify the analysis and make it more tractable, Fisher's Z transformation can be applied to r. This transformation maps r to a variable *ZZ* that is approximately normally distributed with a known variance, making it easier to perform hypothesis tests and construct confidence intervals.

The Fisher's Z transformation is given by:

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

The standard error of Z is:

$$SE_Z = \frac{1}{\sqrt{n-3}}$$

This transformation is particularly useful for large samples, as Z is approximately normally distributed with mean 0 and standard deviation SEz when ρ=0.

Hypothesis Testing:

Null and Alternative Hypotheses:

- Null Hypothesis (H0): ρ=0, meaning there is no linear relationship between the variables in the population.
- Alternative Hypothesis (Ha): ρ≠0, indicating there is a linear relationship.

Testing Procedure:

1. Calculate the Sample Correlation Coefficient rr:
   Use the formula for r based on the sample data.
2. Calculate the Test Statistic:
   Using the t-distribution approach:

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

   Using Fisher's Z transformation:

$$Z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$$
$$Z_{score} = \frac{Z}{SE_Z} = \frac{Z}{\frac{1}{\sqrt{n-3}}}$$

3. Determine Degrees of Freedom or Standard Error:

- For the t-distribution, degrees of freedom df =n−2.
- For Fisher's Z, the standard error is $SE_Z = \frac{1}{\sqrt{n-3}}$

4.Find Critical Value or p-Value:

- Use a t-distribution table or calculator to find the critical value or p-value for the t-statistic.
- For Fisher's Z, use a standard normal distribution to find the p-value associated with the Z-score.

5.Make a Decision:

- If the p-value is less than the chosen significance level αα (e.g., 0.05), reject H0*H*0 and conclude there is a significant linear relationship
- Otherwise, fail to reject H0, suggesting no significant linear relationship at the chosen α.

Example

Suppose we have a sample of n=25 observations and calculate the sample correlation coefficient r=0.40 between variables x and y.

Using the t-Distribution:

1.Calculate the Test Statistic:

$$t = 0.40\sqrt{\frac{25-2}{1-0.40^2}} = 0.40\sqrt{\frac{23}{1-0.16}} = 0.40\sqrt{\frac{23}{0.84}} \approx 0.40 \times 5.48 \approx 2.19$$

2.Determine Degrees of Freedom:
df $=n-2=25-2=23$

3.Find Critical Value or p-Value:

- Using a t-distribution table or calculator with df=23, we find the critical value for a two-tailed test at $\alpha=0.05$ is approximately $t_{0.025,23} \approx 2.069$.
- Alternatively, calculate the p-value associated with $t=2.19$

4.Make a Decision:

- Since $t=2.19 > t_{0.025,23} \approx 2.069$, or if the p-value is less than 0.05, we reject H_0 and conclude there is significant linear relationship.

Using Fisher's Z Transformation

1. **Transform r to Z:**
$$Z = \frac{1}{2}\ln\left(\frac{1+0.40}{1-0.40}\right) = \frac{1}{2}\ln\left(\frac{1.40}{0.60}\right) \approx \frac{1}{2}\ln(2.333) \approx 0.40$$
2. **Calculate the Standard Error:**
$$SE_Z = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{25-3}} = \frac{1}{\sqrt{22}} \approx 0.21$$
3. **Calculate the Z-score:**
$$Z_{score} = \frac{Z}{SE_Z} = \frac{0.40}{0.21} \approx 1.90$$

4.Find the p-Value:

1. Using a standard normal distribution, the p-value for a two-tailed test with $Zscore=1.90$ is approximately 0.057.

5.Make a Decision:

1. Since the p-value is greater than 0.05, we fail to reject H0 using this method, suggesting no significant linear relationship at $\alpha=0.05$.

This example illustrates how different methods (t-distribution vs. Fisher's Z transformation) can lead to slightly different conclusions due to their underlying assumptions and approximations.

Conclusion

Understanding the distribution of the sample correlation coefficient in the null case is essential for statistical inference about linear relationships. Both the t-distribution and Fisher's Z transformation provide useful tools for hypothesis testing, each with its own advantages and assumptions. The choice between these methods depends on the specific context, sample size, and desired level of precision.