## 7.2 One-Sample Correlation ($\rho$ = a)

Introduction

Correlation analysis measures the strength and direction of association between variables. In this chapter we will test whether the population correlation coefficient, *p* (**rho**) is equal to some specific value, ***a***. There are many reasons why we might need to test whether the population correlation coefficient is different from what it traditionally is. Reasons why the strength and direction of an association between two population characteristics has change may be: 1. factors that are under the control of the researcher or 2. factors that are unknown to the researcher but may be the natural evolution of a population change process. As always, it is good practice to construct and examine a scatterplot when examining correlation between variables.

In this section we will examine whether the population correlation coefficient, $\boldsymbol{\rho}$ (*rho*), obtained from an estimated sample correlation coefficient, *r* is significantly different from a hypothesized population correlation coefficient of ***a***. So, we will test the hypotheses that

$H_0$: $\boldsymbol{\rho}$ = a and

$H_a$: $\boldsymbol{\rho} \neq$ a or $H_a$: $\boldsymbol{\rho} >$ a or $H_a$: $\boldsymbol{\rho} <$ a

Most hypothesis testing for correlation significance examine the $H_a$: $\boldsymbol{\rho} \neq 0$ (this was examined in the last section); however, here we examine the null hypothesis whether the population correlation is equal to a specific value, ***a***.

## The Case: $\rho$ = a

The underlying distribution for the correlation study is the *t* distribution. As the sample size gets large, the test statistics approaches that of the normal distribution, so the normal distribution may be used. Approaches to testing that $H_0$: $\boldsymbol{\rho}$ **=** a, will involve transforming the test statistics into a standard normal distribution that uses the z score statistics to make inference about the critical region of the test. This transformation uses the **Fisher's *Z* transformation** that normalized the distribution of the correlation sampling distribution. The degree of freedom, ***df***, for the correlation hypothesis analysis is ***n* – 2**; this is because of the two samples or idealized variables being compared.

## Two Approaches

In this chapter we present two approaches or strategies to evaluate the hypothesis that the population correlation coefficient is equal to a specific value, ***a***. Any one of these approaches will yield the same result. Therefore, after you have used the Pearson or Spearman technique to compute the sample correlation coefficient, you test the null hypothesis that $\rho$ = **a** using any of the following strategies: 1. the critical value from the normalized test statistics *z* distribution, 2. the confidence interval for the sampling distribution of the correlation coefficient, *r*. Both of these approaches use the Fisher's *Z* transformation. The natural logarithm function (ln) is used to aid in this transformation. After transformation to an approximately standard normal distribution, the *z*-score statistics is used to compute the critical region for testing any null hypothesis about the population correlation coefficient or manipulated to compute the confidence interval around *r*. Table A8 shows the transformation of the Pearson *r* to the Fisher's *Z*. The following formulas are used to compute back and forth between *r* and Fisher's *Z*:

Convert r to Fisher's Z: $Z = \frac{1}{2}\ln(\frac{1+r}{1-r})$

Convert Fisher's Z to r: r = $\frac{e^{2z'}-1}{e^{2z'}+1}$

The standard error of the Fisher's *Z* for the sample correlation coefficient, *r* computed by the following formula:

standard error: Fisher's *Z* for r = $S_{zr} = \frac{1}{\sqrt{n-3}}$

Critical Value for *z* Distribution Approach

After the Fisher's transformation of the correlation coefficient, the new distribution is a standard normal distribution. The test statistics for this normalized distribution is

test statistics: $Z = \frac{Z_r - Z_\rho}{S_{zr}}$, where $Z_r$ is Fisher's *Z* for *r* and $Z_\rho$ is Fisher's *Z* for $\rho$

To test $H_0$: $\boldsymbol{\rho} = 0.70$ using the sample data in Table 7.2.1, we first transforms the correlation coefficients using the Fisher's *Z* transformation formula (or Table A8) and then use the test statistics formula to compute the transformed *z* score. The Pearson *r* for the data in Table 7.2.1 is 0.775. The *z*-score is then compared against the critical value of a standard normal distribution at a predetermined significance level, *a* = 0.05.

Fisher's $Z_r = 1.0315$ for $r = 0.0775$ and $Z_\rho = 0.8673$ for $\rho = 0.70$ (Table A8)

$$S_{zr} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{12-3}} = 0.33$$

$$z = \frac{Z_r - Z_\rho}{S_{zr}} = \frac{1.0315 - 0.8673}{0.33} = \frac{0.1642}{3} = 0.50$$

The critical value of the standard normal distribution, $a = 0.05$, for a two-tailed distribution is $z = \pm 1.96$ (to test $H_a$: $\boldsymbol{\rho} \neq$ a) and for an upper one-tailed standard normal distribution is $z = + 1.64$ (to test $H_a$: $\boldsymbol{\rho} >$ a). The null hypothesis is not rejected because $z_{stat} = 0.50$ is within the acceptance region of the test: $-1.96 < z < 1.96$.

Table 7.2.1 *Correlation Study 1: ρ = a*

| ***X*** | ***Y*** |
|---|---|
| 82 | 42 |
| 98 | 46 |
| 87 | 39 |
| 40 | 37 |
| 116 | 65 |
| 113 | 88 |
| 111 | 86 |
| 83 | 56 |
| 85 | 62 |
| 126 | 92 |
| 106 | 54 |
| 117 | 81 |

Below is a summary of the procedure that uses the critical value of the *z* distribution to evaluate the hypotheses that $H_0$: $\rho =$ **a** and $H_a$: $\rho \neq$ **a**. Although a directional alternative hypothesis may be used to evaluate the null hypothesis that the population correlation coefficient is not a specific value, we will examine only the non-directional alternative hypothesis here.

Critical Value of *z* score Distribution Approach for Hypothesis Testing: $H_0$: $\rho =$ a.

Sample Data: Table 7.2.1

**Step 1: State the hypotheses**

$H_0$: $\rho = 0.70$ and $H_a$: $\rho \neq 0.70$ (non-directional)

**Step 2: Select the significance level**

$a = 0.05$ (0.95 confidence interval)

**Step 3: Compute the test statistics**

Correlation coefficient (Pearson *r*) = 0.775

$$z = \frac{Z_r - Z_\rho}{S_{zr}} = \frac{1.0315 - 0.8673}{0.33} = \frac{0.1642}{3} = 0.50$$

The *p* value = 0.309 i.e., $\Pr(z > 0.50) = 0.309 = 1 - 0.691$

**Step 4: Determine the critical region**

The critical value for *z* distribution: $z_{CV} = \pm$ **1.96** (two-tailed, $a = 0.05$)

**Step 5: Make a decision**

We don't reject the null hypothesis since

The *z* statistics = 0.50 is within the acceptable region: $-1.96 < z =$ **0.50** $< 1.96$ or
$p = 0.31 > a = 0.05$.
Therefore, there is no reason to reject the null hypothesis that $\rho = 0.70$.

Confidence Interval Approach

After transformation to an approximately standard normal distribution, the *z*-score statistics is used to compute the critical values for the sampling distribution of the correlation coefficient. Table A8 (See Appendix A) may be used to transformation *r* to the Fisher's *Z* or the following formula can be used to do the same:

$$\text{Fisher's Z= 0.775: } Z_r = \frac{e^{2Z} - 1}{e^{2Z} + 1} = \frac{e^{2(0.775)} - 1}{e^{2(0.775)} + 1} \simeq 1.0315$$

The standard error of the Fisher's *Z* is 0.33 for the sample correlation coefficient, *r* is computed earlier, $S_{zr} = 0.33$. The confidence interval, *CI = statistics* ± *(critical* *value)(standard error)*. The critical value for a two-tailed standard normal distribution with $a = 0.05$ is 1.96. The Fisher's Z for $r$ from the formula for converting $r$ to Fisher's $Z$ or Table A8 is about 1.03. The 95% confidence interval for the Fisher's $Z$ for $r$ ($Z_r$) is 0.38 to 1.68.

$CI_{95}$ (for $Z$) = 1.03 ± 1.96(0.33) = 0.38 to 1.68

Next, we take the confidence interval for the Fisher's $Z$ and convert back to $r$ to determine the corresponding 95% confidence interval for $r$: $CI_{95}$ for $r$ is **0.36** to **0.93**. Since this interval contains $\rho = 0.70$, we don't reject $H_0$. We either use Table A8 or the Fisher's transformation formula to convert from Fisher's $Z$ to $r$.

$Z = 0.38$, $r =$ **0.36** and $Z = 1.68$, $r =$ **0.93** (Table A8)

Below is a summary of the procedure that uses the $p$ value of the confidence interval for $r$ to evaluate the hypotheses that $H_0$: $\rho = 0.70$ and $H_a$: $\rho \neq 0.70$.

Confidence Interval Approach for Hypothesis Testing: $H_0$: $\rho = a$.

Sample Data: Table 7.2.1

**Step 1: State the hypotheses**

$H_0$: $\rho = 0.70$ and $H_a$: $\rho \neq 0.70$ (non-directional)

**Step 2: Select the significance level**

$a = 0.05$ (0.95 confidence interval)

**Step 3: Compute the test statistics**

Correlation coefficient (Pearson $r$) = 0.775

$$\text{standard error} = S_{zr} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{12-3}} = 0.33$$

For $r$ = 0.775, Fisher's $Z$ = 1.0315 (Table A8)

**Step 4: Determine the criterion for rejecting $H_0$**

$CI_{95}$ (for $Z$) = 1.03 ± 1.96(0.33) = 0.38 to 1.68

$CI_{95}$ for $r$ is **0.36** to **0.93**

**Step 5: Make a decision**

We don't reject the null hypothesis since

The confidence interval for $r$ contains $\rho = 0.70$
Therefore, there no reason to reject the null hypothesis that $\rho = 0.70$.

The following is a summary of the procedure that shows both of the two approaches to evaluate the hypotheses that $H_0$: $\rho = -0.95$ and $H_a$: $\rho \neq -0.95$. The null hypothesis is rejected because of either of the following: 1. the test statistics is within the critical region and 2. the confidence interval for $r$ does not contain $p = -0.95$. All the correlations and Fisher's $Z$ values are negative even though Table A8 values are positive.

Table 7.2.2 *Correlation Study 2: $\rho = a$*

| ***X*** | ***Y*** |
|---|---|
| 24 | 1 |
| 20 | 2 |
| 18 | 3 |
| 16 | 4 |
| 17 | 5 |
| 14 | 6 |
| 12 | 7 |
| 10 | 8 |
| 9 | 9 |
| 8 | 10 |
| 6 | 11 |
| 5 | 12 |

Critical Value or Confidence Interval Approaches for Hypothesis Testing: $H_0$: $\rho$ = a.

Sample Data: Table 7.2.1

**Step 1: State the hypotheses**

$H_0$: $\rho = -0.95$ and $H_a$: $\rho \neq -0.95$ (non-directional)

**Step 2: Select the significance level**

$a = 0.05$ (0.95 confidence interval)

**Step 3: Compute the test statistics**

Correlation coefficient (Spearman $r$) = -0.99

$$\text{standard error} = S_{zr} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{12-3}} = 0.33$$

$Z_r = -2.54$ and $Z_\rho = -1.83$ ($Z$ negative since <u>negative</u> correlation)

$$z = \frac{Z_r - Z_\rho}{S_{zr}} = \frac{-2.8262-(-1.8318)}{0.33} = \frac{-0.7181}{3} = -2.18$$

**Step 4: Determine the criterion for rejecting $H_0$**

The critical value for the $z$ distribution: $z_{CV} = \pm$ **1.96** (two-tailed, $a = 0.05$) ***or***

$CI_{95}$ for $Z_r = -2.54 \pm 1.96(0.33) = -3.2033$ to $-1.8966$

$CI_{95}$ for $r$ is -**0.997** to -**0.956**

**Step 5: Make a decision**

We **reject** the null hypothesis for either of the following reasons:

1. The confidence interval for $r$ <u>does not</u> contain $\rho = -0.95$ (note: $-0.95 > -0.956$) ***or***
2. The $z_{stat} = -2.18 < z_{CV} = -1.96$ (lower boundary of $-1.96 < z < 1.96$)
   The $p = 0.015$, $Pr(z < -2.18) = 0.015$

Therefore, there is no significant reason to conclude that the population correlation coefficient, $\rho$ is -0.95.