CHAPTER SEVEN

One-Sample Correlation ($\rho = 0$)

7.1 Introduction

In an early chapter we learned how to compute the correlation coefficient for

measuring associations between two variables. The Pearson Correlation is a measure of

the strength and direction of the relationship or degree of association between two

variables with interval or ratio scales of measurements. The Spearman Correlation is a

measure of association between variables that are ordinal scales of measurements. In this

chapter we will test the strength of the association between two variables as measured by

the *correlation coefficient*, ***r***. There are many factors that contribute to whether the

correlation or the strength of the association between two variables is significant or not.

The sample size is the single most important contributor to the test for the significance of

associated variables. The nature of the relationship, such as the degree of linearity, is also

a major factor in determining if there is a linear relationship or not. We may have a

correlation which test significant, but the visual relationship may be non-linear.

Therefore, it is always good practice to construct and examine a scatterplot when

examining linear relationships between variables.

In this section we will examine whether the population correlation coefficient, ρ

(*rho*), obtained from an estimated sample correlation coefficient, r is significantly

different from a hypothesized population correlation coefficient of zero. So, we will test the hypotheses that

H_0: $\rho = 0$ and H_a: $\rho \neq 0$

Most hypothesis testing for correlation significance, examine the above hypotheses; however, there are times when we must examine the null hypothesis whether the population correlation is equal to a specific value, a. This case will be studied in the next section of this text.

The Case: $\rho = 0$

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. However, in practice, the t distribution is used for both small and large sample size when testing correlation significance. 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.

Four Approaches

In this chapter we present four approaches or strategies to evaluate the hypothesis that the population correlation coefficient is equal to zero. 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 $\rho = 0$ using any of the following strategies: 1. the correlation coefficient's critical value

table, 2. the critical value from the test statistics t distribution, 3. The probability of the test or the p value of the statistics, or 4. the confidence interval for the statistics.

Correlation Critical Value Approach

Numerous researchers have generated the critical values for the correlation coefficient, r with $n-2$ degree of freedom. Table A5 of the appendix shows the critical values for the alpha 0.1, 0.05, 0.02 and 0.01 levels of significance. The critical value found in Table A5 is the minimum r needed to be **1- a** confident that a relationship exists. For example, the critical value for the correlation coefficient with $df = 8$ and $a = 0.05$ is 0.632. If the absolute value of the sample correlation coefficient is equal to or greater than 0.632, reject the null hypothesis, H_0 (there is no relationship) and accept the alternative hypothesis, H_a. In this case, there is a statistical significant relationship between the two variables at the 0.95 confidence interval ($\rho < 0.05$). If the absolute value of r is below 0.632, do not reject H_0. Therefore, conclude that there is not a statistical significant relationship between the two variables at the $\rho < 0.05$ significance level. Table 7.1.1 shows scores for two variables for which the Pearson r is 0.775. The $df = 12 - 2 = 10$ and if alpha is set at 0.05 (0.95 confidence interval), the critical value for the correlation coefficient is 0.576. Since $r = 0.775 > 0.576$, we reject the null hypothesis and conclude that there is population correlation is not 0 or that there is a linear relationship between the variables. A scatterplot of the scores of the two variables supports this conclusion (Figure 7.1.1). Figure 7.1.1 shows a scatterplot of the correlation study data in Table 7.1.1 with a best fit regression line. The concept of the regression line is developed in later chapter of this text. Including the line on the plot gives a graphical indication of the strength and direction of the linear relationship between variables.

Table 7.1.1 *Correlation Study 1*

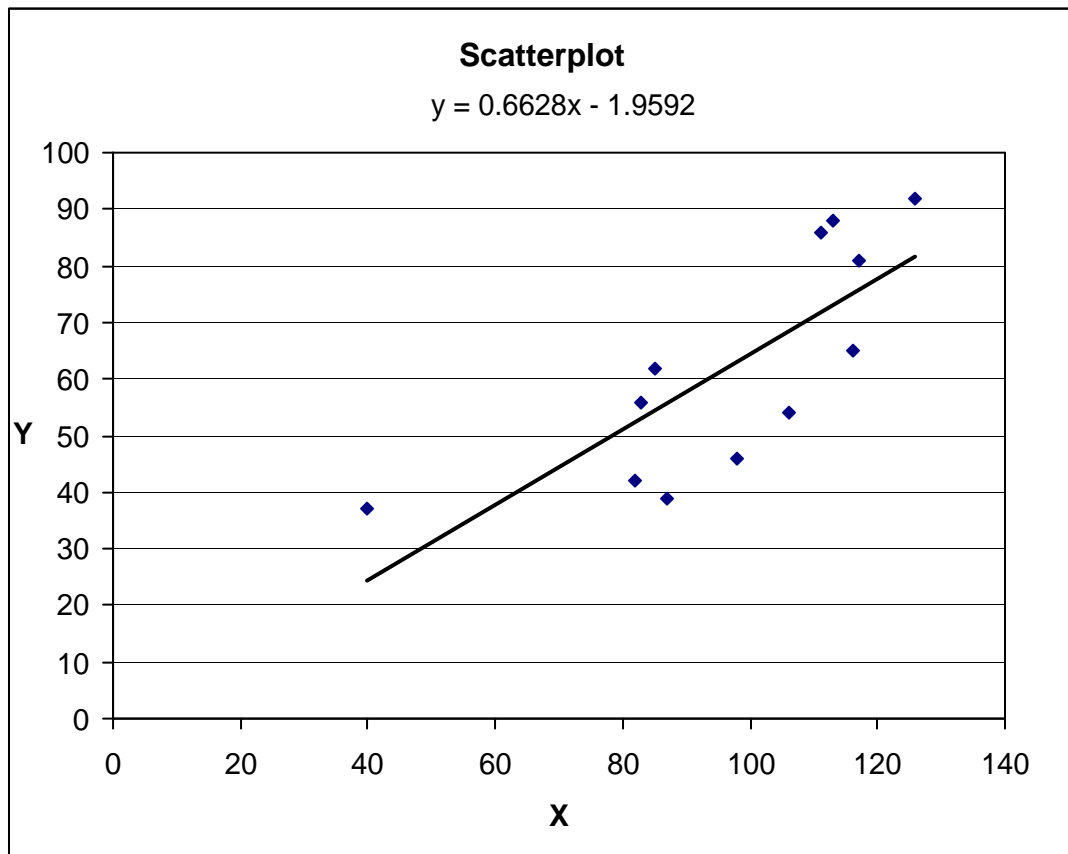| X	Y
82	42
98	46
87	39
40	37
116	65
113	88
111	86
83	56
85	62
126	92
106	54
117	81

Scatterplot

$y = 0.6628x - 1.9592$

Figure 7.1.1 Scatterplot of Correlation Study 1

Below is a summary of the procedure that uses the correlation critical value to test the hypothesis that $\rho = 0$. This is the simplest of the four approaches use to evaluate H_0: $\rho = 0$.

Correlation Critical Value Approach for Hypothesis Testing: H_0: $\rho = 0$.

Sample Data: Table 7.1.1

Step 1: State the hypotheses

H_0: $\rho = 0$ and H_a: $\rho \neq 0$ (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

Step 4: Determine the critical region

Correlation coefficient critical value is **0.576** (Table A5, $df = 10$, $a = 0.05$)

Step 5: Make a decision

We **reject** the null hypothesis since

Sample correlation coefficient, $r = 0.775 > r_{CV} = 0.576$
Therefore, there is a positive significant linear relationship between X and Y.

Critical Value for t Distribution Approach

The underlying distribution of the correlation analysis test statistics is the t distribution with $df = n - 2$ and two tailed for H_a: $\boldsymbol{\rho} \neq 0$. The test statistics is

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

Using the test statistics formula above, the computed test statistics, $t = 3.872$.

Below is a summary of the procedure that uses the critical value of the t distribution to

evaluate the hypotheses that H_0: $\rho = 0$ and H_a: $\rho \neq 0$.

Critical Value of t Distribution Approach for Hypothesis Testing: H_0: $\rho = 0$.

Sample Data: Table 7.1.1

Step 1: State the hypotheses

H_0: $\rho = 0$ and H_a: $\rho \neq 0$ (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

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0.775\sqrt{\frac{12-2}{1-0.60}} = 0.775\sqrt{\frac{10}{0.40}} = 3.875$$

Step 4: Determine the critical region

The critical value for t distribution: $t_{CV} = \mathbf{2.228}$ (Table A2, $df = 10$, $a = 0.05$)

Step 5: Make a decision

We **reject** the null hypothesis since

The t statistics = 3.875 > t_{CV} = 2.228 (absolute value of t statistics)
Therefore, there is a positive significant linear relationship between X and Y.

The Probability of the Statistics Approach

If the probability or p value of the test statistics is within the critical region of the

sampling distribution, we reject the null hypothesis that the population correlation

coefficient is zero. The *p* value is in the critical region if it is less than or equal to the

significance level. Below is a summary of the procedure that uses the *p* value of the *t*

statistics to evaluate the hypotheses that H_0: $\rho = 0$ and H_a: $\rho \neq 0$.

Statistics Probability, *p* value Approach for Hypothesis Testing: H_0: $\rho = 0$.

Sample Data: Table 7.1.1

Step 1: State the hypotheses

H_0: $\rho = 0$ and H_a: $\rho \neq 0$ (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

$$t = r\sqrt{\frac{n-2}{1-r^2}} = 0.775\sqrt{\frac{12-2}{1-0.60}} = 0.775\sqrt{\frac{10}{0.40}} = 3.875$$

From Table A2, $t_{CV} = 2.228$ and 3.169 for a = 0.05 and 0.01 respectively.

Since $t_{stat} = 3.785$ is > any of these t_{CV}, we conclude that *p* value is < 0.01

Also, from Excel we compute the *p* value = **0.003** [TDIST(3.875, 10, 2)]

Step 4: Determine the criterion for rejecting H_0

Reject the null hypothesis if *p* value = a = **0.05**

Step 5: Make a decision

We **reject** the null hypothesis since

The *p* value = 0.003 < *a* = 0.05
Therefore, there is a positive significant linear relationship between *X* and *Y*.

Confidence Interval Approach

The sampling distribution of the correlation coefficient, r can be made approximately normal by a transformation known as 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. 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 $= S_{zr} = \frac{1}{\sqrt{n-3}}$

The confidence interval, $CI = statistics \pm (critical\ value)(standard\ error)$. 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.

$Z_r = \frac{1}{2} \ln(\frac{1+r}{1-r}) = \frac{1}{2} \ln(\frac{1+0.775}{1-0.775}) = 1.03$ and $S_{zr} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{12-3}} = 0.33$

CI_{95} (for Z) $= 1.03 \pm 1.96(0.33) = 0.38$ 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 (CI_{95}) for r is **0.36** to **0.93**. Since this interval does not contain $\rho = 0$, we reject H_0 and say there is a linear relationship.

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

Below is a summary of the procedure that uses the confidence interval for the sampling distribution to evaluate the hypotheses that H_0: $\rho = 0$ and H_a: $\rho \neq 0$.

Confidence Interval Approach for Hypothesis Testing: H_0: $\rho = 0$.
Sample Data: Table 7.1.1

Step 1: State the hypotheses

H_0: $\rho = 0$ and H_a: $\rho \neq 0$ (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) $= 0.775$

$$Z_r = \frac{1}{2}\ln(\frac{1+r}{1-r}) = \frac{1}{2}\ln(\frac{1+0.775}{1-0.775}) = 1.03 \text{ and } S_{zr} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{12-3}} = 0.33$$

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

Compute CI_{95} for r is **0.36** to **0.93** (convert back to r from Fisher's Z)

Step 4: Determine the criterion for rejecting H_0

Reject the null hypothesis if CI_{95} for r <u>does not</u> contain **0**.

Step 5: Make a decision

We **reject** the null hypothesis since CI_{95} for r does not contain 0.
Therefore, there is a positive significant linear relationship between X and Y.

Table 7.1.2 shows the scores for two variables. The Spearman *rho* for these two variables is **-0.92**. Test the hypothesis that H_0: $\rho = 0$. Figure 7.1.2 shows a scatterplot with a linear regression line; a negative correlation is indicated by this plot.

Table 7.1.2 *Correlation Study 2*

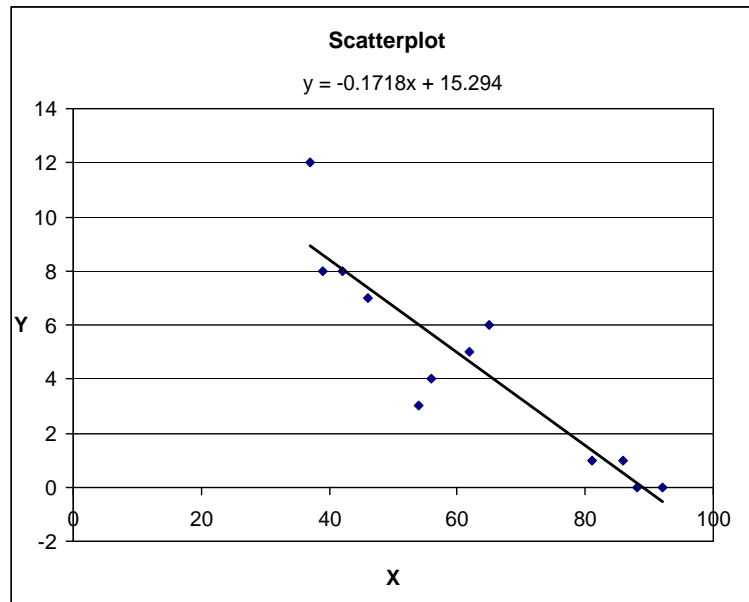| X	Y
42	8
46	7
39	8
37	12
65	6
88	0
86	1
56	4
62	5
92	0
54	3
81	1

Figure 7.1.2 Scatterplot of Correlation Study 2

The following is a summary of all four approaches used to test the hypotheses

that H_0: $\rho = 0$ and H_a: $\rho \neq 0$. We test H_0 at alpha $= 0.01$ (1%).

One-Sample Correlation Hypothesis Testing: H_0: $\rho = 0$.
Sample Data: Table 7.1.2

Step 1: State the hypotheses

    H_0: $\rho = 0$ and H_a: $\rho \neq 0$ (non-directional)

Step 2: Select the significance level

    $a = 0.01$ (0.99 confidence interval)

Step 3: Compute the test statistics

    Correlation coefficient (Spearman rho) = **-0.92**

$$t = r\sqrt{\frac{n-2}{1-r^2}} = \text{-}0.92\sqrt{\frac{12-2}{1-0.85}} = -0.92\sqrt{\frac{10}{0.15}} = -7.43$$

$$S_{zr} = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{12-3}} = 0.33 \text{ and } Z_r = -1.59$$

    CI_{95} $(Z_r) = \text{-}1.59 \pm 2.5758(0.33) = \text{-}2.45$ to $\text{-}0.73$

    Compute CI_{95} for r is **-0.99** to **-0.62** (convert back to r from Fisher's Z)

Step 4: Determine the critical region

    Correlation coefficient critical value is **0.708** (Table A5, $df = 10$, $a = 0.01$)

    The critical value for t distribution: $t_{CV} = $ **3.17** (Table A2, $df = 10$, $a = 0.01$)

    The p value $= $ **0.000** [TDIST(7.43, 10, 2)]

Step 5: Make a decision
Reject the null hypothesis because of any of the following:
1. The absolute value of r is $0.92 > r_{CV} = $ **0.708** (Table A5)
2. The absolute value of the $t_{stat} = 7.43 > t_{CV} = $ **3.17** (Table A2)
3. The p value $= 0.000 < a = $ **0.01**
4. CI_{95} for r <u>does not</u> contain **0**.

Therefore, there is a negative significant linear relationship between X and Y