**A9Wa Testing the significance of linear correlation between the two variables**

In the textbook Example 8.1 we found that the correlation coefficient was 0.83, which indicates a strong correlation between the two variables. Unfortunately, the size of the sample to provide this value of 0.83 is relatively small (sample size n = 50 months). Does this correlation coefficient of 0.83 apply only to this period of 50 months, or is it something that defines the employment of 16 and over and travel abroad in general?

As in previous textbook Chapters on hypothesis testing, we would now like to check whether this provides evidence of a significant association between the two variables for the overall population (a much longer period). To put it the way the statisticians, phrase this question: Is this value of 0.83 due to sampling error from a population where no real association exists?

To answer this question, we need to conduct the appropriate hypothesis test to check if the population value of association is zero. In this hypothesis test we are assessing the possibility that the true population value of association ρ = 0 (in other words, zero correlation).

For the sample size of n ≥ 50, we will assume that r is distributed as a t-distribution with the number of degrees of freedom df = n – 2 (we could have also used the z values from normal distribution). It can be shown that the relationship between r, ρ, and n, is given by equation (1):

(1)

This will be our calculated statistic to test the hypothesis. As per previous chapters on hypothesis testing, testing of the significance, we will use the five-step procedure we are familiar with to conduct this test progresses:

**Step 1 - State hypothesis**

**Step 2 - Select test**

**Step 3 - Set the level of significance**

**Step 4 - Extract relevant statistic**

**Step 5 - Make a decision**

**Example 1**

We’ll use again the data set from the textbook Example 1.



Figure 1 The datasets for the hypothesis testing (rows 10-39 hidden)

Figure 2 illustrates the Excel solution.

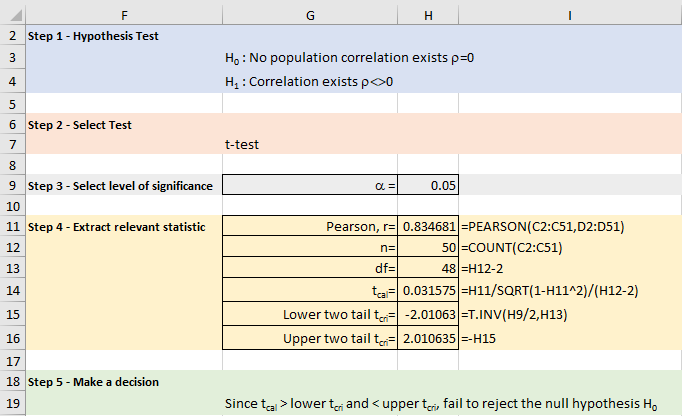


Figure W8.2 The hypothesis testing

**Excel solution**

Significance level = Cell H9 Value

Pearson coefficient = Cell H11 Formula:=PEARSON(C2:C51, D2:D51)

n = Cell H12 Formula:=COUNT(C2:C51)

df= Cell H13 Formula:=H12-2

tcal = Cell H14 Formula:=H11/SQRT((1-H11^2)/(H12-2))

Upper two tail t-critical = Cell H15 Formula:=T.INV (H9/2, H13)

Lower two tail t-critical = Cell H16 Formula:=-H15

**Step 1 - State hypothesis**

Null hypothesis H0: ρ = 0 no population correlation exists

Alternative hypothesis H1: ρ ≠ 0 correlation exists

**Step 2 - Select test**

We are testing the significance of linear correlation and we will use a t-test to test for significance. Because ρ ≠ 0, the direction is not relevant, so we will use the two-tail test.

**Step 3 - Set the level of significance** (α = 0.05)

**Step 4 - Extract relevant statistic**

Calculate the value of r. From Excel, r = 0.83.

Calculate test statistic, tcal

If H0 true, then ρ = 0, and equation (1) simplifies to equation (2).

(2)

with n – 2 degrees of freedom.

We note that the alternative hypothesis is ≠ (i.e. the correlation exists) and, therefore, we have not implied direction for the value of ρ. All we know is that it could be a significant correlation and that ρ > 0 or ρ < 0. In this case we have two directions where ρ would be deemed significant and this is called a two-tailed test. From Excel:

Using a significance level of 0.05 with 46 degrees of freedom the two-tail critical t value = T.INV(H9/2, H13) = ± 2.01.

**Step 5 - Make decision**

**Excel solution using the critical t test statistic**

The calculated value of the t-test statistic (tcal = 0.03157) is smaller than the critical t statistic value (tcri = 1.67722) and larger than -1.67722. We conclude that we should fail to reject the null hypothesis H0.



There is no evidence to suggest linear correlation between the two variables at the level of significance of 0.05. We now have evidence that this correlation is not relevant just for the 50 months window (the sample), but for the total population, i.e. the percentage of not employed age 16 and over and the number of visits abroad in the UK are associated in general.

The preceding example illustrates a two-tailed test, but one-tail tests can exist and will denote confidence in a specific relationship between X and Y.

For example, in the previous example we are quite certain that we would expect the percentage of employed age 16 and over and the number of visits abroad during the examined period of time to be related and the association to be positive (as X increases, Y increases).

In this case we would conduct H0: ρ = 0 and H1: ρ > 0. If we then tested at 5% then all this 5% would be allocated to the right-hand tail of the decision graph, and tcri would be positive. In this example, the Excel solution would give tcri =T.INV.2T(0.05,48) = + 2.31.

If we reversed the test and assumed that the association was negative (as X increases, Y decreases) then the alternative hypothesis would read H0: ρ = 0 and H1: ρ < 0, with a critical t value of tcri = - 2.31.

Test for negative correlation

H0: ρ = 0

H1: ρ < 0

Left-tailed test

Test for positive correlation

H0: ρ = 0

H1: ρ > 0

Right-tailed test