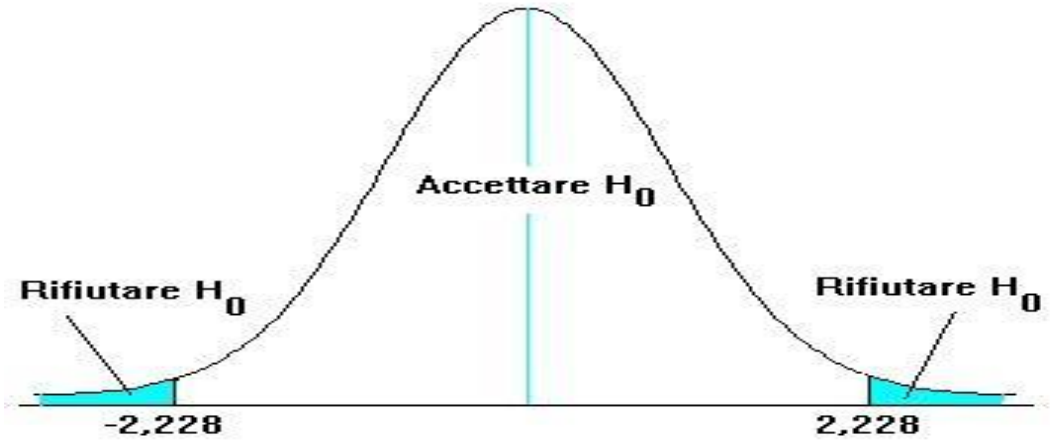
Biostatistics

Test (T)

and how to use it in data analysis



Introduction

The test (T) is one of the most important and most common statistical significance tests, it is one of the parametric measures that are used in both psychological, social and educational research, and the credit for the emergence of the test (T) to the research conducted by the scientist (Studdent), and this test was named after the most frequent letters in the name of the world, which is the letter T, and the test (T) is also used to measure the significance of hypotheses between the associated and independent averages of equal and unequal samples.

Through this post on the LinkedIn page, we will learn about the (T) test, its most important types, conditions of use, and the most important steps in calculating the (T) test, through several points, which are as follows:

- **Definition of Test (T)**
- **The most important conditions and assumptions for the use of the test (T)**
- **How to calculate a test (T)**
- **Types of test (T) and the most important conditions of their use.**
- **Definition of Test(T)**

Definition of the test T

The (T) test is one of the most important statistical tests and the most common and used in many different research and studies, which aims to reveal the significance of statistical differences between the averages of two samples, that is, the statistical test (T) is used to test the hypothesis that is related to the arithmetic mean, and the (T) test is divided into three types, which are as follows：

1- **Test (T) per sample.**
2- **Test (T) for independent samples (for two independent samples).**
3- **Test (T) for double samples (the two samples are related).**

The most important conditions and assumptions for the use of the test (T)

There are several conditions that must be met when using the researcher for the statistical (T) test, and perhaps the most important of these conditions are the following:

- **The data of the variable must be quantitative, i.e. its level of measurement is relative or categorical.**
- **The researcher should use random samples when selecting samples from the study population.**
- **There should be no association between any member of the sample and other groups, and only one group should be associated.**
- **There should be moderation in the distribution of data, data should be free of outliers and anomalous, and the data curve should be moderate.**
- **Samples must be homogeneous, i.e. they all belong to the same origin, as belonging to different origins are considered heterogeneous samples.**
- **The sample size used in the T test must be at least five individuals and preferably more than 30 individuals.**

The (T) test is calculated through five stages, which are as follows:

1. Pose the problem .

2. Formulation of hypotheses .

3. Perform calculations .

4. Comparison and decision-making .

5. Interpretation of the decision .

1 - Pose the problem

By exploiting the question posed in the exercise presented, for example, are there significant differences or a significant relationship between both the variable (x) and the variable (z) ?.

2 - Formulation of hypotheses

There are two types of hypotheses (null hypothesis and alternative hypothesis):

Null hypothesis: It is formulated by answering the negative to the problem, and it is written in a statistical way in the first place, and in a linguistic way to clarify the meaning, for example, it is written as follows: There are no statistically significant differences between both the variable (x) and the variable (z), or there is no statistically significant relationship between both the variable (x) and the variable (z).

Alternative hypothesis: It is formulated by answering the proof of the problem and is written in a statistically mainly statistical way, and in a linguistic way to illustrate the same method of writing the null hypothesis, for example, it is written as follows: There are statistically significant differences between both the variable (x) and the variable (z), or there is a statistically significant relationship between both the variable (x) and the variable (z).

3 - Calculation procedures

Calculations are performed in four steps, which are as follows:

Determine the appropriate test based on the exercise data and conditions of use Calculate the calculated value of the test by performing a set of calculations necessary for the application of the law for this test ، Calculate the degree of freedom for this test by relying on the sample size ، Determine the tabular value that is determined from the test distribution table at the level of statistical significance ($\alpha \leq 0.01$ or $\alpha \leq 0.05$).

4 - Comparison and decision-making

The comparison is made by the calculated value with the tabular value according to the degree of freedom and the level of statistical significance (α ≤ 0.01 or α ≤ 0.05), when the calculated value is greater than the tabular value (positive or negative), in this case we use the alternative hypothesis, but if the calculated value is less than the tabular value (positive or negative), this case the null hypothesis is used, and the graph can be used for the test in order to compare the calculated value and the tabular value, for By observing the location of the calculated value, if it is located in the rejection area, we reject the null hypothesis and accept the alternative hypothesis, and if it falls in the acceptance area, we accept the null hypothesis.

5 - Interpretation of the decision

Through the previous steps, the researcher is sure by 95% to 99% that there are or no statistically significant differences between both the variable (x) and the variable (z) based on the result of the comparison in the previous stage, with an error ranging from 1% to 5%.

Types of test (T) and the most important conditions for their use The T test is divided into three types:

1- **Test (T) for one sample**
2- **Test (T) for independent samples (for two independent samples)**
3- **Test (T) for double samples (the two samples are related)), and they will be explained in detail as follows:**

1-Test (T) for one sample

It is intended to calculate the differences for one sample by measuring one, and this test is used when comparing the arithmetic mean of the sample with an assumed value for the original community, which is the arithmetic average of the population, and this type of test is used on the basis of several conditions, which are as follows:

- **The dependent variable must be measured based on the quantitative level (quantitative data).**
- **The dependent variable must follow the equinox distribution.**
- **Views must be independent.**
- **The sample must be randomly selected.**

The mathematical equation is as follows:

$$t_{\bar{X}} = \frac{\bar{X} - \mu}{s_{\bar{X}}}$$

An example of testing a hypothesis about a community parameter

A one-sample t-test is used to test a hypothesis about a community parameter, such as the claim of employees of an organization that the average working hours in it differs from The general average weekly working hours specified by (40 hours).

To test this claim (the hypothesis We do the following:

- Formulating the null hypothesis and the alternative hypothesis.

- Determining the appropriate test to test the null hypothesis.

- Determining the highest percentage of error that the researcher allows (level of significance)

- Collecting information and conducting the test.

- Decision making.

Conditions for using one sample test

1- **That the dependent variable is measured at the quantitative level**
2- **That the dependent variable follow the moderate distribution**
3- **The independence of views**
4- **The sample is randomly selected**

To answer the question, the researcher collected data on a sample of 80 workers in the company as well as the number of hours each of them worked in the past week.

Zero hypothesis and alternative hypothesis.

- H_0: $\mu = 40$
- H_a: $\mu \neq 40$

$$t_{\bar{X}} = \frac{\bar{X} - \mu}{s_{\bar{X}}}$$

- **Significance level** $(0.05 = \alpha)$
- **Tests for one group and its law .**
- **Test and decision making .**

• The following table shows some information about the sample

Number of sample members (80)

The average $(\bar{x})$ **is 47.30**

Standard deviation (S) is 13.659

Standard error $(SE_{\bar{x}})$ **means the standard deviation of the mean as a statistic from the parameter and is calculated from the following equation: And is equal to 1.527**

.

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

Number of hours worked last	
N	80
Mean	47.30
Std. Deviation	13.659
Std. Error Mean	1.527

It turns out that the mean of the sample is not equal to 40 (the assumed value), but what is the probability that the value we obtained for the average (47.3) differs from The assumed value (40) only due to the random error «coincidence factor)?

To answer this question we use the T test for one sample

• The value of the T test is equal to (4.78) It means the ratio of the difference of the mean of the sample from the average of the assumed population to the expected difference in the light of chance only, and this ratio is accompanied by Probability less than 5%

Disposition: In light of the above information rejects the null hypothesis that the mean of the population is equal to 40

• There is sufficient statistical evidence that the population mean is not equal to 40 at the level of significance 5%.

One-Sample Test

Test Value = 40

t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
				Lower	Upper
4.780	79	0.000	7.300	4.26	10.34

2-Test (T) for independent samples (for two independent samples)

This type of test is used to compare the averages of two independent samples, and the two samples are independent if there is a difference between them in terms of individuals, and the characteristics that relate to the variable measured by the researcher, and they are homogeneous if the two samples are equal in number, and the variance of one of them does not differ from the variance of the other sample by more than twice, and if the two samples differ in number, the homogeneity must be tested by testing (F).

Terms of use of the T test for two independent samples:

The researcher must ensure that the conditions for using the (T) test for two independent samples are met, which are as follows:

1- Convergence of the size of the two samples:

It is preferable that the size of the two samples of the study be close, and the difference between the size of the first sample and the size of the second sample should not be spaced, for example (500) and (50) because the degrees of freedom intervene significantly and directly in determining the level of significance of the differences, so it is preferable that the size of each of the two samples exceeds (5) and preferably more than (30), but if the size of any of the two samples is less than (5), it is not possible in this case to use the test (T).

2- Homogeneity of the two samples:

If the samples belong to one origin, they are homogeneous and if the samples do not belong to one origin, they are heterogeneous, which is difficult for researchers to determine the origins of the samples in order to determine their homogeneity, so he can use the missed ratio to determine homogeneity, using the Harley's F max test and it is calculated manually, or the (Leven's Test) and it is calculated automatically when calculating (T) for two independent samples using (SPSS).

3- Moderation of distribution for the two study samples:

What is meant by moderation is the extent to which the distribution is free from twisting, and the torsion may be negative or positive, while the equinox distribution does not twist it, and the torsion coefficient extends from 3 to +3 and whenever the torsion coefficient approaches zero, the distribution is moderate, in the moderate distribution the arithmetic mean = median, and the Kolmogorov Smirnov test can be used for large samples (larger than or equal to 50), and the Shapiro-Wilk test for small samples (less than 50), through the use of (SPSS).

There are two types of T-test for two independent samples and they are shown in the following table:

1- **Test (T) of two homogeneous independent samples .**
2- **Test (T) for two independent heterogeneous samples .**

1- Terms of use of the T test for independent samples

•**<u>Assumptions</u>:**

The independent variable should be a two-level taxonomic variable (male – female or educated – uneducated) Group autonomy (if this condition is not met, such as when a person is measured twice, we need a test for associated samples)

The distribution of the dependent variable is moderate

The variations of the dependent variable of the sets are homogeneous (we can use another method to calculate the value of t)

Randomly selected samples

• Basic steps for statistical tests

To test this claim (hypothesis) we do the following:

- Formulation of the null hypothesis and the alternative hypothesis .

- Determine the appropriate test for the null hypothesis test .

- Determine the highest error rate allowed by the researcher (level of significance) Information and test procedure .

- Decision .

Example

A researcher wanted to study the difference between average student achievement and average female students' achievement on a math test.

- Randomly selected samples (male sample, female sample and each sample not less than 30)

- Zero and alternative imposition mode .

- $H_0: \mu 1 = \mu 2$ أو $H_0: \mu 1 - \mu 2 = 0$
- $H_1: \mu 1 \neq \mu 2$ أو $H_1: \mu 1 - \mu 2 \neq 0$

-The appropriate test is the T test for independent samples and its code:

$$t_{\bar{X}_1 - \bar{X}_2} = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{X}_1 - \bar{X}_2}}$$

- The level of statistical significance is equal to 5% (which means the highest error rate of the first type allowed by the researcher)

- Gather information and make a decision.

- **Number of male (56) and female (44) sample members**
- **Mean male sample (42.55) and mean female sample (44.09)**
- **Standard deviation (S) for male (10.976) and female (11.448) sample**
- **The standard error of the mean of males ($SE_{\bar{x}}$) is equal to (1.467) and females (1.726) and means the standard deviation of the mean as a statistic for The parameter is calculated from the following equation:**

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

Group Statistics					
	Respondent's Sex	N	Mean	Std. Deviation	Std. Error Mean
Age of Respondent	Male	56	42.55	10.976	1.467
	Female	44	44.09	11.448	1.726

To confirm the homogeneity of variance, we use Levene's Test.

-Zero hypothesis: • $H_0: \sigma_1^2 = \sigma_2^2$

-Research hypothesis: • $H_a: \sigma_1^2 \neq \sigma_2^2$

From the table, it Is clear that the p-value of the Levine test for the equality of variances for the two groups Is greater than (0.05).

- **Therefore, we accept the null hypothesis that the two societies are homogeneous .**

Independent Samples Test			
		Levene's Test for Equality of Variances	
		F	Sig.
Age of Respondent	Equal variances assumed	.047	.828
	Equal variances not assumed		

The following table shows test results for independent samples

The value of the T test is equal to (6.82) and means the ratio of the observed difference between the averages of the samples to the expected difference due to the coincidence (error

random).

- The p-value associated with the value of (T) is equal to 0.497 which is greater than the significance level (0.05).

- Decision: There is insufficient statistical evidence that there are differences between the average achievement of students and the average achievement of female students in mathematics .

Independent Samples Test

		t-test for Equality of Means		
		t	df	Sig. (2-tailed)
Age of Respondent	Equal variances assumed	-.682-	98	.497
	Equal variances not assumed	-.679-	90.595	.499

Mohamed abu libdah

Statistician and Data Analyst at the
NCI - Cairo University - Egypt