Data Analysis in Excel

CIS 1902103: Computer Skills for Medical Students
Dr. Raja Alomari, Tamara Almarabeh and Lama Rajab
Dept of Computer Information Systems
King Abdullah II School for Information Technology
The University Of Jordan

Fall 2014

1

Hypothesis Testing with Excel

- Researchers are interested in answering many types of questions. For example,
 - Does a new medication lower blood pressure?
 - Does smoking cause death?
 - Does practicing sports reduce heart attacks?
- These types of questions can be addressed through statistical hypothesis testing.
- Hypothesis testing is a decision-making process for evaluating claims about a population.

2

Hypothesis Testing : Null and Alternative Hypothesis

Research studies and testing usually formulate two hypotheses : One will describe the prediction while the other will describe all other possible outcomes. For example, you predict that there is no difference between A and B (null hypothesis). The only other possible outcome is that there is a difference (alternative hypothesis) between A and B .

Rejecting the null hypothesis is a central scientific task in all scientific research by proving the significance of the relationship between the phenomena.

- $\bullet\,$ The Null hypothesis , denoted by H_0 (starting point in the hypothesis testing) .
- The alternative hypothesis, denoted by H₁
 Both null and alternative hypothesis are necessary.



Hypothesis Testing Approach

- Neyman and Pearson approach: selects between the null hypothesis H₀ And an alternative hypothesis H₁
- -- The alternative hypothesis H_1 is the complementary hypothesis of H_0 .

This Approach has become the standard in scientific research.



Example:

Imagine that you have a **population** to which **no treatment** has been applied, and you already know the (mean and the standard deviation for this population). **Another population** exists same as the first but some treatment has been applied. You don't know the mean and the standard deviation for this population. Now samples are drawn from the second population and the statistics are derived from the sample.

5

Example (cont.)

Hypothesis Testing steps:

• Step 1: State the hypotheses:

H0: the treatment has **no effect**.

H1: there will be **an effect of** treatment and there is a difference between the treated and untreated populations.

• Step 2: Compute the test statistic.

The value of the test statistic (p-value) is used to make a decision regarding the null hypothesis. we will use the Ttest to compute the p-value. which can be computed in Excel by using the Ttest or Tdist functions.

Step 3: Make a decision depending on table 1 in the next slide



Interpreting significance probabilities

| Significance probability (p) | Interpretation |
|------------------------------|--|
| $p \le 0.01$ | Strong evidence to reject H ₀ |
| 0.01 | Significant evidence to reject H_0 |
| 0.05 | weak evidence against H_0 |
| p > 0.10 | Insignificant evidence to reject |
| | H_0 |

Table 1: Interpreting Significance Probabilities

7

1. T-Test Function.

(T-test Function Syntax in Excel)

• Ttest (array1, array2, tails, type)

1 : One tailed distribution
2 :Two tailed distribution

Kind of Test :

- 1 : paired(one-sample)
- 2 : Two-sample equal variance3: Two-sample unequal variance
- The second data set

The first data set

• In this course, we will use paired for the one sample test and two-sample equal variance for the two samples test

8

2. Tdist Function Syntax in Excel

• Tdist(t,deg_freedom,tails)

1 : one tailed distribution2 : two tailed distribution

the t-value computed from the following formula :

In one sample test: n-1In two sample test: 2n-2Where n is the sample size

 $t = \frac{\bar{x} - \bar{y}}{\sigma_{x,y} * \sqrt{\frac{2}{n}}}$

9

One sample & Two Samples Tests

• One sample test(paired): The mean of a population has a specific value (estimate) specified by a null hypothesis:

$$H_0: \mu = \mu_0$$
 Specific value

• Two-sample test: the mean of both samples **x** and **y** are equal:

$$H_0: \mu_x = \mu_y$$

One sided & Two Sided Test

• In one-sided test (one – tailed test), the alternative hypothesis takes only the values greater than the theoretical value $\mu_{_{0}}$, or less than the theoretical value $\mu_{_{0}}$, this is written as:

$$H_{_{\scriptscriptstyle{0}}}$$
 : μ = $\mu_{_{\scriptscriptstyle{0}}}$ $H_{_{\scriptscriptstyle{1}}}$: μ > $\mu_{_{\scriptscriptstyle{0}}}$ Right tailed

$$H_{_{\scriptscriptstyle{0}}}$$
: $\mu = \mu_{_{\scriptscriptstyle{0}}}$ $H_{_{\scriptscriptstyle{1}}}$: $\mu < \mu_{_{\scriptscriptstyle{0}}}$ Left tailed

11

One sided & Two Sided Tests

• In **two-sided test** (**two-tailed test**), the alternative hypothesis takes either values greater than or less than but not equal to the theoretical value μ_{a} , this is written as:

$$H_{\scriptscriptstyle 0}$$
: $\mu = \mu_{\scriptscriptstyle 0}$

$$H_{\scriptscriptstyle 1}: \mu \neq \mu_{\scriptscriptstyle 0}$$

How to solve (one sample test) using TTest

Step 1 State the hypotheses and identify the claim.

Step 2 To Find the *P*-value using **Ttest** in Excel (Put the in µcolumn and repeat it with the same number of samples).

Step 3 Make the decision depending on Table 1

13

How to solve (one sample test) using TDist

Step 1 State the hypotheses and identify the claim.

Step 2 To find p-value using the **Tdist** function:

$$t = \frac{x - \mu_0}{\sigma / \sqrt{n}}$$

 \underline{C} ompute the test value.

X

Where: is the sample mean, σ sample standard deviation, n is the sample size, and μ_0 is the population mean. Distribution of the sample is assumed to be normal then you'll use \mathbf{Tdist} function to find P-value with degree of freedom = n -1.

Step 3 Make the decision depending on Table 1



How to solve (two samples test) using Ttest

- **Step 1** State the hypotheses and identify the claim.
- **Step 2** Find the *P*-value using **Ttest** in Excel
- Step 3 Make the decision depending on Table 1

15

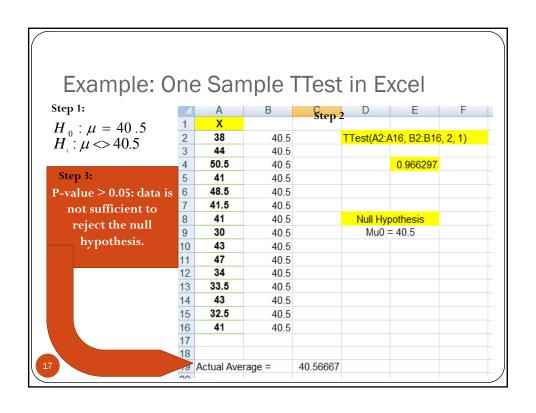
How to solve (two samples test) using TDist

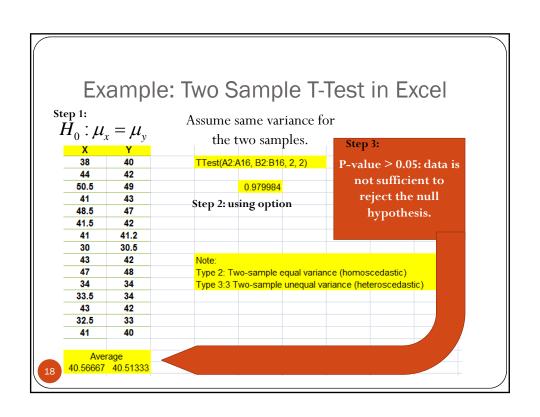
- **Step 1** State the hypotheses and identify the claim.
- Step 2 To find p-value

Compute the test value.
$$t = \frac{\bar{x} - \bar{y}}{\sigma_{x,y} * \sqrt{\frac{2}{n}}} \quad \text{where:} \quad \sigma_{xy} = \sqrt{\frac{1}{2} (\sigma_x^2 + \sigma_y^2)}$$

Then use Tdist function with degree of freedom = 2n - 2

Step 3 Make the decision depending on Table 1





Example: One Sample using Tdist in Excel

- A hospital decided to perform a study about the appropriateness of the total discharge cost for the patients. Based on other well-founded studies, the charge schedule is appropriate if the average discharge cost is \$6000. A sample of 10 discharged patients had a mean of \$6586.3 and a standard deviation of \$5262.73 was withdrawn randomly from the hospital database. Assume that the sample is normally distributed.
- What can you conclude about appropriateness of the the observed costs?

19

Solution

- Step 1: hypothesis
 - H_0 : mean = 6000 (claim)
 - H_1 : mean <> 6000

• Step 2: t-statistic,
$$t = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{6586.3 - 6000}{5262.73 / \sqrt{10}} = 0.3523$$

P-value=Tdist(0.3523, 9, 2) = 0.733

• Step 3:

According to table 1 :data is not sufficient to reject the null hypothesis.



Example: One Sample using Tdist in Excel

- A study on a children hospital found that the average number of infections per week is 16.3. An investigator wanted to test such claim by recording the results of the infections within a ten week period. He found that the sample mean of this 10 week period was 17.7 infections with a standard deviation of 1.8. Assume that the sample is normally distributed.
- Can you decided to reject of not reject the claim of the 16.3 infections per week recorded by the original study?

21

Solution

• Step 1: hypothesis $H_0: \mu = 16.3 \text{ (claim)}$ and $H_1: \mu \neq 16.3$

• **Step 2:** Compute the t-statistic:

$$t = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = \frac{17.7 - 16.3}{1.8 / \sqrt{10}} = 2.46$$

P-value = TDIST(2.46, 9, 2) = 0.036159

• **Step 3:** Significant evidence to reject H₀ (claim)

Note: here we can't use Ttest because we don't have the records in details

As in table1: we can reject the null hypothesis. Hence, the mean infections is not 16.3.

