

Dear Students.....

- I would like to give you a pre preparation to my course unit
ECON 4160 Applied Econometrics
- It is expected to make your knowledge raised on “Mean comparison” methods through this handout.
- I will post you practice questions once you are ready with this material.
- Please note that a comprehensive “Summary of this handout” is provided in Sinhala pdf.
- This material will be discussed in the lecture and in the practice classes answering all your quarries once normal academic activities are commenced.

Comparison of Groups

Common inferential statistics for testing and comparing means.

Dr. B.W.R. Damayanthi

- [One Sample t Test](#): One sample t tests (Analyze > Compare Means > One Sample T Test) are used to test if the mean of a continuous numeric variable is equal to a hypothesized value of the population mean.
- [Paired-Samples T Test](#): Paired t tests (Analyze > Compare Means > Paired-Samples T Test) are used to test if the means of two paired measurements, such as pretest/posttest scores, are significantly different.
- [Independent Samples T Test](#): Independent samples t tests (Analyze > Compare Means > Independent-Samples T Test) are used to test if the means of two independent groups are significantly different.
- [One-Way ANOVA](#): One-way ANOVA is used to test if the means of two or more groups are significantly different.

One Sample t Test

- The One Sample t Test compares a sample mean to a hypothesized population mean to determine whether the two means are significantly different.
- The One Sample t Test determines whether the sample mean is statistically different from a known or hypothesized population mean.

- The One Sample t Test is a parametric test.
- This test is also known as:
 - Single Sample t Test
- The variable used in this test is known as:
 - Test variable
- In a One Sample t Test, the test variable is compared against a "test value", which is a known or hypothesized value of the mean in the population.

Common Uses

- The One Sample t Test is commonly used to test the following:

- 1) testing a sample a against a pre-defined value

Statistical difference between a sample mean and a known or hypothesized value of the mean in the population/Statistical difference between the sample mean and the sample midpoint of the test variable.

the hypothetical mean value can be a generally assumed or pre-defined value.

For example

A researcher wants to disprove that the average retiring age of 3 wheeler driver is 68. The researcher would draw a representative sample of 3 wheeler drivers entering retirement and collecting their ages when they did so. The 1-sample t-test compares the mean score obtained in the sample (e.g., 63) to the hypothetical test value of 68. The t-test analyzes whether the difference we find in our sample is just due to random effects of chance or if our sample mean differs systematically from the hypothesized value.

2) testing a sample against an expected value

the hypothetical mean can be some derived expected value.

For instance, consider the example that the researcher observes a coin toss and notices that it is not completely random. The researcher would measure multiple coin tosses, assign one side of the coin a 0 and the flip side a 1. The researcher would then conduct a 1-sample t-test to establish whether the mean of the coin tosses is really 0.5 as expected by the laws of chance.

3) testing a sample against common sense

1-sample t-test can also be used to test for the difference against a commonly established and well known mean value.

For example, a researcher might suspect that the village she was born in is more intelligent than the rest of the country. She therefore collects IQ scores in her home village and uses the 1-sample t-test to test whether the observed IQ score differs from the defined mean value of 100 in the population.

4) testing the results of a replicated experiment against the original study

the 1-sample t-test can be used to compare the results of a replicated experiment or research analysis. In such a case the hypothesized value would be the previously reported mean score. The new sample can be checked against this mean value.

- **Note:** The One Sample t Test can only compare a single sample mean to a specified constant. It can not compare sample means between two or more groups.
- compare the means of multiple groups to each other, run an Independent Samples t Test (to compare the means of two groups) or a One-Way ANOVA (to compare the means of two or more groups).

Data Requirements

1. Test variable that is continuous (i.e., interval or ratio level)
2. Scores on the test variable are independent (i.e., independence of observations)
 - There is no relationship between scores on the test variable
 - Violation of this assumption will yield an inaccurate p value
3. Random sample of data from the population
4. Normal distribution (approximately) of the sample and population on the test variable
 - Non-normal population distributions, especially those that are thick-tailed or heavily skewed, considerably reduce the power of the test
 - Among moderate or large samples, a violation of normality may still yield accurate p values

Data Requirements Contd.

5. Homogeneity of variances (i.e., variances approximately equal in both the sample and population)

6. No outliers

- **Note:** When one or more of the assumptions for the One Sample t Test are not met, you may want to run the nonparametric Binomial Test of Equality instead.

Hypotheses

- $H_0: \mu = x$ ("the sample mean is equal to the [proposed] population mean")
- $H_1: \mu \neq x$ ("the sample mean is not equal to the [proposed] population mean")

where μ is a constant proposed for the population mean and x is the sample mean.

Test Statistic

- The test statistic for a One Sample t Test is denoted t , which is calculated using the following formula:

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}$$

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

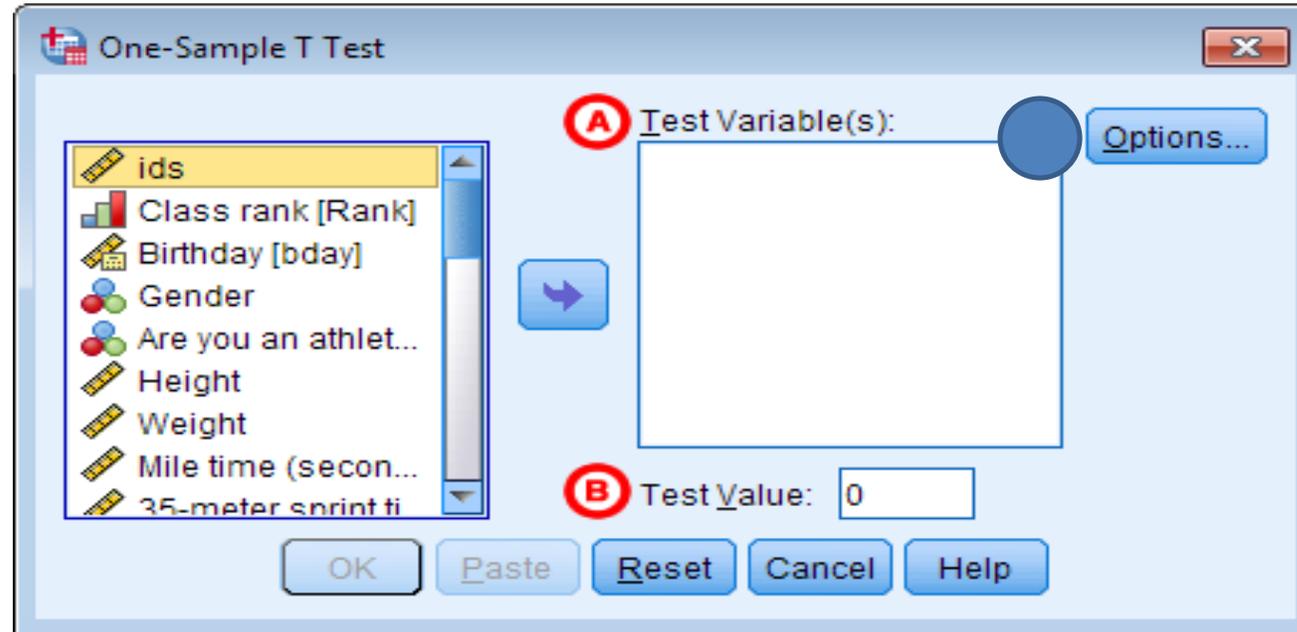
- μ = Proposed constant for the population mean
 \bar{x} = Sample mean
 n = Sample size (i.e., number of observations)
 s = Sample standard deviation
 $s_{\bar{x}}$ = Estimated standard error of the mean (s/\sqrt{n})
- The calculated t value is then compared to the critical t value from the t distribution table with degrees of freedom $df = n - 1$ and chosen confidence level. If the calculated t value $>$ critical t value, then we reject the null hypothesis.

Data Set-Up

- Your data should include one continuous, numeric variable (represented in a column) that will be used in the analysis. The variable's measurement level should be defined as Scale in the Variable View window.

Run a One Sample t Test

- **Analyze > Compare Means > One-Sample T Test.**



A Test Variable(s): The variable whose mean will be compared to the hypothesized population mean (i.e., Test Value). You may run multiple One Sample t Tests simultaneously by selecting more than one test variable. Each variable will be compared to the same Test Value.

B Test Value: The hypothesized population mean against which your test variable(s) will be compared.

Example: PROBLEM STATEMENT

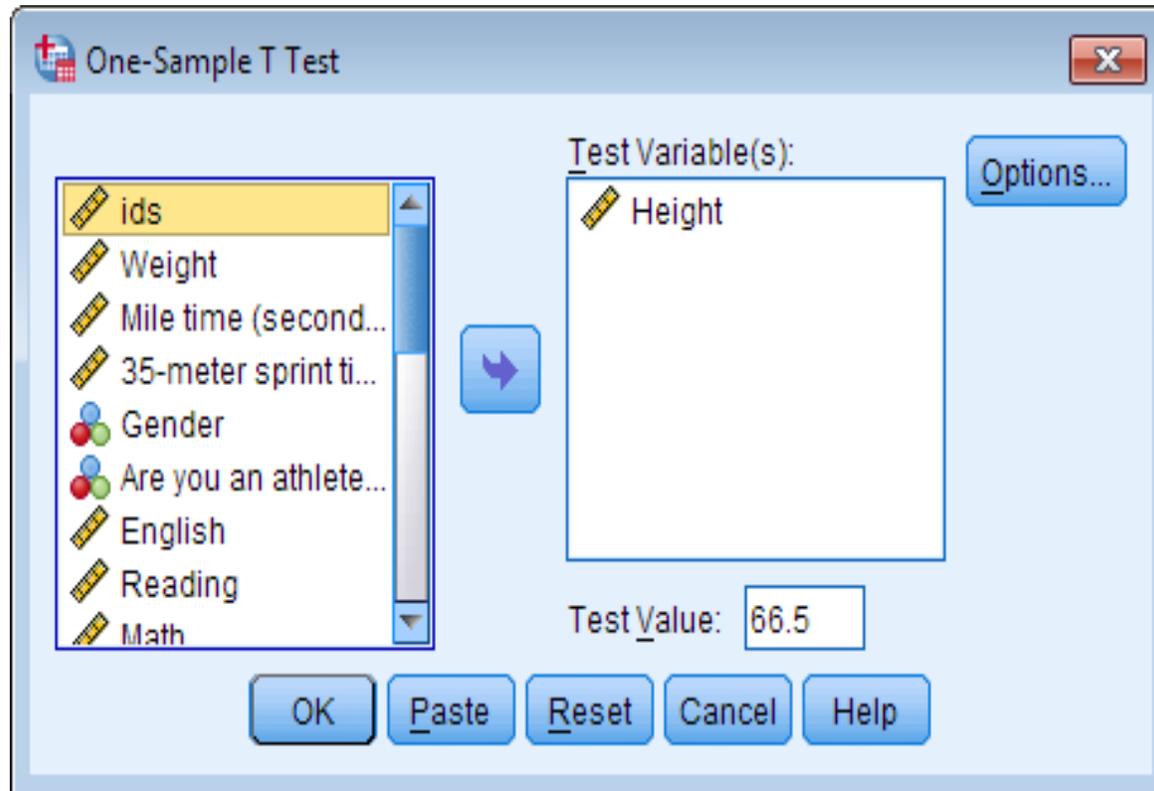
- The mean height of adults ages 20 and older is about 66.5 inches (69.3 inches for males, 63.8 inches for females). **Test** whether the mean height of the sample data is significantly different than 66.5 inches using a one-sample t test.
- The null and alternative hypotheses of this test are:
- $H_0: 66.5 = \bar{x}_{\text{Height}}$
- ("the mean height of the sample is equal to 66.5")
- $H_1: 66.5 \neq \bar{x}_{\text{Height}}$
- ("the mean height of the sample is not equal to 66.5")
- where 66.5 is the DCS's estimate of average height for adults, and \bar{x}_{Height} is the mean height of the sample.

BEFORE THE TEST

- In the sample data, we will use the variable *Height*, which is a continuous variable representing each respondent's height in inches. The heights exhibit a range of values from 55.00 to 88.41
(Analyze > Descriptive Statistics > Descriptives).
- Let's create a histogram of the data to get an idea of the distribution, and to see if our hypothesized mean is near our sample mean. Click **Graphs > Legacy Dialogs > Histogram**. Move variable Height to the Variable box, then click **OK**.

RUNNING THE TEST

- **Analyze > Compare Means > One-Sample T Test.** Move the variable *Height* to the **Test Variable(s)** area. In the **Test Value** field, enter 66.5



OUTPUT

- Two sections (boxes) appear in the output: **One-Sample Statistics** and **One-Sample Test**.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Height	408	68.0318	5.32566	.26366

- **One-Sample Statistics**, provides basic information about the selected variable, *Height*, including the valid (non missing) sample size (n), mean, standard deviation, and standard error. In this example, the mean height of the sample is 68.03 inches, which is based on 408 non missing observations.

- **One-Sample Test**, displays the results most relevant to the One Sample t Test.

One-Sample Test

	Test Value = 66.5 A					
	B t	C df	D Sig. (2-tailed)	E Mean Difference	95% Confidence Interval of the Difference F	
					Lower	Upper
Height	5.810	407	.000	1.53176	1.0135	2.0501

- A Test Value:** The number we entered as the test value in the One-Sample T Test window.
- B t Statistic:** The test statistic of the one-sample t test, denoted t . In this example, $t = 5.810$. Note that t is calculated by dividing the mean difference (E) by the standard error mean (from the One-Sample Statistics box).
- C df:** The degrees of freedom for the test. For a one-sample t test, $df = n - 1$; so here, $df = 408 - 1 = 407$.
- D Sig. (2-tailed):** The two-tailed p-value corresponding to the test statistic.
- E Mean Difference:** The difference between the "observed" sample mean (from the One Sample Statistics box) and the "expected" mean (the specified test value (A)). The sign of the mean difference corresponds to the sign of the t value (B). The positive t value in this example indicates that the mean height of the sample is greater than the hypothesized value (66.5).

DECISION AND CONCLUSIONS

- Since $p < 0.001$, we reject the null hypothesis that the sample mean is equal to the hypothesized population mean and conclude that the mean height of the sample is significantly different than the average height of the overall adult population.
- Based on the results, we can state the following:
- There is a significant difference in mean height between the sample and the overall adult population ($p < .001$).
- The average height of the sample is about 1.5 inches taller than the overall adult population average.

