Statistical Background:

Types of Statistics

Two broad types of statistics exists which are descriptive and inferential. Descriptive statistics describe the basic characteristics of the data in a study. Usually generated through an Exploratory Data Analysis (EDA), they provide simple numerical and graphical summaries about the sample and the measures. Inferential statistics allows you to make conclusions regarding the data i.e. significant differences, relationships between variables, etc.

Here are some examples of descriptive and inferential statistics:

Descriptive	Inferential
Frequencies	T-tests
Means	Chi-squares
Standard Deviations	ANOVA
Ranges	Friedman
Medians	
Modes	

T-test and ANOVA

The t-test and analysis of variance (ANOVA) compare group means. A t-test may examine gender differences in average salary or racial (white versus black) differences in average annual income etc.

While the t-test is limited to comparing means of two groups, one-way ANOVA can compare more than two groups. Therefore, the t-test is considered a special case of one-way ANOVA. It is Important to note that, these analyses do not necessarily imply any causality (i.e., a causal relationship between the left-hand and right-hand side variables).

There are three types of t-tests:

1. One-sample t-test
2. Paired sample t-test
3. Independent sample t-test

One-sample t-test

A one sample t-test procedure tests whether the mean of a single variable differs from a specified constant.

T-test in STATA

The .ttest command conducts t-tests. For a One-sample t-test, the command requires that a hypothesized value be explicitly specified. The level() option indicates the confidence level as a percentage. The 99 percent confidence level is equivalent to the .01 significance level.

Example:

In this example we will use the data < `http://www.ats.ucla.edu/stat/data/hs0.dta`> we will determine if the test score for writing is significantly different 50.

So the hypothesis we want to test is:

H_0: $\mu_{Write} = 0$
H_A: $\mu_{Write} \neq 0$

To run the t-test type **`ttest write=50, level(99)`**

```
. summarize write

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       write |       200      52.775    9.478586        31         67

. ttest write=50, level(99)

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [99% Conf. Interval]
---------+--------------------------------------------------------------------
   write |     200      52.775    .6702372    9.478586    51.03187    54.51813
------------------------------------------------------------------------------
    mean = mean(write)                                            t =   4.1403
Ho: mean = 50                                     degrees of freedom =      199

    Ha: mean < 50               Ha: mean != 50                 Ha: mean > 50
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0001          Pr(T > t) = 0.0000
```

STATA first lists descriptive statistics of the variable `write`. The mean and standard deviation of the 200 observations are 52.775 and 9.478, respectively.

There are three t-tests at the bottom of the output. The first and third are one-tailed tests, whereas the second is a two-tailed test. The t-statistic 4.14 and its small p-value reject the null hypothesis that the population mean of the writing score from 200 students is 50 at the .01 level.

Note that the hypothesized value 50 does not fall into the 99 percent confidence interval 51.031-54.518 indicating that mean of the writing score is greater than 50.

Paired samples t-test

The paired samples t-test is also referred to as the dependent or related samples t-test. It is useful for testing if a significant difference occurs between the means of two variables that represent the same group at different times (before or after) or related groups (husband and wife). For example in medical research, a paired t-test is used to compare the means on a measure before (pre) and after (post) a treatment. Looking at market research, this test could be used to compare the rating an individual gives a product they usually purchase and a competing product on some characteristic.

Example:

In this example we use the data we use the same dataset `http://www.ats.ucla.edu/stat/data/hs0.dta`. In this data the same students took both the writing and the reading test. Hence, you would expect there to be a relationship between the scores provided by each student. The paired t-test accounts for this.

So the hypothesis we want to test is:

H_0: $\mu_{Write} = \mu_{Read}$
H_A: $\mu_{Write} \neq \mu_{Read}$

To run the t-test type **`ttest write=read, level(99)`**

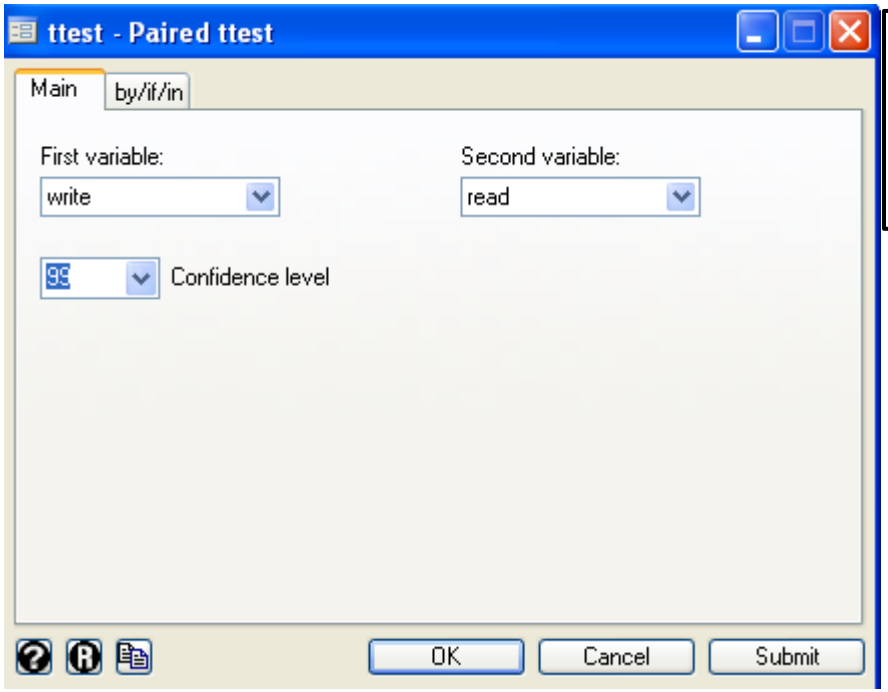
```
Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [99% Conf. Interval]
---------+--------------------------------------------------------------------
   write |     200      52.775    .6702372    9.478586    51.03187    54.51813
    read |     200       52.23    .7249921    10.25294    50.34447    54.11553
---------+--------------------------------------------------------------------
    diff |     200        .545    .6283822    8.886666   -1.089271    2.179271
------------------------------------------------------------------------------
    mean(diff) = mean(write - read)                               t =   0.8673
Ho: mean(diff) = 0                     degrees of freedom =      199

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.8066         Pr(|T| > |t|) = 0.3868          Pr(T > t) = 0.1934
```

In this example, the t-statistic is 0.8673 with 199 degrees of freedom. The corresponding two-tailed p-value is 0.3868, which is greater than 0.05. We conclude that the mean difference of write and read is not different from 0.

To perform the paired samples t-test by using the command menu

click `Statistics => Summaries, tables and tests => Classical tests of hypotheses => mean comparison test, paired data`



Choose the 2 variables that you want to test in this example: write and read. Select the 99 percent confidence interval.

Click OK.

Independent sample t-test

The independent samples t-test is also referred to as unpaired or unrelated samples t-test. It allows for us to compare the means observed for one variable for two independent samples.

Example:

In this example, we compare the mean writing score between the group of female students and the group of male students. Ideally, these subjects are randomly selected from a larger population of subjects. The test assumes that variances for the two populations are the same.

To run the t-test type **`ttest write, by(gender)`**

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
1	91	50.12088	1.080274	10.30516	47.97473	52.26703
2	109	54.99083	.7790686	8.133715	53.44658	56.53507
combined	200	52.775	.6702372	9.478586	51.45332	54.09668
diff		-4.869947	1.304191		-7.441835	-2.298059

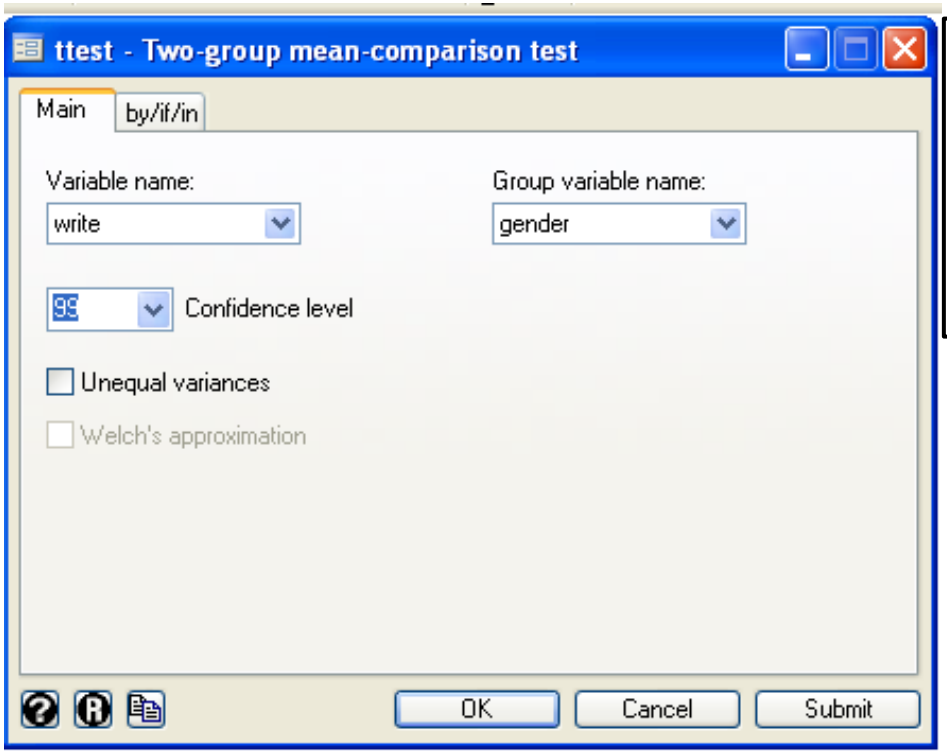
diff = mean(1) - mean(2) t = -3.7341
Ho: diff = 0 degrees of freedom = 198

Ha: diff < 0 Ha: diff != 0 Ha: diff > 0
Pr(T < t) = 0.0001 Pr(|T| > |t|) = 0.0002 Pr(T > t) = 0.9999

In this example, the t-statistic is -3.7341 with 198 degrees of freedom. The corresponding two-tailed p-value is 0.0002, which is less than 0.05. We conclude that the difference of means in **write** between males and females is different from 0.

To perform the paired samples t-test by using the command menu

click on: **Statistics => Summaries, tables and tests => Classical tests of hypotheses =>Two-group mean-comparison test**



Choose the variable that you want to compare the mean this example: write and the grouping variable in this example is gender. Select the 99 percent confidence interval.

Click OK.