Two Sample t-tests in Stata

Example: In the Framingham cohort, we want to examine the distribution of heart rate at exams 1 and 2. Specifically, we wish to test whether there is a difference in mean heart rate between exam 1 and exam 2. Additionally, we are interested in whether the mean heart rate differs between men and women at exam 2. We sample 100 people from the Framingham cohort. For this example, use the dataset `heartrate.dta` on this webpage, which contains the random sample of 100 participants.

Hypothesis testing with paired data in Stata:

```
. ttest heartrte1 == heartrte2

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
heartr~1 |     100       75.03    1.290247    12.90247    72.46987    77.59013
heartr~2 |     100       76.17    1.293031    12.93031    73.60435    78.73565
---------+--------------------------------------------------------------------
    diff |     100       -1.14    1.344125    13.44125   -3.807035    1.527035
------------------------------------------------------------------------------
     mean(diff) = mean(heartrte1 - heartrte2)                      t =  -0.8481
 Ho: mean(diff) = 0                               degrees of freedom =       99

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.1992         Pr(|T| > |t|) = 0.3984          Pr(T > t) = 0.8008


. gen hdiff = heartrte2 - heartrte1

. ttest hdiff== 0

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
   hdiff |     100        1.14    1.344125    13.44125   -1.527035    3.807035
------------------------------------------------------------------------------
    mean = mean(hdiff)                                           t =   0.8481
Ho: mean = 0                                     degrees of freedom =       99

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.8008         Pr(|T| > |t|) = 0.3984          Pr(T > t) = 0.1992
```

The commands `ttest heartrte2 == heartrte1` and `ttest hdiff==0` lead to the same test.

This command can be found through the following drop-down menus: Statistics / Summaries, tables, and tests / Classical tests of hypotheses / Mean-comparison test, paired data.

Hypothesis testing with unpaired data and equal variances in Stata:

```
. ttest heartrte2, by(sex1)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    Male |      39    76.82051    2.042025    12.75244    72.68665    80.95438
  Female |      61     75.7541    1.681246    13.13095    72.39111    79.11709
---------+--------------------------------------------------------------------
combined |     100       76.17    1.293031    12.93031    73.60435    78.73565
---------+--------------------------------------------------------------------
    diff |            1.066414    2.662326               -4.216884    6.349713
------------------------------------------------------------------------------
    diff = mean(Male) - mean(Female)                              t =   0.4006
Ho: diff = 0                                     degrees of freedom =       98

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.6552         Pr(|T| > |t|) = 0.6896          Pr(T > t) = 0.3448
```

Hypothesis testing with unpaired data and unequal variances in Stata:

```
. ttest heartrte2, by(sex1) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    Male |      39    76.82051    2.042025    12.75244    72.68665    80.95438
  Female |      61     75.7541    1.681246    13.13095    72.39111    79.11709
---------+--------------------------------------------------------------------
combined |     100       76.17    1.293031    12.93031    73.60435    78.73565
---------+--------------------------------------------------------------------
    diff |            1.066414    2.645081               -4.194674    6.327503
------------------------------------------------------------------------------
    diff = mean(Male) - mean(Female)                              t =   0.4032
Ho: diff = 0                     Satterthwaite's degrees of freedom =  82.8637

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.6561         Pr(|T| > |t|) = 0.6879          Pr(T > t) = 0.3439
```

This command can be found through the following drop-down menus: Statistics / Summaries, tables, and tests / Classical tests of hypotheses / Two-group mean-comparison test.

Instead of the data structure above, suppose that, in your dataset, you have heart rate for men in one variable/column and heart rate for women in another variable/column (instead of our situation where we have heart rate in one variable and sex as another variable). How do you perform a t-test then? Use the command `ttest heartratew == heartratem, unpaired unequal`, where heartratew is the heart rate variable for women and heartratem is the heart rate for men. It is important to use the option `unpaired`. If you do not use this option, Stata will perform a paired t-test. You may also choose the leave out the `unequal` option if you wish to assume equal variances.

The following 4 lines of code transform the data to the situation where we have heart rate for men in one variable (heartrtem) and heart rate for women in another variable (heartrtew). It is not necessary to memorize or understand this portion of code. It is simply included for completeness. The fifth line of code runs the two sample t-test.

```
. gen id = _n
. reshape wide heartrte2, i(id) j(sex1)
. rename heartrte21 heartrtem
. rename heartrte22 heartrtew

. ttest heartrtew = heartrtem, unpaired unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
heartr~w |      61     75.7541    1.681246    13.13095    72.39111    79.11709
heartr~m |      39    76.82051    2.042025    12.75244    72.68665    80.95438
---------+--------------------------------------------------------------------
combined |     100       76.17    1.293031    12.93031    73.60435    78.73565
---------+--------------------------------------------------------------------
    diff |           -1.066414    2.645081               -6.327503    4.194674
------------------------------------------------------------------------------
    diff = mean(heartrtew) - mean(heartrtem)                      t =  -0.4032
Ho: diff = 0                     Satterthwaite's degrees of freedom =  82.8637

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.3439         Pr(|T| > |t|) = 0.6879          Pr(T > t) = 0.6561
```

This command can be found through the following drop-down menus: Statistics / Summaries, tables, and tests / Classical tests of hypotheses / Two-sample mean-comparison test.

Exercises

1. Calculate the sample mean and sample standard deviation of heart rate at exam 1 and exam 2 in the Framingham cohort.
2. Are these data dependent or independent?
3. Generate a new variable for the difference in heart rate between exam 1 and exam 2. Make a histogram of this new variable.
4. Perform a hypothesis test at the $\alpha = 0.05$ level.
   (a) What test are you using?
   (b) State your null and alternative hypothesis.
   (c) Perform the hypothesis test. What are:
      i. your test statistic,
      ii. the degrees of freedom,
      iii. the p-value,
      iv. your decision, and
      v. your interpretation?

Now, assume that you are interested in whether the mean heart rate differs between men and women at exam 2.

5. Are these data dependent or independent?
6. Calculate the sample mean and sample standard deviation of heart rate at exam 2 for men and women.
7. Perform a hypothesis test at the $\alpha = 0.05$ level, assuming unequal variances.
   (a) What test are you using?
   (b) State your null and alternative hypothesis.
   (c) Perform the hypothesis test. What are:
      i. your test statistic,
      ii. the degrees of freedom,
      iii. the p-value,
      iv. your decision, and
      v. your interpretation?
8. Given the 95% confidence intervals, would you expect the hypothesis test to be significant?