Stata How-to:
Test Hypotheses

Patrick Blanchenay

2021-12-27

Contents

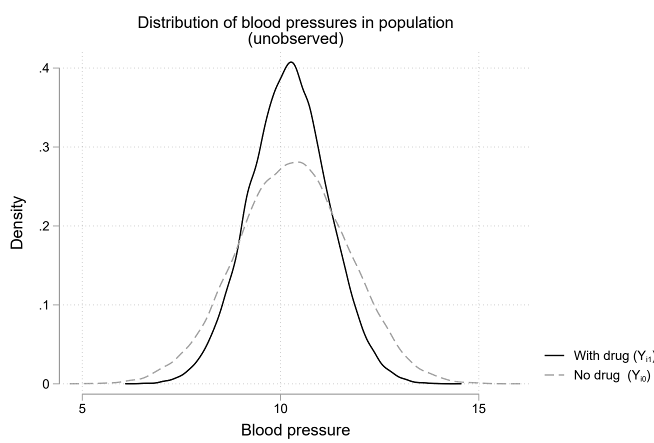
1. Testing equality of means

In Randomized Controlled Trials, individuals are typically allocated in a treatment and a control group. The treatment group receives the treatment, while the control group doesn't. Estimating the effectiveness of the treatment is done by testing whether the mean of the outcome variable is different in the treatment versus control group.

This is done using a *t*-test.

1.1. Intuition behind the *t*-test

Let's take an example: we are testing a drug designed to lower blood pressure, which we denote Y_i. People naturally vary in their blood pressure. For an individual i, let's denote Y_{i0} the potential blood pressure they would have if they do not take the drug, and Y_{i1} if they take the drug. The dashed line on the graph below shows what the hypothetical distribution of blood pressures would be in the population if no one received the drug: the distribution of Y_{i0}. The solid line shows hypothetical blood pressures if everyone in the population received the drug: the distribution of Y_{i1}. We do not observe either of these distributions.



The drug would be considered effective if the two distributions differed. However, since it is hard to compare curves, let's restrict our question: do people who take the drug have *on average* a lower blood pressure than those who don't take the drug? In other words: is it the case that $\mathbb{E}[Y_{i1}] \neq \mathbb{E}[Y_{i0}]$? Let's denote these μ_1 and μ_0. We are testing the null hypothesis H_0 : "$\mu_0 = \mu_1$" against the research hypothesis H_1 : "$\mu_0 \neq \mu_1$".

We never get to observe the full population so never get to know μ_0 and μ_1 so we have to rely on a sample.

We randomly allocate individuals into a treatment group who receives the drug ($D_i = 1$) and a control group that doesn't ($D_i = 0$). For the treatment group we observe their blood pressure after taking the drug, that is, we observe $Y_i = Y_{i1}$, while for the control group we observe their blood pressure without the drug: $Y_i = Y_{i0}$. Let's denote the average blood pressure in each group $\text{Avg}[Y_i \mid D_i = 0]$ and $\text{Avg}[Y_i \mid D_i = 1]$; some books use the notation $\bar{Y}_0$ and $\bar{Y}_1$.

The law of large numbers implies that if we take a large number of observations from the distribution of Y_{i0}, the average of these observations will be "close to" μ_0. In other words, $\text{Avg}[Y_i \mid D_i = 0]$ should be close to μ_0. It will never be exactly equal to μ_0: in a sample, it might be slightly higher, but if we were to redraw a sample of blood pressures, the average would be different again, and might be slightly lower than μ_0 this time.

Similarly, $\text{Avg}[Y_i \mid D_i = 1]$ should be close to μ_1.

So if indeed $\mu_1 = \mu_0$, the group averages $\text{Avg}[Y_i \mid D_i = 0]$ and $\text{Avg}[Y_i \mid D_i = 1]$ should be close to each other.

More precisely, if indeed $\mu_1 = \mu_0$ (we say "under the null"), the Central Limit Theorem implies that the rescaled difference in averages $\frac{\text{Avg}[Y_i|D_i=1]-\text{Avg}[Y_i|D_i=0]}{sd(\text{Avg}[Y_i|D_i=1]-\text{Avg}[Y_i|D_i=0])}$ follows a Student t distribution: a bell-shaped distribution centered around 0, and with slightly fatter tails than the Standard Normal distribution. We denote this rescaled difference t and call it a t-statistic.

This means that if indeed $\mu_1 = \mu_0$, the t-statistic should be in the vicinity of zero. *Conversely*, if we do observe that t is far from zero, we take it as sufficient evidence to reject the null, in other words that $\mu_1 \neq \mu_0$. Testing whether t is far from zero is called a t-test.

What do we mean by "far from zero"? It depends on how many observations you have. But if you have very large samples (eg $n = 500$), then far from zero means less than -1.96 or more than 1.96.

1.2. *t*-test in Stata

Testing whether averages of a variable (e.g. blood pressure) are "sufficiently" different between two groups (e.g. treatment an control) is done by using the **`ttest`** command. Let's use a fictitious dataset `Blood_pressure_fictitious.dta`. Blood pressure is recorded in `bloodpressure`, and whether the person takes the drug is the dummy `drug`. So the dummy drug defines who is in the treatment group (`drug == 1`) and who is in the control group (`drug == 0`). Testing equality of mean blood pressure in the two groups is done by:

```
ttest bloodpressure, by(drug) unequal
```

This tests whether the variable `bloodpressure` is the same in groups defined by the dummy variable `drug`. The **`unequal`** option is important: it tells Stata that the variance (spread) of blood pressure could be different in the two groups, and should always be included.[1]

[1] If you had access to the population distributions like the graph above, you would be able to tell whether the variances are different: in our case the variance is higher among people who do not take the drug. But in general you do not know anything about the population distribution (otherwise you would not be doing hypothesis testing); so you should assume that this could be the case.

Warning

Always include the **unequal** option when doing a t-test.

```
Two-sample t test with unequal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |     129    10.36465    .1207954    1.371972    10.12563    10.60366
       1 |     119    10.19693    .0992434    1.082618     10.0004    10.39346
---------+--------------------------------------------------------------------
combined |     248    10.28417    .0788636    1.241945    10.12884     10.4395
---------+--------------------------------------------------------------------
    diff |             .167712    .1563355               -.1402508    .4756748
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   1.0728
Ho: diff = 0                     Satterthwaite's degrees of freedom =  240.337

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.8578         Pr(|T| > |t|) = 0.2845          Pr(T > t) = 0.1422
```

In this output, **Group 0** denotes individuals for whom `drug == 0`, and **Group 1** denotes individuals for whom `drug == 1`. So we have 248 observations, 129 of whom did not take the drugs and 119 who did. Average blood pressure in the control group is 10.36, while average blood pressure in the treatment group is 10.20. The difference between control and treatment (in that order) is 0.168. The corresponding t statistic is given in the lower right of the table, here $t = 1.07$. Make sure you know how to find all these figures in the output above.

The t-statistic is not sufficiently far from zero to reject the null: we do not have sufficient evidence to claim that drug-takers have a lower blood pressure than non-drug-takers.

2. t-test and regressions

Once we cover regressions, you will see that difference in blood pressure could have been tested using a regression, by estimating:

$$\text{BloodPressure}_i = \alpha + \beta \text{Drug}_i + \varepsilon_i \quad (1)$$

(See Stata How-to: OLS regressions .)

In the above model, β captures how much higher (or lower) blood pressure is among those who take the drug ($\text{Drug}_i = 1$) versus those who do not ($\text{Drug}_i = 0$). Testing whether blood pressures differ in the two groups or not ($\mu_0 = \mu_1$) is the same as testing whether $\beta = 0$.

```
reg bloodpressure drug, robust // robust option equivalent to unequal option in ttest
```

```
Linear regression                               Number of obs     =        248
                                                F(1, 246)         =       1.15
                                                Prob > F          =     0.2844
                                                R-squared         =     0.0046
                                                Root MSE          =     2.2416

------------------------------------------------------------------------------
             |               Robust
bloodpress~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        drug |   -.167712   .1563395    -1.07   0.284    -.4756467    .1402227
       _cons |   10.36465   .1208145    85.79   0.000     10.12668    10.60261
------------------------------------------------------------------------------
```

Notice how the regression output contains the same information as the t-test:

- the estimate for the constant $\hat{\alpha}$ is the average in the no-drug group: 10.36;
- the estimate for the slope $\hat{\beta}$ is the difference in blood pressures -0.168;
- the sum of $\hat{\alpha} + \hat{\beta}$ is the average blood pressure in the drug group: 10.20;
- the t-statistic associated with the coefficient β, testing whether $\beta = 0$, is the same as in our two-sample t-test: -1.07.