Section 7.2 - Comparing Two Means

Statistics 104

Autumn 2004

Comparing Two Means

Two-sample problems

- Want to compare the responses in two groups or treatments
- Each sample is considered to be a sample from a distinct population
- The responses in each group are independent of those in the other group

Setup:

Population	Variable	Mean	Standard Deviation
1	x_1	μ_1	σ_1
2	x_2	μ_2	σ_2

Want to

- Construct confidence intervals for $\mu_1 - \mu_2$

- Examine hypotheses for the form

  $H_0 : \mu_1 = \mu_2$ or $H_0 : \mu_1 - \mu_2 = \delta$ ($\delta = 0$ often)

Inference is based on two independent SRS, one from each population. The sample setups are

Population	Sample Size	Sample Mean	Sample Standard Deviation
1	n_1	$\bar{x}_1$	s_1
2	n_2	$\bar{x}_2$	s_2

An additional assumption underlying most of the following is that the observations are normally distributed (at least approximately).

Two-sample z statistics

These are appropriate if the standard deviations are known. The are based on the facts that $\bar{X}_1 - \bar{X}_2$ has mean

$$\mu_1 - \mu_2$$

and variance

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

If the observations from each group are normally distributed, then so is $\bar{X}_1 - \bar{X}_2$. If not, then $\bar{X}_1 - \bar{X}_2$ will be approximately normally distributed if both sample sizes are large.

This leads to a $100C\%$ CI for $\mu_1 - \mu_2$ of

$$(\bar{x}_1 - \bar{x}_2) \pm z^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where z^* is the same as for the one-sample problem.

The hypothesis test of $H_0 : \mu_1 - \mu_2 = \delta$ is based on the two-sample z statistics

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

which is compared to a $N(0, 1)$.

Note that these procedures are rarely used, since the population variances are rarely known. However they do motivate the t procedures that usually used.

Two-sample t Statistics

The inference on means in the two sample problem is often based on the standardized quantity

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Unfortunately, this doesn't have a t distribution, except in special cases.

However often it will be approximately t distributed. We need to come up with an approximation to the degrees of freedom k. There are two popular approaches

1. Calculate k from the data

$$k = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{s_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{s_2^2}{n_2}\right)^2}$$

   (Satterthwaite's degrees of freedom)

2. Set $k = \min(n_1 - 1, n_2 - 1)$

Most software uses option 1, unless the user requests something else. (e.g. Stata has this and another df formula available)

If you don't have software handy, use option 2.

Note that the second option is conservative. It gives confidence intervals that are slightly too wide and p-values that are a little too big (making it harder to reject).

The reason is that the complicated formula will give you something bigger than option 2.

Two-sample t test

$$H_0 : \mu_1 - \mu_2 = \delta$$

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

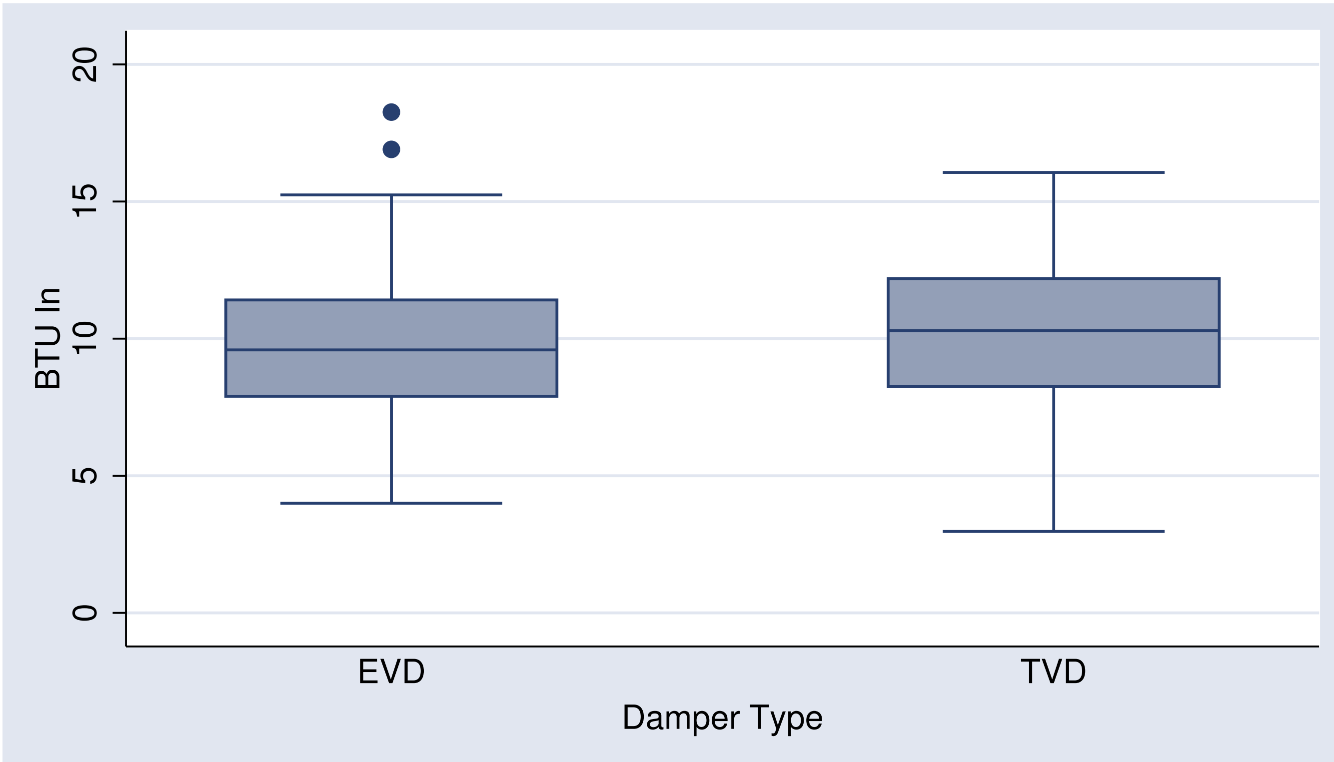
Use critical values and p-values based on the t distribution with k degrees or freedom, with k either calculated by the complicated formula or $\min(n_1 - 1, n_2 - 1)$.

Example: Furnace Dampers

Study performed by Wisconsin Power and Light

Interested whether energy use would be lower with the use of electric vent dampers (EVD) as compared to thermally controlled vent dampers (TVD).

The energy used (`btuin`) was the fuel comsumption, adjusted for weather and house size.



Summary Statistics

EVD: $n_1 = 40$ $\bar{x}_1 = 9.908$ $s_1 = 3.020$

TVD: $n_2 = 50$ $\bar{x}_2 = 10.143$ $s_2 = 2.767$

$$\bar{x}_1 - \bar{x}_2 = 9.908 - 10.143 = -0.235$$

$$SE_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{3.020^2}{40} + \frac{2.767^2}{50}} = 0.617$$

$$t = \frac{-0.235}{0.617} = -0.381$$

$H_0 : \mu_1 = \mu_2$ versus $H_A : \mu_1 \neq \mu_2$

$$p - \text{value} = 2 \times P[T \geq |-0.381|]$$

1. $k = 80.2$ (Satterthwaite)

   p-value $= 2 \times P[T \geq |-0.381|] = 2 \times 0.3521 = 0.7042$

2. $k = \min(40-1, 50-1) = 39$

   p-value $= 2 \times P[T \geq |-0.381|] = 2 \times 0.3526 = 0.7052$

(Often the difference between the two approximations will be a bit bigger.)

To get the p-values in Stata, you can use a command like

```
. generate pvalue = ttail(39,0.381)

. list pvalue in 1

     +----------+
     |   pvalue |
     |----------|
  1. | .3526354 |
     +----------+
```

Note that if you want to reuse the variable for another p-value calculation, do

```
. replace pvalue = ttail(39,0.381)
```

So there is little evidence that the new EVD dampers lead to different energy consumption.

Even though there appears to be little evidence that the two dampers are different, lets look at a confidence interval for the mean difference.

Two-sample t confidence interval for $\mu_1 - \mu_2$

A $100C\%$ CI for $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where t^* is the appropriate t critical value.

For the example, a 95% CI for $\mu_1 - \mu_2$ is

1. $k = 80.2 \rightarrow t^* = 1.990$

$$\begin{aligned} CI &= -0.235 \pm 1.990 \times 0.617 \\ &= -0.235 \pm 1.228 = (-1.463, 0.993) \end{aligned}$$

2. $k = 39 \rightarrow t^* = 2.023$

$$\begin{aligned} CI &= -0.235 \pm 2.023 \times 0.617 \\ &= -0.235 \pm 1.248 = (-1.483, 1.013) \end{aligned}$$

Note the intervals are very similar, regardless of which approach you take to getting the degrees of freedom.

To get critical values in Stata, you can use a command like

```
. gen tcrit = invttail(80,0.025)

. list tcrit in 1

     +----------+
     |    tcrit |
     |----------|
  1. | 1.990063 |
     +----------+
```

Stata Output:

```
. ttest btuin, by(damper) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------
   Group |  Obs      Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------
     EVD |   40  9.90775    .4774831    3.019868    8.941949    10.87355
     TVD |   50   10.143    .3913156    2.767019    9.356622    10.92938
---------+--------------------------------------------------------------
combined |   90 10.03844    .3023127     2.86799    9.437755    10.63913
---------+--------------------------------------------------------------
    diff |     -.2352499    .6173476               -1.463766    .9932662
------------------------------------------------------------------------
Satterthwaite's degrees of freedom:  80.1897

                    Ho: mean(EVD) - mean(TVD) = diff = 0

     Ha: diff < 0              Ha: diff != 0              Ha: diff > 0
       t =  -0.3811              t =  -0.3811              t =  -0.3811
   P < t =   0.3521        P > |t| =   0.7042          P > t =   0.6479
```

Study Design

When designing a study, you usually want to have equal sample sizes for both groups as this will usually make the standard error small. In addition, equal sample sizes helps with robustness of the procedures where there is non-normality of the observations within each group.

Exceptions to this can occur due to

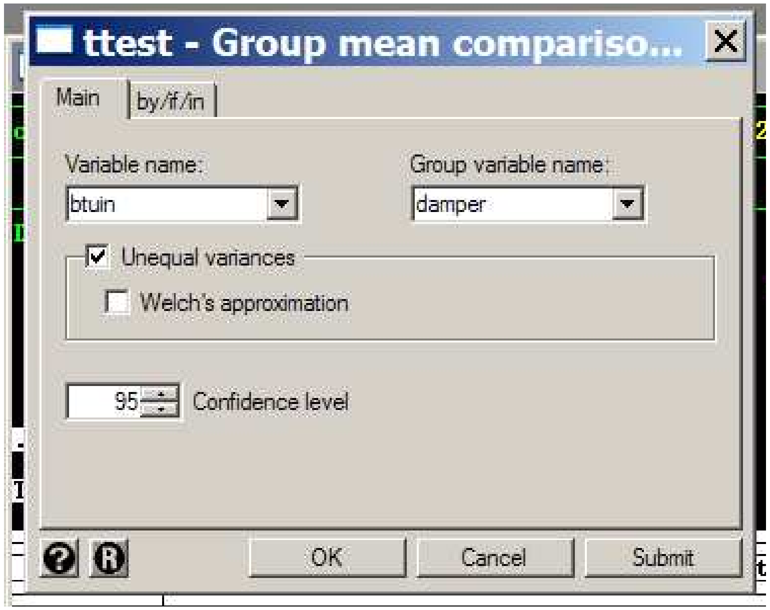
- Cost:

  If one more treatment is much more expensive than the other, you may get a more efficient design, for a given outlay of resources, by having an unbalanced allocation.

- Within group standard deviations:

  If one group is much more variable (higher σ) than the other, you want to place more observation in that group to lower the standard error of the differences of the mean.

Pooled two-sample t procedures



If you use the menu versions of the Stata command you will get a dialog box similar to the left.

Notice the check box for `Unequal variances`.

In the earlier analyses we made no assumptions about the relationship between σ_1 and σ_2.

The default approach in Stata is to assume that $\sigma_1 = \sigma_2 = \sigma$.

That was the reason for the unequal option in the earlier Stata output.

If we make this constant variance assumption, we need to change our analysis slightly.

First,

$$\mathrm{Var}(\bar{X}_1 - \bar{X}_2) = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)$$

So we need to estimate the common variance for the two groups.

The sample variances for the two groups, s_1^2 and s_2^2 both estimate this quantity. We need to combine them to given a better estimate. The usual approach is to take a weighted average with weights equal to their degrees of freedom.

This gives the estimator

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

which is known as the pooled estimator of variance. This estimator as $n_1 + n_2 - 2$ degrees of freedom.

The standard error of $\bar{X}_1 - \bar{X}_2$ based on s_p^2 is

$$SE_{\bar{x}_1 - \bar{x}_2} = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

This gives the following pooled procedures

A $100C\%$ CI for $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) \pm t^* s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

To test the hypothesis $H_0 : \mu_1 - \mu_2 = \delta$, use the test statistic

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - \delta}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

For critical values and p-values, use the t distribution with $n_1 + n_2 - 2$ degrees of freedom.

For the damper example

$$\begin{aligned} s_p^2 &= \frac{39 \times 3.020^2 + 49 \times 2.767^2}{40 + 50 - 2} = 8.305 \\ s_p &= 2.882 \\ df &= 88 \end{aligned}$$

Note that s_p must lie between s_1 and s_2 (as it does here).

Then

$$SE_{\bar{x}_1 - \bar{x}_2} = 2.882\sqrt{\frac{1}{40} + \frac{1}{50}} = 0.611$$

Therefore a 95% CI for $\mu_1 - \mu_2$ is

$$\begin{aligned} CI &= -0.235 \pm 1.987 \times 0.611 \\ &= -0.235 \pm 1.215 = (-1.450, 0.980) \end{aligned}$$

To test $H_0 : \mu_1 = \mu_2$

$$t = \frac{-0.235}{0.611} = -0.385$$

$$p - \text{value} = 2 \times P[T \geq |-0.385|] = 2 \times 0.3506 = 0.7013$$

Stata Output of pooled analysis

```
. ttest btuin, by(damper)

Two-sample t test with equal variances
------------------------------------------------------------------------
   Group |  Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------
     EVD |   40    9.90775    .4774831    3.019868    8.941949    10.87355
     TVD |   50     10.143    .3913156    2.767019    9.356622    10.92938
---------+--------------------------------------------------------------
combined |   90   10.03844    .3023127     2.86799    9.437755    10.63913
---------+--------------------------------------------------------------
    diff |      -.2352499    .6113255               -1.450131     .979631
------------------------------------------------------------------------
Degrees of freedom: 88

                    Ho: mean(EVD) - mean(TVD) = diff = 0

     Ha: diff < 0               Ha: diff != 0              Ha: diff > 0
       t =  -0.3848               t =  -0.3848               t =  -0.3848
   P < t =   0.3506         P > |t| =   0.7013           P > t =   0.6494
```

The pooled procedure is more powerful if the equal variance assumption is reasonable. However if there is a difference in variances, particularly if the sample sizes are different, the pooled procedures break down (not very robust). In general, use the unpooled procedures, unless the sample sizes are similar.