9. Hypothesis testing: two groups

LPO.8800: Statistical Methods in Education Research

Sean P. Corcoran

Hypothesis testing thus far

- Hypothesis tests for a population mean μ (σ known or unknown)
- Hypothesis tests for a population proportion π
- Null and alternative hypotheses (H_0 and H_1)
- One vs. two-sided alternative hypotheses
- Significance levels (α) and p-values
- Type I and Type II errors
- Power of a test
- Practical significance and effect size

Statistical tests for comparing two groups

Hypothesis tests are frequently used to make inferences about how two population parameters compare:

- Do female executives earn less on average than males?
- Do 4th graders in an experimental reading program perform differently on standardized reading tests than 4th graders not in the program?
- Are women more likely to vote for Democratic candidates than men?
- Do subjects participating in a 6-week weight loss program lose more weight over time than those who do not participate in the program?
- Has obesity among children aged 10-12 increased between 2000 and 2016?
- Are COVID infection rates higher in counties without a mask mandate than counties with them?

Statistical tests for comparing two groups

One can think of these examples as **bivariate analyses** involving two variables:

- Group identifier: a binary *explanatory variable*
- Outcome: the *response*

Statistical tests for comparing two groups

Each of the above examples is a comparison of two parameters. For example:

- A comparison of means across groups (μ_1 and μ_2)
- A comparison of proportions across groups (π_1 and π_2)
- A comparison of means or proportions *over time*
- A comparison of means of the *same group* pre- and post-treatment ("within subject")

When making comparisons of two parameters we usually construct a test for the *difference* in those parameters (e.g., $\mu_2 - \mu_1$, or $\pi_2 - \pi_1$).

Statistical tests for comparing two groups

The steps for conducting a test comparing two groups are the same as those for the test of a single parameter. The most common null hypothesis is that there is *no difference* between the two population means:

$$H_0 : \mu_1 = \mu_2$$

Or equivalently,

$$H_0 : \mu_2 - \mu_1 = 0$$

Note: it doesn't matter which mean you subtract from the other, as long as you keep track and are consistent in your ordering throughout the test.

Statistical tests for comparing two groups

The alternative hypothesis H_1 is that there *is* a difference (a two-sided alternative) or that one mean is greater than the other (a one-sided alternative):

Two-sided alternative:

$$H_1 : \mu_2 - \mu_1 \neq 0$$

One-sided alternatives:

$$H_1 : \mu_2 - \mu_1 > 0$$

$$H_1 : \mu_2 - \mu_1 < 0$$

Hypothesis test steps

1. Determine H_0 and H_1.
2. The estimator for a difference in population means is the difference in sample means: $\bar{x}_2 - \bar{x}_1$. Determine its sampling distribution under H_0. (This requires knowing the standard error of $\bar{x}_2 - \bar{x}_1$).
3. Determine the probability of obtaining your observed test statistic if H_0 is true (the p-value), and draw a conclusion.

The standard error of $\bar{x}_2 - \bar{x}_1$ depends on how the two samples were drawn (next slide).

Independent vs. dependent samples

The design of a study—and in particular, whether the two samples being compared are independent or dependent—is important to statistical testing, because standard errors depend on how the samples were drawn.

- Samples are **dependent** when there is a natural matching between subjects in each sample. Examples include repeated measures on the *same* subjects (a **longitudinal study**), siblings, marriage partners, etc.
- Samples are **independent** when there is no such matching (e.g., random draws from two populations). Selection of subjects into one sample has no effect on the selection of subjects into the second.

With dependent samples, pairs of outcomes do not represent independent draws from a population—they are likely to be correlated.

Experimental vs. observational group assignment

- In **experimental** designs, subjects are *randomly* assigned to groups (e.g., treatment and control).
- In **observational** designs, subjects are in naturally-occurring groups that they may or may not have control over (e.g., gender vs. political affiliation).

The distinction is important for causal inference and interpretation, though not for statistical inference (e.g., comparing means in the population).

Sampling distribution of $\bar{x}_2 - \bar{x}_1$

Over repeated samples, the difference in two means drawn from independent samples will have a mean of $\mu_2 - \mu_1$ (the difference in the *true* means—it is an *unbiased* estimator) and a standard error of:

$$se_{\bar{x}_2 - \bar{x}_1} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Notice this is a larger number than either of the individual standard errors alone (for $\bar{x}_1$ or $\bar{x}_2$)! Also note this is expression is: $\sqrt{(se_1)^2 + (se_2)^2}$

Intuitively, $(\bar{x}_2 - \bar{x}_1)$ is often further from $(\mu_2 - \mu_1)$ than $\bar{x}_1$ is from μ_1 or $\bar{x}_2$ is from μ_2.

Sampling distribution of $\bar{x}_2 - \bar{x}_1$

If the samples are independent *and* the sample sizes n_1 and n_2 are sufficiently large, then we can also say the sampling distribution of $\bar{x}_2 - \bar{x}_1$ (divided by its standard error) has an approximately normal distribution.

In practice, σ_1 and σ_2 are unknown, so the sample standard deviations are used in their place (s_1 and s_2):

$$se_{\bar{x}_2 - \bar{x}_1} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

We then refer to the t-distribution instead of the standard normal. (Which df to use is discussed momentarily).

Confidence interval for $\mu_2 - \mu_1$

With this information, a $(1-\alpha)\%$ confidence interval for the difference in population means $\mu_2 - \mu_1$ is:

$$(\bar{x}_2 - \bar{x}_1) \pm t_{\alpha/2}(se_{\bar{x}_2 - \bar{x}_1})$$

As before, a $(1-\alpha)\%$ confidence interval can be used to test two-sided alternative hypotheses with significance level α.

Hypothesis test for $\mu_2 - \mu_1$

Alternatively, a test statistic can be calculated for a specific hypothesis. For example:

$$H_0 : \mu_2 - \mu_1 = 0$$

$$H_1 : \mu_2 - \mu_1 \neq 0$$

Under the assumption that H_0 is true, the test statistic is as follows, using the standard error formula given above:

$$t = \frac{(\bar{x}_2 - \bar{x}_1) - 0}{se_{\bar{x}_2 - \bar{x}_1}}$$

Use the t distribution to find the p-value associated with this t. If the two sample sizes are large enough, the standard normal (z) distribution can be used instead of t.

Hypothesis test for $\mu_2 - \mu_1$

The degrees of freedom for the t-statistic in this case is complex (can use the Welch-Satterthwaite approximation). However, if we assume that the two distributions from which the samples are drawn have *equal variances*, the degrees of freedom will be $df = n_1 + n_2 - 2$.

Stata's default is to assume equal variances and use the simplified df. However, can opt for unequal variance assumption.

Welch-Satterthwaite approximation

FYI the Welch-Satterthwaite approximation for the degrees of freedom in a two-sample independent t-test is

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

Hypothesis test for $\mu_2 - \mu_1$

Note: if we assume equal variances, then we can also use a pooled estimate of the variance (s_p^2), and the standard error for $\bar{x}_2 - \bar{x}_1$ simplifies to:

$$se_{\bar{x}_2 - \bar{x}_1} = \sqrt{s_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

The pooled variance estimate is a weighted average of s_1^2 and s_2^2:

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

Note: s_p^2 is close—but not the same as—the combined variance (i.e., if you were to combine the data and treat it like one sample).

Example 1

What is the impact of alcoholism during pregnancy on the IQ of infants? IQs of infants of 6 women with alcoholism (group 1) were compared with those of infants of 46 women without alcoholism (group 2).

- $n_1 = 6$, $\bar{x}_1 = 78$, $s_1^2 = 361$
- $n_2 = 46$, $\bar{x}_2 = 99$, $s_2^2 = 256$

$$H_0 : \mu_2 - \mu_1 = 0$$
$$H_1 : \mu_2 - \mu_1 \neq 0$$

Example 1

The 95% confidence interval is

$$(\bar{x}_2 - \bar{x}_1) \pm t_{\alpha/2}(se_{\bar{x}_2-\bar{x}_1})$$

Assuming equal variances, the t-statistic here is $t(50, 0.025) = 2.01$, where $df = n_1 + n_2 - 2$. The standard error is:

$$s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{(n_1-1)+(n_2-1)} = \frac{(5)361 + (45)256}{5+45} = 266.5$$

$$se_{\bar{x}_2-\bar{x}_1} = \sqrt{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{266.5\left(\frac{1}{6}+\frac{1}{46}\right)} = 7.086$$

Example 1

So:

$$(\bar{x}_2 - \bar{x}_1) \pm t_{\alpha/2}(se_{\bar{x}_2-\bar{x}_1})$$

$$(99-78) \pm 2.01(7.086) = (6.757, 35.242)$$

The confidence interval does not contain zero so we can reject H_0 at the $\alpha = 0.05$ significance level.

Example 1

Alternatively, can calculate the test statistic:

$$t = \frac{\bar{x}_2 - \bar{x}_1 - 0}{se_{\bar{x}_2 - \bar{x}_1}}$$

$$t = \frac{99 - 78 - 0}{7.086} = 2.964$$

The reference distribution for looking up the p-value is $t(50)$. In Stata: `display ttail(50,2.964) = 0.0023`. Since $p < \alpha$, we reject H_0.

Example 1: Using Stata `ttesti`

Can use the t-test calculator in Stata to obtain these results. Syntax below or use the drop-down menus. Default assumes equal variances, and *n, m, s* below refer to the two group sample sizes, means, and standard deviations.

`ttesti` *`n1 m1 s1 n2 m2 s2`*

```
. ttesti 6 78 19 46 99 16

Two-sample t test with equal variances
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |       6          78    7.756718          19    58.06072    97.93928
       y |      46          99    2.359071          16    94.24859    103.7514
---------+--------------------------------------------------------------------
combined |      52    96.57692    2.430458    17.52628    91.69758    101.4563
---------+--------------------------------------------------------------------
    diff |                 -21    7.085912               -35.23247   -6.767527
------------------------------------------------------------------------------
    diff = mean(x) - mean(y)                                      t =  -2.9636
Ho: diff = 0                                     degrees of freedom =       50

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0023         Pr(|T| > |t|) = 0.0046          Pr(T > t) = 0.9977
```

Example 1: Using Stata `ttesti`

This is a case where the equal variance assumption makes a big difference:

`ttesti` *n1 m1 s1 n2 m2 s2*, `unequal`

```
. ttesti 6 78 19 46 99 16, unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
         |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       x |       6          78    7.756718          19    58.06072    97.93928
       y |      46          99    2.359071          16    94.24859    103.7514
---------+--------------------------------------------------------------------
combined |      52    96.57692    2.430458    17.52628    91.69758    101.4563
---------+--------------------------------------------------------------------
    diff |                 -21     8.10752                -40.86902   -1.130985
------------------------------------------------------------------------------
    diff = mean(x) - mean(y)                                      t =  -2.5902
Ho: diff = 0                     Satterthwaite's degrees of freedom =  5.96208

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0207         Pr(|T| > |t|) = 0.0414          Pr(T > t) = 0.9793
```


Example 1: Using Stata `ttesti`

Welch-Satterthwaite approximation for the degrees of freedom in this case:

$$df = \frac{\left(\frac{361}{6} + \frac{256}{46}\right)^2}{\frac{1}{6-1}\left(\frac{361}{6}\right)^2 + \frac{1}{46-1}\left(\frac{256}{46}\right)^2}$$

$$df = 5.96$$

(substantially less than $n_1 + n_2 - 2$)

Example 2: Using Stata

Use NELS to test the hypothesis that male and female college-bound high school graduates in the South have equal years of preparation in high school math (*unitmath*). Use α=0.05. Let group 2 be males and group 1 be females.

$$H_0 : \mu_2 - \mu_1 = 0$$

$$H_1 : \mu_2 - \mu_1 \neq 0$$

Note: in the NELS, college-bound is *edexpect*$>=$ 2 and the South is *region*$==$ 3.

Example 2: Using Stata

In Stata use `ttest varname, by(groupvar)`. The default assumes equal variances. Use the `unequal` option otherwise (this will affect the standard error calculation).

```
. ttest unitmath if edexpect>=2 & region==3, by(gender)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    male |      57    3.938596    .0928778    .7012118     3.75254    4.124653
  female |      76    3.747105    .0783084    .6826767    3.591107    3.903104
---------+--------------------------------------------------------------------
combined |     133    3.829173    .0602281    .6945841    3.710036     3.94831
---------+--------------------------------------------------------------------
    diff |            .1914912     .121017               -.0479093    .4308918
------------------------------------------------------------------------------
    diff = mean(male) - mean(female)                              t =   1.5823
Ho: diff = 0                                     degrees of freedom =      131

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9420         Pr(|T| > |t|) = 0.1160          Pr(T > t) = 0.0580
```

Example 2: Using Stata

From the above Stata output:

$$\bar{x}_2 - \bar{x}_1 = 0.1915$$

$$se_{\bar{x}_2 - \bar{x}_1} = 0.1210$$

$$t = 1.582$$

$$p = 0.116$$

Conclusion: do not reject H_0, since $p > \alpha$.

Example 2: Using Stata

Note: the above Stata output reports the combined standard deviation (0.6945841). The square of this (0.482447) is close, but is not the same as the pooled variance used to calculate the standard error of the difference:

$$s_p^2 = \frac{(57-1) * 0.7012^2 + (76-1) * 0.6827^2}{(57-1) + (76-1)} = 0.4770$$

Then:

$$se_{\bar{x}_2 - \bar{x}_1} = \sqrt{0.4770 \left(\frac{1}{57} + \frac{1}{76} \right)} = 0.121017$$

Example 2: Using Stata

In practice, standard errors, test statistics and p-values will be similar whether assuming equal variances or not, **if**:

- n_1 and n_2 are similar or
- s_1^2 and s_2^2 are similar

Example 2: Using Stata

Same example, using `unequal` option

```
. ttest unitmath if edexpect>=2 & region==3, by(gender) unequal

Two-sample t test with unequal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    male |      57    3.938596    .0928778    .7012118     3.75254    4.124653
  female |      76    3.747105    .0783084    .6826767    3.591107    3.903104
---------+--------------------------------------------------------------------
combined |     133    3.829173    .0602281    .6945841    3.710036     3.94831
---------+--------------------------------------------------------------------
    diff |            .1914912    .1214845                 -.04906    .4320424
------------------------------------------------------------------------------
    diff = mean(male) - mean(female)                              t =   1.5763
Ho: diff = 0                     Satterthwaite's degrees of freedom =  119.011

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9412         Pr(|T| > |t|) = 0.1176          Pr(T > t) = 0.0588
```

Standard error in this case is $\sqrt{(se_1)^2 + (se_2)^2}$, not the calculation with s_p^2

Sampling distribution of $\hat{\pi}_2 - \hat{\pi}_1$

Over repeated samples, the difference in two proportions drawn from independent samples will have a mean of $\pi_2 - \pi_1$ (the *true* difference in means) and a standard error of:

$$se_{\hat{\pi}_2 - \hat{\pi}_1} = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

If the samples are independent *and* the sample sizes n_1 and n_2 are sufficiently large, then we can also say the sampling distribution of $\hat{\pi}_2 - \hat{\pi}_1$ has an approximately normal distribution. (I assume this below, in using z in the confidence interval).

Confidence interval for $\pi_2 - \pi_1$

With this information, a $(1-\alpha)$% confidence interval for the difference in population proportions $\pi_2 - \pi_1$ is:

$$(\hat{\pi}_2 - \hat{\pi}_1) \pm z_{\alpha/2}(se_{\hat{\pi}_2 - \hat{\pi}_1})$$

A $(1-\alpha)$% confidence interval can be used to test two-sided alternative hypotheses with significance level α.

Hypothesis test for $\pi_2 - \pi_1$

Alternatively, a test statistic can be calculated for a specific hypothesis. For example:

$$H_0 : \pi_2 - \pi_1 = 0$$

$$H_1 : \pi_2 - \pi_1 \neq 0$$

Under the assumption that H_0 is true, the test statistic is as follows, using a modified standard error se_0 (see next slide).

$$z = \frac{\hat{\pi}_2 - \hat{\pi}_1 - 0}{se_0}$$

Hypothesis test for $\pi_2 - \pi_1$

Under H_0 the two population proportions are the same. This directly implies equal variances, since for proportions the variance is $\pi(1-\pi)$. If H_0 is true, we can use a *pooled estimate* of π in the standard error formula that uses the combined samples rather than separate estimates of π_2 and π_1. The standard error using the pooled estimate is:

$$se_0 = \sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

where $\hat{\pi}$ is the overall proportion equal to one in the combined sample. This is used in the test statistic (z) above.

Example 3

In the latest YouGov poll (Oct 25, 2020), 44% of 674 men and 36% of 826 women stated that they intend to vote for Donald Trump. Construct a 95% confidence interval for the difference in these two population proportions (men - women).

$$(\hat{\pi}_2 - \hat{\pi}_1) \pm z_{\alpha/2}(se_{\hat{\pi}_2 - \hat{\pi}_1})$$

$$(0.44 - 0.36) \pm 1.96(se_{\hat{\pi}_2 - \hat{\pi}_1})$$

$$se_{\hat{\pi}_2 - \hat{\pi}_1} = \sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}} = \sqrt{\frac{0.36(0.64)}{826} + \frac{0.44(0.56)}{674}} = 0.0254$$

Example 3

So:

$$(\hat{\pi}_2 - \hat{\pi}_1) \pm z_{\alpha/2}(se_{\hat{\pi}_2 - \hat{\pi}_1})$$

$$(0.44 - 0.36) \pm 1.96(0.0254)$$

$$0.08 \pm 1.96(0.0254) = (0.030, 0.130)$$

The confidence interval does not contain zero so we can conclude there is a statistically significant difference in support for Donald Trump between men and women.

Example 3: using Stata prtesti

Can use the t-test calculator in Stata (for proportion) to obtain these results. Syntax below or use the drop-down menus. n, p refer to the two group sample sizes and proportions.

`prtesti` *`n1 p1 n2 p2`*

```
. prtesti 674 0.44 826 0.36

Two-sample test of proportions                    x: Number of obs =      674
                                                  y: Number of obs =      826
------------------------------------------------------------------------------
             |       Mean   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           x |        .44   .0191201                       .4025253    .4774747
           y |        .36   .0167013                        .327266     .392734
-------------+----------------------------------------------------------------
        diff |        .08   .0253873                       .0302419    .1297581
             |  under Ho:   .0253853     3.15   0.002
------------------------------------------------------------------------------
        diff = prop(x) - prop(y)                                  z =   3.1514
    Ho: diff = 0

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(Z < z) = 0.9992         Pr(|Z| > |z|) = 0.0016          Pr(Z > z) = 0.0008
```


Example 4: Using Stata

Use NELS to test the hypothesis that male and female high school students differ in their propensity to binge drink. The variable *alcbinge* is equal to 1 if the student has ever binged on alcohol, and equal to 0 otherwise. Use α=0.05. Let group 2 be males and group 1 be females.

$$H_0 : \pi_2 - \pi_1 = 0$$

$$H_1 : \pi_2 - \pi_1 \neq 0$$

Example 4: Using Stata

In Stata use `ttest` *varname*, `by(`*groupvar*`)`. (Do not use the `unequal` option in this case).

```
. ttest alcbinge, by(gender)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    male |     227    .2907489    .0302068    .4551114    .2312259    .3502719
  female |     273    .1538462    .0218768    .3614638    .1107768    .1969155
---------+--------------------------------------------------------------------
combined |     500        .216    .0184219    .4119264    .1798059    .2521941
---------+--------------------------------------------------------------------
    diff |            .1369027    .0365263                .0651382    .2086673
------------------------------------------------------------------------------
    diff = mean(male) - mean(female)                              t =   3.7481
Ho: diff = 0                                     degrees of freedom =      498

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


Example 4: Using Stata

From the above Stata output:

$$\hat{\pi}_2 - \hat{\pi}_1 = 0.1369$$

$$se_{\hat{\pi}_2 - \hat{\pi}_1} = 0.0365$$

$$t = 3.748$$

$$p = 0.0002$$

Conclusion: Reject H_0, since $p < \alpha$. There is a statistically significant difference in the propensity to binge drink between males and females.

Paired sample t-test

- Suppose instead our samples are *dependent*, or paired.
- One advantage of a paired design is the ability to control for some external differences between the two samples.
- Independent samples will differ for a lot of idiosyncratic reasons ("noise")
  - Paired samples allow you to "difference out" some of the noise that produces differences in independent sample means
  - One example: the *pre-post* design. Comparing the same individuals before and after some intervention eliminates many of the external differences that cause samples to differ

Paired sample t-test

With paired samples, we can conduct a test for the the average *within pair* difference, rather than the difference in two sample means. Here a null hypothesis of a zero difference in means is stated as:

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d \neq 0$$

where μ_d is the mean within-pair difference, or mean "difference score." A test of this hypothesis is simply the one-sample t-test for $\bar{x}_d$, the sample mean within-pair difference (calculated using each matched pair).

Paired sample t-test

The t-statistic for the paired sample t-test is:

$$t = \frac{\bar{x}_d - 0}{se_{\bar{x}_d}}$$

with $n-1$ degrees of freedom, where n is the number of matched pairs. The standard error of the mean difference is calculated as:

$$se_{\bar{x}_d} = s_{\bar{x}_d}/\sqrt{n}$$

$s_{\bar{x}_d}$ is the standard deviation of the paired differences.

Example 5

In Everitt (1994), 17 girls treated for anorexia were weighed before and after treatment. Difference scores were calculated for each participant, with the following results: $\bar{x}_d = 7.26$, $s_{\bar{x}_d} = 7.16$. Test the null hypothesis that there was no change in weight.

$$H_0 : \mu_d = 0$$
$$H_1 : \mu_d \neq 0$$

Example 5

The test statistic is:

$$t = \frac{7.26 - 0}{7.16/\sqrt{17}} = 4.17$$

With df=16 the probability of obtaining by chance a t-statistic of 4.17 or larger is $p < 0.01$. With a significance level of $\alpha = 0.05$ we reject H_0. The change in weight was statistically significant.

Example 5 - confidence interval approach

Assuming a significance level of $\alpha = 0.05$ we can alternatively construct a 95% confidence interval as:

$$\bar{x}_d \pm t(df)_{\alpha/2}(se_{\bar{x}_d})$$

$$\bar{x}_d \pm t(16)_{0.025}(s_{\bar{x}_d}/\sqrt{n})$$

$$7.26 \pm 2.12\left(\frac{7.16}{\sqrt{17}}\right)$$

(3.57, 10.95) - this interval does not contain zero, so we reject H_0. (The change in weight was statistically significant, or significantly different from zero).

Example 6: Using Stata

Using the Agresti & Finlay dataset anorexia.dta, test for an effect of cognitive behavioral therapy on the weight of anorexia patients. $n = 29$ subjects had `therapy==b`; *before* is the subject's weight before therapy and *after* is the subject's weight after therapy.

```
. ttest before=after if therapy=="b"

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  before |      29    82.68966    .8997857    4.845494    80.84653    84.53278
   after |      29    85.69655    1.550913    8.351924    82.51965    88.87345
---------+--------------------------------------------------------------------
    diff |      29   -3.006896    1.357155    7.308504   -5.786902   -.2268896
------------------------------------------------------------------------------
     mean(diff) = mean(before - after)                            t =  -2.2156
 Ho: mean(diff) = 0                              degrees of freedom =       28

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0175         Pr(|T| > |t|) = 0.0350          Pr(T > t) = 0.9825
```


Example 6: Using Stata

Notice the change in syntax in the `ttest` command. From the above Stata output:

$$\bar{x}_d = -3.001$$

$$se_{\bar{x}_d} = 1.357$$

$$t = -2.216$$

$$p = 0.035$$

Conclusion: Reject H_0, since $p < \alpha$. The CBT therapy had a statistically significant effect on the subjects' weight.

Stata `ttest` syntax recap

Stata `ttest` syntax:

- Hypothesis test about a single population mean: `ttest` *varname* `==[#]`, where `[#]` is the population mean under H_0.
- Hypothesis test comparing two population means: `ttest` *varname*`, by(`*group*`)`, where *group* is the group variable. The null hypothesis H_0 is that the means are equal.
- Paired sample hypothesis test: `ttest` *varname1==varname2*, where the null hypothesis H_0 is that the means are equal.
- Using Stata as a t-test calculator, for a single population mean: `ttesti` *#obs #mean #sd #val* `[, level(#)]`

Power calculation in Stata: two sample test

In Stata: Power analysis for a two-sample means test (independent samples). H_0 : no difference. You select:

- Effect size
- Equal variances or group-specific variances
- Sample size and "allocation ratio" n_2/n_1
- Significance level (α)
- 2- or 1-sided test

Effect size

How should we express an *effect size* for a difference in means, in order to assess practical significance? Can use a Cohen's d type measure, expressing the estimated difference in means as a proportion of the overall standard deviation:

$$d = \frac{\bar{x}_2 - \bar{x}_1}{s}$$

Practical vs. statistical significance, revisited

	A	B	C	D
Sample size	10,000	10,000	9	1,000
Overall mean test score	200	200	200	200
Standard deviation	25	25	100	25
Girls' mean test score	175	199	175	199
Boys' mean test score	225	201	225	201
Δ = Difference (Boys - Girls)	**50**	**2**	**50**	**2**
Effect size (Δ/SD)	2	0.08	0.5	0.08
Practically significant?	Yes	No	Yes (if true)	No
Standard error (se) of difference in means	0.5	0.5	66.6	1.58
t-statistic for difference in means (Δ/se)	100	4	0.75	1.26
p-value	p<0.0001	p<0.001	p>0.40	p>0.20
Statistically significant?	Yes	Yes	No	No
Confidence interval	$50 \pm 1.96 * 0.5$ (49.02, 50.98)	$2 \pm 1.96 * 0.5$ (1.02, 2.98)	$50 \pm 2.31 * 66.6$ (-103.8, 203.85)	$2 \pm 1.96 * 1.58$ (-1.10, 5.10)

Source: Remler & Van Ryzin ch. 8. Note standard error for difference in means is $2 * (SD/\sqrt{n})$ where $n/2$ is the number in each group. Assumes the standard deviation is the same for boys and girls, and an equal number of boys and girls.