Two Sample t Test

Nathaniel E. Helwig

Associate Professor of Psychology and Statistics
University of Minnesota

October 17, 2020

Table of Contents

Table of Contents

Paired Data Situation

Suppose that we have a random sample of bivariate data $(x_i, y_i) \overset{\text{iid}}{\sim} F$

- variable measured from subject before and after some intervention
- variable recorded from two people who are paired (e.g., twins)

We want to test a null hypothesis about mean differences:

- $H_0 : \mu_x - \mu_y = \mu_0$ versus $H_1 : \mu_x - \mu_y \neq \mu_0$ (two-sided H_1)
- $H_0 : \mu_x - \mu_y \geq \mu_0$ versus $H_1 : \mu_x - \mu_y < \mu_0$ (less than H_1)
- $H_0 : \mu_x - \mu_y \leq \mu_0$ versus $H_1 : \mu_x - \mu_y > \mu_0$ (greater than H_1)

where μ_0 is the null hypothesized mean difference (typically $\mu_0 = 0$).

Testing Hypotheses about Difference Scores

Define the difference score such as $z_i = x_i - y_i$ for $i = 1, \ldots, n$.

We could rewrite the null hypotheses as a one sample t test (Student, 1908) on the difference score:

- $H_0 : \mu_z = \mu_0$ versus $H_1 : \mu_z \neq \mu_0$ (two-sided H_1)
- $H_0 : \mu_z \geq \mu_0$ versus $H_1 : \mu_z < \mu_0$ (less than H_1)
- $H_0 : \mu_z \leq \mu_0$ versus $H_1 : \mu_z > \mu_0$ (greater than H_1)

where $\mu_z = E(z_i)$ is the expected value of the difference scores.

Paired samples t test simply involves conducting a one sample t test on the difference score.

Example 1: Psychiatric Patients (data)

Suppose that $n = 9$ psychiatric patients were treated with a tranquilizer drug, which was meant to reduce their suicidal tendencies.

Let X and Y denote the suicidal tendencies of the patients (as measured by the Hamilton Depression Scale) before and after the tranquilizer treatment.

Consequently, we will test the null hypothesis $H_0 : \mu_z \leq 0$ versus the alternative hypothesis $H_1 : \mu_z > 0$, where $Z = X - Y$ is the difference score (before minus after).

Example 1: Psychiatric Patients (parametric)

```
> # data
> pre <- c(1.83, 0.50, 1.62, 2.48, 1.68, 1.88, 1.55, 3.06, 1.30)
> post <-  c(0.878, 0.647, 0.598, 2.050, 1.060, 1.290, 1.060, 3.140, 1.290)

> # paired samples t test with greater than alternative
> t.test(pre, post, alternative = "greater", paired = TRUE)

Paired t-test

data:  pre and post
t = 3.0354, df = 8, p-value = 0.008088
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.1673028       Inf
sample estimates:
mean of the differences
              0.4318889
```

Example 1: Psychiatric Patients (nonparametric)

We could conduct a nonparametric version of this test using the `np.loc.test` function in the **nptest** R package (Helwig, 2020):

```
> library(nptest)
> np.loc.test(pre, post, alternative = "greater", paired = TRUE)

Nonparametric Location Test (Paired t-test)
alternative hypothesis: true difference of means is greater than 0
t = 3.0354,  p-value = 0.0137
sample estimate:
mean of the differences
              0.4318889
```

Note that this is an exact test because there are only $2^9 = 512$ elements of the permutation distribution.

Example 2: Twin Test Scores (data)

Suppose that we have collected psychological test scores from $n = 13$ pairs of dizygotic twins.

Want to test the null hypothesis that there is no difference between the first born twin's scores (X) and the second born twin's scores (Y).

In other words, we will test the exact null hypothesis $H_0 : \mu_z = 0$ versus the two-sided alternative hypothesis $H_1 : \mu_z \neq 0$, where $Z = X - Y$ is the difference score (twin 1 minus twin 2).

Example 2: Twin Test Scores (parametric)

```
> # data
> x <- c(277, 169, 157, 139, 108, 213, 232, 229, 114, 232, 161, 149, 128)
> y <- c(256, 118, 137, 144, 146, 221, 184, 188, 97, 231, 114, 187, 230)

> # paired samples t test with two-sided alternative
> t.test(x, y, paired = TRUE)

Paired t-test

data:  x and y
t = 0.34787, df = 12, p-value = 0.734
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -22.26777  30.72931
sample estimates:
mean of the differences
               4.230769
```

Example 2: Twin Test Scores (nonparametric)

We could conduct a nonparametric version of this test using the `np.loc.test` function in the **nptest** R package.

```
> np.loc.test(x, y, paired = TRUE)

Nonparametric Location Test (Paired t-test)
alternative hypothesis: true difference of means is not equal to 0
t = 0.3479,  p-value = 0.7415
sample estimate:
mean of the differences
               4.230769
```

Note that this is an exact test because there are only $2^{13} = 8192$ elements of the permutation distribution.

Table of Contents

Independent Data Situation

$x_i \overset{\text{iid}}{\sim} N(\mu_x, \sigma_x^2)$ and $y_i \overset{\text{iid}}{\sim} N(\mu_y, \sigma_y^2)$, where the X and Y observations are assumed to be independent of one another.

- n_x and n_y denote the number of observations for each group
- no natural pairing between the X and Y observations

We want to test a null hypothesis about mean differences:

- $H_0 : \mu_x - \mu_y = \mu_0$ versus $H_1 : \mu_x - \mu_y \neq \mu_0$ (two-sided H_1)
- $H_0 : \mu_x - \mu_y \geq \mu_0$ versus $H_1 : \mu_x - \mu_y < \mu_0$ (less than H_1)
- $H_0 : \mu_x - \mu_y \leq \mu_0$ versus $H_1 : \mu_x - \mu_y > \mu_0$ (greater than H_1)

where μ_0 is the null hypothesized mean difference (typically $\mu_0 = 0$).

Test Statistic Assuming Equal Variances

If the variances of the two populations are assumed to be equal, i.e., if $\sigma_x^2 = \sigma_y^2$, then the test statistic can be defined as

$$T_0 = \frac{\bar{x} - \bar{y} - \mu_0}{s_p \sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

where $\bar{x} = \frac{1}{n_x} \sum_{i=1}^{n_x} x_i$ and $\bar{y} = \frac{1}{n_y} \sum_{i=1}^{n_y} y_i$ are the sample means and

$$s_p^2 = \frac{(n_x - 1)s_x^2 + (n_y - 1)s_y^2}{n_x + n_y - 2}$$

is the *pooled variance estimate* with $s_x^2 = \frac{1}{n_x - 1} \sum_{i=1}^{n_x} (x_i - \bar{x})^2$ and $s_y^2 = \frac{1}{n_y - 1} \sum_{i=1}^{n_y} (y_i - \bar{y})^2$ denoting the variance for each group.

$T_0 \sim t_{n_x + n_y - 2}$ under the assumptions that H_0 is true and $\sigma_x^2 = \sigma_y^2$

Test Statistic Assuming Unequal Variances

When the variances are unequal, the previous test statistic's sampling distribution will *not* be a $t_{n_x+n_y-2}$ distribution even if H_0 is true.

Whether or not the type I error rate is too small versus too large will depend on the direction of the alternative hypothesis, the sample sizes n_x and n_y, and the ratio of the true variance σ_x^2/σ_y^2 (see Helwig, 2019).

Instead, we could define the test statistic as

$$T_0 = \frac{\bar{x} - \bar{y} - \mu_0}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}}$$

which is the test statistic that was proposed by Welch (1938, 1947).

Null Sampling Distribution for Unequal Variances

Assuming that the null hypothesis is true, the sampling distribution of Welch's t test statistic can be well approximated by Student's t distribution with the degrees of freedom parameter

$$\nu = \frac{\left(\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}\right)^2}{\frac{s_x^4}{n_x^2(n_x-1)} + \frac{s_y^4}{n_y^2(n_y-1)}}$$

which is known as the Welch-Satterthwaite approximation to the degrees of freedom (Satterthwaite, 1946).

Using this approach is the safer thing to do in practice, given that it can produce reasonable results even if the two variances are equal.

Example 3: Social Skills Training (data)

Suppose we are interested in assessing the effectiveness of a Social Skills Training (SST) program for alcoholics in a rehabilitation program.

Assume that $n_x = 12$ patients (the control group) were in the normal treatment program, and $n_y = 11$ patients (the test group) were in the SST supplement in addition to the normal treatment program.

Goal is to test the null hypothesis $H_0 : \mu_x - \mu_y \leq 0$ versus the alternative hypothesis $H_1 : \mu_x - \mu_y > 0$, where μ_x and μ_y denote the average amount of alcohol consumed in the first year after the program.

Example 3: Social Skills Training (parametric)

```
> # data
> x <- c(1042, 1617, 1180, 973, 1552, 1251, 1151, 1511, 728, 1079, 951, 1319)
> y <- c(874, 389, 612, 798, 1152, 893, 541, 741, 1064, 862, 213)

> # welch t test with greater than alternative
> t.test(x, y, alternative = "greater")

Welch Two Sample t-test

data:  x and y
t = 3.9747, df = 20.599, p-value = 0.0003559
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 258.5566      Inf
sample estimates:
mean of x mean of y
1196.1667  739.9091
```

Example 3: Social Skills Training (nonparametric)

We could conduct a nonparametric version of this test using the `np.loc.test` function in the **nptest** R package (Helwig, 2020):

```
> set.seed(0)
> np.loc.test(x, y, alternative = "greater")

Nonparametric Location Test (Welch Two Sample t-test)
alternative hypothesis: true difference of means is greater than 0
t = 3.9747,  p-value = 4e-04
sample estimate:
difference of the means
               456.2576
```

Note that this is a Monte Carlo test using $R = 9999$ resamples because there are $\binom{23}{11} = 1352078$ elements of the permutation distribution.

Table of Contents

Confidence Interval Definition

The parameter of interest is $\delta = \mu_x - \mu_y$, which denotes the true population mean difference.

To form a $100(1-\alpha)\%$ confidence interval for the mean difference $\delta = \mu_x - \mu_y$, we will use the Welch version of the t test. Note that

$$1 - \alpha = P\left(t_{\alpha/2} < \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} < t_{1-\alpha/2} \right)$$

where t_α denotes the α-th quantile of the t_ν distribution.

Confidence Interval Derivation

Rearranging the terms inside the probability statement reveals that

$$
\begin{aligned}
1-\alpha &= P\left(t_{\alpha/2}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}} < \bar{x}-\bar{y}-\delta < t_{1-\alpha/2}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}} \right) \\
&= P\left(t_{\alpha/2}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}} - (\bar{x}-\bar{y}) < -\delta < t_{1-\alpha/2}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}} - (\bar{x}-\bar{y}) \right) \\
&= P\left(\bar{x}-\bar{y}-t_{\alpha/2}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}} > \delta > \bar{x}-\bar{y}-t_{1-\alpha/2}\sqrt{\frac{s_x^2}{n_x}+\frac{s_y^2}{n_y}} \right)
\end{aligned}
$$

Confidence Interval in Practice

Previous result implies that the $100(1-\alpha)\%$ confidence interval is

$$\left[\bar{x} - \bar{y} - t_{1-\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}, \;\; \bar{x} - \bar{y} - t_{\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}\right]$$

Given that $-t_{\alpha/2} = t_{1-\alpha/2}$ the confidence interval can be written as

$$\bar{x} - \bar{y} \pm t_{1-\alpha/2}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}$$

Lower and Upper Confidence Bounds

Confidence intervals make sense to use if H_1 is two-sided.

For one-sided tests, it would make more sense to use a confidence bound, which puts the uncertainty in the direction relevant to H_1.

For $H_1 : \mu < \mu_0$, use an upper confidence bound:
$[-\infty, \quad \bar{x} - \bar{y} + t_{1-\alpha}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}]$

For $H_1 : \mu > \mu_0$, use a lower confidence bound:
$[\bar{x} - \bar{y} - t_{1-\alpha}\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}, \quad \infty]$

Example 3: Social Skills Training (revisited)

The `t.test` function automatically outputs the one-sided lower confidence bound. Note that we get this lower confidence bound (instead of an interval) because our alternative was greater than.

With 95% confidence, we conclude that (on average) the control group drinks at least 258.5566 more centiliters of alcohol than the SST group during the first year after the treatment program.

Reminder: this confidence bound assumes that the data are normally distributed for each group.

Example 3: Social Skills Training (bootstrap setup)

For a nonparametric confidence interval, would could use the `np.boot` function in test **nptest** R package (Helwig, 2020).

First, we need to setup the data in a data frame, and the we need to define the test statistic function.

```
> # setup data
> z <- c(x, y)
> g <- factor(rep(c("ctrl", "sst"), c(length(x), length(y))))
> data <- data.frame(alcohol = z, group = g)

> # define statistic function
> statfun <- function(x, data){
+     means <- with(data[x,], tapply(alcohol, group, mean))
+     means[1] - means[2]
+ }
```

Example 3: Social Skills Training (bootstrap results)

```
> # bootstrap data
> set.seed(0)
> np.boot(1:length(z), statfun, data)

Nonparametric Bootstrap of Univariate Statistic
using R = 9999 bootstrap replicates

  t0: 456.2576
  SE: 113.4205
Bias: 0.9594

BCa Confidence Intervals:
       lower    upper
90% 276.3364 648.3219
95% 243.5856 685.4309
99% 175.8862 756.5396
```

For a 95% lower confidence bound, we would use the lower limit of the 90% confidence interval, i.e., $[276.3364, \infty]$

References

Helwig, N. E. (2019). Statistical nonparametric mapping: Multivariate permutation tests for location, correlation, and regression problems in neuroimaging. *WIREs Computational Statistics 2*, e1457.

Helwig, N. E. (2020). *nptest: Nonparametric Tests*. R package version 1.0-2.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin 2*(6), 110–114.

Student (1908). The probable error of a mean. *Biometrika 6*(1), 1–25.

Welch, B. L. (1938). The significance of the difference between two means when the population variances are unequal. *Biometrika 39*(3/4), 350–362.

Welch, B. L. (1947). The generalization of 'Student's' problem when several different population variances are involved. *Biometrika 34*(1–2), 28–35.