The Two-Sample Independent Sample t Test

James H. Steiger

Department of Psychology and Human Development
Vanderbilt University

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The Two-Sample Independent Sample t Test

Introduction

- In previous lectures, we dealt with simplest kind of t statistic, the 1-sample t test for a single mean.
- This test statistic is actually just the simplest special case of an entire class of test statistics.
- In this lecture, we examine perhaps the best known of the t statistics, the 2-Sample Independent Sample t test for comparing the means of two groups.
- Some of the things we learned about the 1-sample t will generalize directly to this new situation.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Introduction

- The 2-sample t test is used to assess whether two different populations have the same mean.
- The most popular use for the test is in the context of the 2-group, Experimental-Control group design, in which a group of subjects is randomly divided into two groups, one of which receives the experimental treatment, the other a control (often some kind of placebo).
- Let μ_1 be the mean of the Experimental group, and μ_3 the mean of the control group. Then the statistical null hypothesis is

$$H_0: \quad \mu_1 = \mu_2$$

- Notice that this hypothesis is true, if and only if $\mu_1 - \mu_2 = 0$, so in a sense it is a hypothesis about the mean difference.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The 2-Sample Statistic

- We are interested in whether or not $\mu_1 - \mu_2$ is equal to zero. So we take independent samples of size n_1 and n_2, and compute sample means M_1 and M_2.
- We then examine $M_1 - M_2$ and see if it is "different enough from zero to be statistically significant."
- But how do we do *that*?

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The 2-Sample Statistic
A 2-Sample Z

- Although the equations are more complicated, the basic idea is the same.
- The numerator includes two sample statistics, but ultimately reduces them to a single number, $M_1 - M_2$.
- We *standardize* the mean difference $M_1 - M_2$ by subtracting its null-hypothesized mean, and dividing by its standard error.
- It turns out that the standard error of $M_1 - M_2$ is given by the following formula

$$\sigma_{M_1-M_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad (1)$$

- Notice that, in the formula, groups can come from populations with unequal variances, and the groups can have unequal sample sizes.
- The quantity in Equation 1 is called the *standard error of the difference between means*.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The 2-Sample Statistic
A 2-Sample Z

- If we knew the two population variances, we could construct a Z-statistic of the form

$$Z = \frac{M_1 - M_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{2}$$

- This Z-statistic would be the two-sample equivalent of the Z-statistic we saw earlier.
- But in practice we don't know the two population variances! So what can we do?

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The 2-Sample Statistic
Substituting Sample Variances

- We could simply substitute the two sample variances for their population counterparts in the above formula.
- This would give us the following Z-statistic.

$$Z = \frac{M_1 - M_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{3}$$

- This statistic would have a distribution that gets closer and closer to a standard normal distribution as the sample sizes get larger and larger.
- However, it would *not* have a t distribution at small samples.
- Is there some way we could modify the statistic so that it *would* have a t distribution? Read on . . .

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The 2-Sample Statistic
Getting to a t-Statistic

- Advanced statistical theory tells us that simply substituting sample variances in Equation 2 above will not be enough to get us to a t-distribution, unless $n_1 = n_2$.
- However, it turns out that making two simple modifications in the Z-statistic formula *will* result in a statistic that does have a t distribution.
- These two modifications occur in the denominator of the formula.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The 2-Sample Statistic
Getting to a t-Statistic

- First, let us incorporate an *assumption* of *equal variances*, that is, that both experimental populations have the same variance.
- If we are using an experimental-control design, this amounts to an assumption that the experimental effect acts additively, that is, scores in the experimental group have, in effect, a constant added to them. (Remember from the early days of the course that adding a constant to a group of scores does not change the variance of those scores.)

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The 2-Sample Statistic
Getting to a t-Statistic

- What would our original Z-statistic look like if we assumed equal variances?
- One way of approaching this is to simply drop the subscript on σ_1^2 and σ_2^2, since they are now assumed to be the same σ^2.
- We'd get

$$Z = \frac{M_1 - M_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \tag{4}$$

- But note, there is a common σ^2 that can be factored out, resulting in the equation below.

$$Z = \frac{M_1 - M_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2}} \tag{5}$$

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The 2-Sample Statistic
Getting to a t-Statistic

- Now, of course, we are no more likely to know the common σ^2 than we would be to know the individual group variances if they were not equal.

- We have to estimate them. So I simply substitute a symbol, $\hat{\sigma}^2$ indicating that an estimate of σ^2 will be used in place of the actual value.

$$Z = \frac{M_1 - M_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{\sigma}^2}} \tag{6}$$

- But *which* estimate should we use?

The 2-Sample Statistic
Getting to a t-Statistic

- There are many possible estimates of σ^2 that we could construct from the two sample variances.
- Remember, we are *assuming* each of the two sample variances is estimating the same quantity.
- It turns out that *one particular estimate*, if substituted for σ^2, yields a statistic that has an exact Student t distribution.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The 2-Sample Statistic
Getting to a t-Statistic

- This estimate goes by a variety of names.
- It is sometimes called "the pooled unbiased estimator," and also sometimes called "Mean Square Within" or "Mean Square Error." Gravetter and Walnau call it Mean Square Within.
- I'll give the specific formula for two groups

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \tag{7}$$

- Note that this can also be written

$$\hat{\sigma}^2 = \frac{SS_1 + SS_2}{df_1 + df_2} \tag{8}$$

where SS_1 stands for the sum of squared deviations inside group 1, and SS_2 stands for the sum of squared deviations inside group 2, and df_j is equal to $n_j - 1$ for each group.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The 2-Sample Statistic
Getting to a t-Statistic

- Note that $\hat{\sigma}^2$ is a weighted average of the two sample variances. Each variances is weighted by its degrees of freedom divided by the total degrees of freedom.
- This version of the formula is

$$\hat{\sigma}^2 = \left(\frac{df_1}{df_1 + df_2}\right) s_1^2 + \left(\frac{df_2}{df_1 + df_2}\right) s_2^2 \qquad (9)$$

- Notice that $\hat{\sigma}^2$ is a weighted average of the two variances in which the weights are positive and add up to 1. Consequently, it must be somewhere between the two s^2 values.
- What would the formula reduce to if the two sample sizes are equal, and consequently both df are the same? (C.P.)

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The 2-Sample Statistic
Getting to a t-Statistic

- That's right! If the two *df* are the same, then

$$\hat{\sigma}^2 = \frac{s_1^2 + s_2^2}{2} \qquad (10)$$

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The General Formula
The Equal-n Formula

Computing the 2-Sample t
The General Formula

- The equations we gave above yield a general formula that works whether or not the sample sizes are equal for the two groups. First compute

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

- Then plug your obtained value of $\hat{\sigma}^2$ into

$$t_{n_1+n_2-2} = \frac{M_1 - M_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{\sigma}^2}} \qquad (11)$$

- Note that the resulting t-statistic has $n_1 + n_2 - 2$ degrees of freedom.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The General Formula
The Equal-n Formula

Computing the 2-Sample t
An Example

Example (2-Sample t-Test)

A travel agent wants to examine the notion that people lose equal amounts of money in casinos in Reno than in Las Vegas.

She takes a sample of 20 clients who visited Las Vegas, and found that they lost an average of \$1435.65 with a standard deviation of \$265.14. On the other hand, 24 clients who visited Reno lost an average of \$1354.34 with a standard deviation of \$249.98.

Is this a 1-tailed or a 2-tailed test? Compute the t-statistic and test the null hypothesis with $\alpha = 0.05$.

(Continued on next slide ...)

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The General Formula
The Equal-n Formula

Computing the 2-Sample t
An Example

Example (2-Sample t-Test)

The null hypothesis is that $\mu_1 = \mu_2$, and the test is 2-tailed. We first compute $\hat{\sigma}^2$ as

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \\
&= \frac{(20 - 1)265.14^2 + (24 - 1)249.98^2}{24 + 20 - 2} \\
&= \frac{(19)70299.22 + (23)62490.00}{42} \\
&= 66022.7
\end{aligned}
$$

(Continued on next slide . . .)

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The General Formula
The Equal-n Formula

Computing the 2-Sample t
An Example

Example (2-Sample t-Test)

Next, we substitute the obtained value of 66022.7 for $\hat{\sigma}^2$ in the formula below.

$$
\begin{aligned}
t_{n_1+n_2-2} &= \frac{M_1 - M_2}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{\sigma}^2}} \\[2mm]
&= \frac{1435.65 - 1354.34}{\sqrt{\left(\frac{1}{20} + \frac{1}{24}\right)66022.7}} \\[2mm]
&= \frac{81.31}{\sqrt{\left(\frac{24+20}{24\times20}\right)66022.7}} \\[2mm]
&= \frac{81.31}{\sqrt{6052.081}} \\[2mm]
t_{42} &= 1.045181
\end{aligned}
$$

(Continued on next slide . . .)

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The General Formula
The Equal-n Formula

Computing the 2-Sample t
An Example

Example (2-Sample t-Test)

From our knowledge of normal distribution rejection points, and the fact that those for the t are always larger than the normal distribution, we know that this obtained value of the t-statistic is not significant.

The Z-statistic would have a critical value of 1.96, and the t will be somewhat higher. How much higher? We can compute it with R as

```
> qt(0.975,42)
[1] 2.018082
```

Introduction
The 2-Sample Statistic
Computing the Two-Sample *t*
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

The General Formula
The Equal-*n* Formula

Computing the 2-Sample *t*
The Equal-*n* Formula

- We have already seen that, when the sample sizes are equal, the formula for $\hat{\sigma}^2$ simplifies substantially.
- Suppose both groups have the same sample size, n. Then the t-statistic becomes

$$
\begin{aligned}
t_{2(n-1)} &= \frac{M_1 - M2}{\sqrt{\left(\frac{1}{n} + \frac{1}{n}\right)\hat{\sigma}^2}} \\
&= \frac{M_1 - M2}{\sqrt{\left(\frac{2}{n}\right)\frac{s_1^2 + s_2^2}{2}}} \\
&= \frac{M_1 - M2}{\sqrt{\frac{s_1^2 + s_2^2}{n}}} \quad\quad (12)
\end{aligned}
$$

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Building an R Routine

- Computing the t-statistic with two unequal sized samples is hard work.
- This is what computers are for.
- So let's construct an R function to compute the 2-sample t, and then we may never have to compute it "by hand" again!

Building an R Routine

- We begin by giving names to all the quantities required to compute the statistic and the critical values. They are
  `m.1,m.2,s.1,s.2,n.1,n.2,alpha,tails`
- We tell R with the following structural form that
  `t.2.sample` is a function to compute the two-sample t
  statistic. Note that we have used the syntax to establish
  the default α of 0.05 and default number of tails of 2.

```
> t.2.sample <- function(m.1,m.2,s.1,s.2,n.1,
+                        n.2,alpha=0.05,tails=2){
+ }
```

- All we need to do is put the code to calculate the t inside the braces, and properly return the output.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Building an R Routine

- First we add the code to compute $\hat{\sigma}^2$.

```
> t.2.sample <- function(m.1,m.2,s.1,s.2,n.1,
+                        n.2,alpha=0.05,tails=2){
+ # compute sigma.hat.squared
+   df <- n.1 + n.2 -2
+   sigma.hat.squared <-( (n.1-1)*s.1^2 +(n.2-1)*s.2^2
+
+ }
```

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Building an R Routine

- Next we add the code to compute the t.

```
> t.2.sample <- function(m.1,m.2,s.1,s.2,n.1,
+                        n.2,alpha=0.05,tails=2){
+ # compute sigma.hat.squared
+   df <- n.1 + n.2 -2
+   sigma.hat.squared <-( (n.1-1)*s.1^2 +(n.2-1)*s.2^2 )/df
+ # compute t
+   t <- (m.1 - m.2) / sqrt((1/n.1 + 1/n.2)*sigma.hat.squared)
+   return(t)
+ }
```

- Before continuing, we input the data from the example we looked at earlier. I call the function with the parameter names given explicitly, to help reduce the chance of an error. We get the same value.

```
> t.2.sample(m.1=1435.65,m.2=1354.34,s.1=265.14,s.2=249.98,
+            n.1=20,n.2=24,alpha=0.05,tails=2)

[1] 1.045181
```

James H. Steiger   The Two-Sample Independent Sample t Test

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Building an R Routine

- Next we add the code to compute the critical value from the t.

```
> t.2.sample <- function(m.1,m.2,s.1,s.2,n.1,
+                        n.2,alpha=0.05,tails=2){
+ # compute sigma.hat.squared
+   df <- n.1 + n.2 -2
+   sigma.hat.squared <-( (n.1-1)*s.1^2 +(n.2-1)*s.2^2 )/df
+ # compute t
+   t <- (m.1 - m.2) / sqrt((1/n.1 + 1/n.2)*sigma.hat.squared)
+ # compute critical value
+   if(tails == -1) p <- alpha
+   if(tails == 1) p <- 1-alpha
+   if(tails == 2) p <- c(alpha/2,1 - alpha/2)
+   crit <- qt(p,df)
+ # create a list of named quantities and return it
+   res <- list(t.statistic = t, df = df, alpha = alpha,
+               critical.t.values = crit)
+   return(res)
+ }
```

Introduction
The 2-Sample Statistic
Computing the Two-Sample *t*
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Building an R Routine

- Now, when we run the test problem, we get the full output.

```
> # test problem
> t.2.sample(m.1=1435.65,m.2=1354.34,s.1=265.14,s.2=249.98,
+            n.1=20,n.2=24,alpha=0.05,tails=2)
$t.statistic
[1] 1.045181

$df
[1] 42

$alpha
[1] 0.05

$critical.t.values
[1] -2.018082  2.018082
```

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Confidence Intervals on the Mean Difference

- Just as with the 1-Sample t, we may wish to construct a confidence interval on the quantity of interest.

- With the 1-Sample test, the "quantity of interest" was μ, the mean of the single population.

- In the 2-Sample situation, the quantity of interest is $\mu_1 - \mu_2$, the difference between the two population means. In an experimental-control group design, this mean difference represents the actual effect of the treatment.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Confidence Intervals on the Mean Difference

- The formula for the $1 - \alpha$ confidence interval is

$$M_1 - M_2 \pm t_{1-\alpha/2, n_1+n_2-2} \sqrt{(1/n_1 + 1/n_2)\hat{\sigma}^2} \qquad (13)$$

- Note that the left part of the formula is simply the numerator of the t statistic, and the right side of the formula is a critical value of t multiplied by the denominator of the t-statistic.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Introduction

- The *statistical assumptions* of the classic two-sample independent sample t are

  1. *Independence of observations.* Each observation is independent. The classic formula for the sampling variance of the sample mean, σ^2/n, is based on this assumption.
  2. *Normality.* The distribution of the populations is assumed to be normal.
  3. *Homogeneity of variances.* The populations are assumed to have equal variances.

- We need to consider, in turn,

  1. How violations of these assumptions affect performance of the t-test.
  2. What methods are available to produce reasonable inferential performance when assumptions are violated.
  3. How to detect violations of assumptions.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Independence

- If the n observations are independent, then M has a sampling variance of σ^2/n. Otherwise, the sampling variance may be quite different.
- Since most classic tests assume the formula σ^2/n is correct, they can be seriously in error if this assumption is violated.
- Exactly what the affect of the error is depends on the precise nature of the dependency.
- *If* the pattern of dependency is known, it may be possible to correct for it, using linear combination theory as taught in Psychology 310.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Normality

- A key fact about the normal distribution is that the sample mean and sample variance of a set of observations taken randomly from a normal population are independent.
- This independence of the mean and variance are crucial in the derivation of Student's t distribution.
- When populations are not normal, this lack of independence can lead to poor performance of the t-test.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Normality

- Violations of normality can occur in several distinct ways.
- The general shape of the distribution can be skewed, in some cases for obvious reasons related to the nature of the measurement process.
- There can be *contamination* by *outliers.* These extreme and unusual observations lead to the distribution having tails that are much longer than seen with a normal distribution.
- Yet, if the contamination probability is small, it may be difficult to diagnose outlier problems when they occur. For example, are the outliers the result of:
  1. A mixture of two or more processes (or subgroups) that characterize the population of interest?
  2. A random measurement error?

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Normality

- High skewness or kurtosis can lead to Type I error rates that are either much higher or much lower than the nominal rates.
- Contamination by outliers can lead to a significant loss of power when the null hypothesis is false.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Homogeneity of Variances

- As we saw earlier, the denominator of the 2-Sample t-statistic explicitly assumes equal variances.
- Recall that, with independent samples, the variance of $M_1 - M_2$ is

$$\text{Var}(M_1 - M_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \tag{14}$$

- The t statistic replaces this formula with one that assumes equal variances, i.e.,

$$\text{Var}(M_1 - M_2) = \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \sigma^2 \tag{15}$$

and then substitutes the estimate $\hat{\sigma}^2$ for σ^2, where

$$\hat{\sigma}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \tag{16}$$

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Homogeneity of Variances

- Notice that, in the preceding formula, we are *essentially* substituting the (weighted) average of the two variances for each variance in the formula for the sampling variance of $M_1 - M_2$.
- If the assumption of equal variances is correct, the resulting formula will be a consistent estimate of the correct quantity.
- What will the effect be if the assumption of equal variances is incorrect?
- How can we approximate the impact of a violation of the equal variances assumption on the true Type I error rate of the t-test when the null hypothesis is true?

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Homogeneity of Variances

- One simplified approach would be two assume that there is no sampling error in the sample variances, i.e., that $s_1^2 = \sigma_1^2$ and $s_2^2 = \sigma_2^2$, and measure the result of the violation of assumptions.
- For example, suppose $\sigma_1^2 = 40$, and $\sigma_2^2 = 10$, while $n_1 = 10$ and $n_2 = 20$. What will the approximate effect on the true α?

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Homogeneity of Variances

- Using our simplified assumption that the sample variances would perfectly estimate the population variances, let's compute the ratio of the obtained denominator to the correct denominator.
- First, let's compute $\hat{\sigma}^2$.

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{(10-1)40 + (20-1)(10)}{10 + 20 - 2} \\
&= \frac{360 + 190}{28} \\
&= 19.64286
\end{aligned}
$$

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Homogeneity of Variances

- The obtained denominator is then

$$
\begin{aligned}
\sqrt{\hat{\mathrm{Var}}(M_1 - M_2)} &= \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\hat{\sigma}^2} \\
&= \sqrt{\left(\frac{1}{10} + \frac{1}{20}\right)19.64286} \\
&= \sqrt{2.946429} \\
&= 1.716516
\end{aligned}
$$

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Homogeneity of Variances

- However, the correct denominator is

$$\sqrt{\frac{40}{10} + \frac{10}{20}} = \sqrt{4.5} = 2.12132 \qquad (17)$$

- The obtained denominator is considerably smaller than it should be. So the t statistic will, in general, be larger in value than it should be, and will therefore reject more often than it should.
- The critical value of the t statistic with 28 degrees of freedom is

```
> qt(.975,28)
[1] 2.048407
```

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Homogeneity of Variances

- Since obtained values of the t statistic are, in effect, expanded by the ratio $2.12132/1.71516$, the true α can be approximated as the area outside absolute t values of

```
> 1.71516/2.12132 * qt(.975,28)
[1] 1.656207
```

- This is

```
> 2*pt(1.71516/2.12132 * qt(.025,28),28 )
[1] 0.1088459
```

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Homogeneity of Variances

- The above estimate of .109 was obtained with a simplifying assumption, and is an approximation.
- An alternative approach is Monte Carlo simulation. I ran a t-test 10,000 times under the above conditions and the Type I error rate was .1155.
- This confirms that the true α is more than twice as large as the nominal α of .05.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Independence
Normality
Homogeneity of Variances

Effect of Violations
Homogeneity of Variances

- Using the approaches just demonstrated, we can verify the following general principles):
  1. If the two sample sizes are equal, the difference between nominal and true α will be minimal, unless sample sizes are really small and the variance ratio really large.
  2. If the two sample sizes are unequal, and the variances are inversely related to sample sizes, then the true α will substantially exceed the nominal α.
  3. If the two sample sizes are unequal, and the variances are directly related to sample sizes, then the true α will be substantially lower than the nominal α.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Non-Normality
Unequal Variances

Dealing with Non-Normality

- When data show a recognized non-normal distribution, one has recourse to several options:
    1. *Do nothing.* If violation of normality is not severe, the t-test may be reasonably robust.
    2. *Transform the data.* This seems especially justifiable if the data have a similar non-normal shape. With certain kinds of shapes, certain transformations will convert the distributions to be closer to normality. However, this approach is generally not recommended, for a variety of reasons.
    3. *Trim the data.* By trimming a percentage of the more extreme cases from the data, the skewness and kurtosis may be brought more into line with those of a normal distribution.
    4. *Use a non-parametric procedure.* Tests for equality of means that do not assume normality are available. However, they generally assume that the two samples have equal distributions, not that they simply have equal means (or medians).

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Non-Normality
Unequal Variances

Dealing with Non-Normality

- Although the "jury is still out" on these matters, a number of authors writing on robustness in social science statistical journals (e.g., Algina, Keselman, Lix, Wilcox) have promoted the use of trimmed means.
- In the preceding lecture module, we described a single sample test and confidence interval using a trimmed mean.
- One could examine the data and then choose a trimming proportion γ, but many authors recommend using a fixed value of $\gamma = 0.20$ to avoid the general problems connected with *post hoc* analysis.

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Non-Normality
Unequal Variances

Dealing with Unequal Variances
The Welch Test

- With unequal variances, a standard approach, recommended in many textbooks, is to employ the Welch test, which can be generalized to the analysis of variance setting.

- In the case of the two-sample t statistic, the Welch test employs a modified test statistic,

$$t' = \frac{M_1 - M_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{18}$$

and modified degrees of freedom

$$df' = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}} \tag{19}$$

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Non-Normality
Unequal Variances

Dealing with Unequal Variances
The Welch Test

- With unequal variances, a standard approach, recommended in many textbooks, is to employ the Welch test, which can be generalized to the analysis of variance setting.
- In the case of the two-sample t statistic, the Welch test employs a modified test statistic,

$$t' = \frac{M_1 - M_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \tag{18}$$

and modified degrees of freedom

$$df' = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{s_1^4}{n_1^2(n_1-1)} + \frac{s_2^4}{n_2^2(n_2-1)}} \tag{19}$$

Introduction
The 2-Sample Statistic
Computing the Two-Sample t
Building an R Routine
Confidence Intervals on the Mean Difference
Statistical Assumptions and Robustness
Dealing with Assumption Violations

Non-Normality
Unequal Variances

General Testing Strategy
The Welch Test

- How should one employ the Welch test?
- Some authors advocate a sequential strategy, in which one first tests for equal variances. If the equal variance test rejects, employ the Welch test, otherwise employ the standard t-test.
- This is, at its foundation, an "Accept-Support" strategy, in which one employs the standard test if the null hypothesis is not rejected.
- The fact that tests on variances have low power compromises this strategy.
- As a result, some authors advocate always doing a Welch test.