Two-sample t-tests

Recall the derived population of sample means.

Imagine you had two of these derived populations, as we had for variances when we developed the F-test.

Equality of Population Means

	Population 1	Population 2
Mean	μ_1	μ_2
Variance	σ_1^2	σ_2^2

Given two populations, test for equality. Actually test to see if the difference between two means is equal to some value, usually zero.

H_0: $\mu_1 - \mu_2 = \delta$ where often $\delta = 0$

Derived populations

From population 1 draw all possible samples of size n_1 to get $\bar{Y}_1$.

- Mean μ_1 yields $\mu_1 = \mu_{\bar{Y}_1}$
- Variance σ_1^2 yields $\sigma_1^2 = \sigma_{\bar{Y}_1}^2 = \sigma_1^2 / n_1$
- Derived population size is $N_1^{n_1}$

Likewise population 2 gives $\bar{Y}_2$.

- Mean μ_2 yields $\mu_2 = \mu_{\bar{Y}_2}$
- Variance σ_2^2 yields $\sigma_2^2 = \sigma_{\bar{Y}_2}^2 = \sigma_2^2 / n_2$
- Derived population size is $N_2^{n_2}$

Draw a sample from each population

	Population 1	Population 2
Sample size	n_1	n_2
d.f.	γ_1	γ_2
Sample mean	$\bar{Y}_1$	$\bar{Y}_2$
Sample variance	S_1^2	S_2^2

Take all possible differences $\left(\bar{Y}_1 - \bar{Y}_2\right) = \bar{d}$. That is, all possible means from the first population ($N_1^{n_1}$ means) and all possible means from the second population ($N_2^{n_2}$), and get the difference between each pair. There are a total of ($N_1^{n_1}$)($N_2^{n_2}$) values.

$\left(\bar{Y}_{1i} - \bar{Y}_{2j}\right) = \bar{d}_k$

for i = 1,..., $N_1^{n_1}$, j = 1,..., $N_2^{n_2}$ and k = 1,...,($N_1^{n_1}$)($N_2^{n_2}$)

The population of differences $\left(\bar{Y}_{1i} - \bar{Y}_{2j}\right) = \bar{d}_k$ has a mean equal to $\mu_{\bar{d}}$ or $\mu_{\bar{Y}_1-\bar{Y}_2}$, a variance equal to $\sigma^2_{\bar{d}}$ or $\sigma^2_{\bar{Y}_1-\bar{Y}}$ and a standard error equal to $\sigma_{\bar{d}}$ or $\sigma_{\bar{Y}_1-\bar{Y}}$ (the standard deviation of the mean difference).

Characteristics of the derived population

1) As n_1 and n_2 increase, the distribution of $\bar{d}$ approaches a normal distribution. If the two original populations are normal, the distribution of $\bar{d}$ is normal regardless of n_1 and n_2.

2) The mean of the differences ($\bar{d}_k$) is equal to the difference between the means of the two parent populations. $\mu_{\bar{d}} = \mu_{\bar{Y}_1} - \mu_{\bar{Y}_2} = \mu_{\bar{Y}_1-\bar{Y}_2} = \delta$

3) The variance of the differences ($\bar{d}_k$) is equal to the variance for the difference between the means of the two parent populations.

$$\sigma^2_{\bar{d}} = \sigma^2_{\bar{Y}_1-\bar{Y}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

$$\sigma_{\bar{d}} = \sigma_{\bar{Y}_1-\bar{Y}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

This is a variance for a linear combination. The two samples are assumed independent, so covariance considerations are not needed.

The variance comes from linear combinations. The variance of the sums is the sum of the variance (no covariance if independent).

Linear combination: $\bar{Y}_1 - \bar{Y}_2$

The coefficients are 1, –1 (i.e. $1\bar{Y}_1 + (-1)\bar{Y}_2$)

The variance for this linear combination is $1^2\hat{\sigma}^2_{\bar{Y}_1} + (-1)^2\hat{\sigma}^2_{\bar{Y}_2} = \frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}$ if the two variables are independent. Since they are each separately sampled at random this is an easy assumption.

The two-sample t-test

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

H_1: $\mu_1 - \mu_2 \neq \delta$,

a directional alternative (one tailed test) would specify a difference, either $>\delta$ or $< \delta$.

Commonly, δ is 0 (zero)

If H$_0$ is true, then

$$E(\bar{d}) = \mu_1 - \mu_2$$

$$\sigma^2_{\bar{d}} = \frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}$$

If values of σ^2_1 and σ^2_2 were KNOWN, we could use a Z-test, $Z = \frac{\bar{d} \text{ - } \delta}{\sigma_{\bar{d}}} = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2_1}{n_1} + \frac{\sigma^2_2}{n_2}}}$.

If values of σ^2_1 and σ^2_2 were NOT KNOWN, and had to be estimated from the samples, we would use a t-test, $t = \frac{\bar{d} \text{ - } \delta}{S_{\bar{d}}} = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S^2_1}{n_1} + \frac{S^2_2}{n_2}}}$.

Since the hypothesized difference is usually 0 (zero), the term ($\mu_1 - \mu_2$) is usually zero, and the equation is often simplified to $t = \frac{\bar{d}}{S_{\bar{d}}}$,

Et voila, a two sample t-test!

This is a very common test, and it is the basis for many calculations used in regression and analysis of variance (ANOVA). It will crop up repeatedly as we progress in the course. It is very important!

$t = \frac{\bar{d} \text{ - } \delta}{S_{\bar{d}}} = \frac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S^2_1}{n_1} + \frac{S^2_2}{n_2}}}$, often written just $t = \frac{\bar{d}}{S_{\bar{d}}}$ when δ or $\mu_1 - \mu_2$ is equal to zero.

The two-sample t-test

Unfortunately, this is not the end of the story. It turns out that there is some ambiguity about the degrees of freedom for the error variance, $\frac{S^2_1}{n_1} + \frac{S^2_2}{n_2}$. Is it n$_1$–1, or n$_2$–1, or somewhere in between, or maybe the sum?

Power considerations

POWER! We want the greatest possible power. It turns out that we get the greatest power (and our problems with degrees of freedom go away) if we can combine the two variance estimates into one, single, new and improved estimate! But we can only do this if the two variances are not different.

We can combine the two variance estimates into a single estimate if they are not too different. To determine if they are sufficiently similar we use an F test. Therefore, two-sample t-tests START WITH AN F TEST!

Pooling variances

If the two estimates of variance are sufficiently similar, as judged by the F test of variances (e.g. $\mathrm{H_0}\colon \sigma_1^2 = \sigma_2^2$), then they can be combined. This is called "pooling variances", and is done as a weighted mean (or weighted average) of the variances. The weights are the degrees of freedom.

Weighted means or averages

The usual mean is calculated as $\bar{Y} = \sum_{i=1}^{n} Y_i \Big/ n$. The weighted mean is $\bar{Y} = \sum_{i=1}^{n} w_i Y_i \Big/ \sum_{i=1}^{n} w_i$, or the sum of the variable multiplied by the weights divided by the sum of the weights.

Pooled variances are calculated as Pooled $S^2 = S_P^2 = \sum_{j=1}^{k} \gamma_i S_i^2 \Big/ \sum_{j=1}^{k} \gamma_i$ where *j* will be *j* = 1 and 2 for groups 1 and 2. There could be more than 2 variances averaged in other situations.

Recall that $\gamma_j S_j^2 = SS_j$, so we can also calculate the sum of the corrected *SS* for each variable divided by the sum of the d.f. for each variable $S_p^2 = \sum \gamma_j S_j^2 \Big/ \sum \gamma_j = \sum SS_j \Big/ \sum \gamma_j$

Pooled variance calculation $S_p^2 = \dfrac{\gamma_1 S_1^2 + \gamma_2 S_2^2}{\gamma_1 + \gamma_2} = \dfrac{SS_1 + SS_2}{\gamma_1 + \gamma_2} = \dfrac{SS_1 + SS_2}{(n_1 - 1) + (n_2 - 1)}$

Two sample t-test variance estimates

From linear combinations we know that the variance of the sum is the sum of the variances. This is the GENERAL CASE. $\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}$. But, if we test $\mathrm{H_0}\colon \sigma_1^2 = \sigma_2^2$ and fail to reject, we can pool the variances. The error variance is then $S_p^2\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)$. One additional minor simplification is possible. If $n_1 = n_2 = n$, then we can place the pooled variance over a single n, $2S_p^2 \big/ n$.

So we now have a single, more powerful, pooled variance! What are it's degrees of freedom?

The first variance had a d.f.= $n_1 - 1 = \gamma_1$

The second variance had d.f.= $n_2 - 1 = \gamma_2$

The pooled variance has a d.f equal to the sum of the d.f. for the variances that were pooled, so the degrees of freedom is S_p^2 is $(n_1-1)+(n_2-1) = \gamma_1 + \gamma_2$

Summary: case where $\sigma_1^2 = \sigma_2^2$

Test the variances to determine if they are significantly different. This is an F test of $\mathrm{H_0}\colon \sigma_1^2 = \sigma_2^2$.

If they are not different, then pool the variances into a single estimate of S_p^2.

The t-test is then done using this variance used to estimate the standard error of the difference.

The d.f. are (n$_1$–1) + (n$_2$–1)

The t-test equation is then $t = \dfrac{\bar{d} - \delta}{S_{\bar{d}}} = \dfrac{(\bar{Y}_1 - \bar{Y}_2) - (\mu_1 - \mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

One other detail; we are conducting the test with the condition that $\sigma_1^2 = \sigma_2^2$. This will be a new assumption for this test, equal variances.

Assumptions: NID r.v. (μ, σ^2)

N for Normality; the differences are normally distributed

I for Independence; the observations and samples are independent

Since the variance is specified to be a single variance equal to σ^2, then the variances are equal or the variance is said to be homogeneous. The compliment to homogeneous variance is heterogeneous variance.

Equal variance is also called homoscedasticity and the alternative referred to as heteroscedasticity. Samples characterized as having equal variance can also be referred to as homoscedastic or heteroscedastic.

Case where $\sigma_1^2 \neq \sigma_2^2$

How do we conduct the test if the variances are not equal? On the one had, this is not a problem. The linear combination we used to get the variance does not require homogeneous variance, so we know the variance is $\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}$.

But what are the degrees of freedom?

It turns out the d.f. are somewhere between the smaller of n$_1$–1 and n$_2$–1 and the d.f. for the pooled variance estimate [(n$_1$–1) + (n$_2$–1)].

It would be conservative to just use the smaller of n$_1$–1 and n$_2$–1. This works and is reasonable and it is done. However, power is lost with this solution. (This solution is suggested by your textbook).

The alternative is to estimate the d.f. using an approximation developed by Satterthwaite. This solution is used by SAS in the procedure PROC TTEST.

Satterthwaite's approximation d.f. $= \gamma \approx \dfrac{\left[\dfrac{S_1^2}{n_1} + \dfrac{S_2^2}{n_2}\right]^2}{\left[\dfrac{\left(S_1^2/n_1\right)^2}{n_1 - 1} + \dfrac{\left(S_2^2/n_2\right)^2}{n_2 - 1}\right]}$

This calculation is, of course, an approximation as the name suggests. Note that it does not usually give nice integer degrees of freedom, expect some decimal places. This is not an issue for computer programs that can get P-values for any d.f. It does complicate using our tables a little.

There is one additional "simplification". We know that the d.f. are at least the smaller of n_1–1 and n_2–1. But what if $n_1 = n_2 = n$? In this case the d.f. will be at least n–1. However, Satterthwaite's approximation will still, usually, yield a larger d.f.

Summary

There are two cases in two-sample t-tests. The case where $\sigma_1^2 = \sigma_2^2$ and the case where $\sigma_1^2 \neq \sigma_2^2$.

There are also some considerations for the cases where $n_1 = n_2$ and where $n_1 \neq n_2$.

Each of these cases alters the calculation of the standard error of the difference being tested and the degrees of freedom.

Variance	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
$n_1 \neq n_2$	$S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$	$\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}$
$n_1 = n_2 = n$	$2S_p^2 / n$	$\frac{S_1^2+S_2^2}{n}$

d.f.	$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$
$n_1 \neq n_2$	$(n_1 - 1) + (n_2 - 1)$	$\geq \min[(n_1 - 1), (n_2 - 1)]$
$n_1 = n_2 = n$	$2n - 2$	$\geq n - 1$

For our purposes, we will generally use SAS to conduct two-sample t-tests, and will let SAS determine Satterthwaite's approximation when the variances are not equal?

How does SAS know if the variances are equal? How does it know what value of α you want to use? Good questions. Actually, SAS does not know or assume anything. We'll find out what it does later.

One last thought on testing for differences between two populations. The test we have been primarily discussing is the *t* test, a test of equality of means. However, if we find in the process of checking variance that the variances differ, then there are already some differences between the two populations that may be of interest.

Numerical example

Compare the ovarian weight of 14 fish, 7 randomly assigned to receive injections of gonadotropin (treatment group) and 7 assigned to receive a saline solution injection (control group). Both groups are treated identically except for the gonadotropin treatment. Ovarian weights are to be compared for equality one week after treatment. During the experiment two fish were lost due to causes not related to the treatment, so the experiment became unbalanced.

Raw data

Obs	Treatment	Control
1	134	70
2	146	85
3	104	94
4	119	83
5	124	97
6	*	77
7	*	80

Summary statistics

Statistic	Treatment	Control
n	5	7
ΣY_i	627	586
ΣY_i^2	79,625	49,588
$\bar{Y}$	125.4	83.7
SS	999	531
γ	4	6
S^2	249.8	88.6

Research question: Does the gonadotropin treatment affect the ovarian weight? (Note: this implies a non-directional alternative). First, which of the various situations for two-sample t-tests do we have? Obviously, $n_1 \neq n_2$. Now check the variances.

1) H_0: $\sigma_1^2 = \sigma_2^2$

2) H_1: $\sigma_1^2 \neq \sigma_2^2$

3) Assume Y_i ~ NIDrv, representing the usual assumptions of normality and independence.

4) $\alpha = 0.05$ and the critical value for 4, 6 d.f. is $F_{\alpha/2,4,6} = 6.23$.

5) We have the samples, and know that the variances are 249.8 and 88.6, and the d.f. are 4 and 6 respectively. The calculated value is (given that we have a nondirectional alternative and arbitrarily placing the largest variance in the numerator), F = 249.8/88.6 = 2.82 with 4, 6 d.f.

6) The critical value is larger than the calculated value. We therefore fail to reject the null hypothesis.

7) We can conclude that the two samples have sufficiently similar variances for pooling.

Pooling the variances.

Recall, $S_p^2 = \dfrac{\gamma_1 S_1^2 + \gamma_2 S_2^2}{\gamma_1 + \gamma_2} = \dfrac{SS_1 + SS_2}{\gamma_1 + \gamma_2}$

$$S_p^2 = \frac{4(249.8)\ +\ 6(88.6)}{4+6} = \frac{999+531}{4+6} = \frac{1530}{10} = 153 \text{ with 10 d.f.}$$

Now calculate the standard error for the test, $S_{\bar{d}}$, using the pooled variance.

For this case

$$S_{\bar{d}} = S_{\bar{Y}_1-\bar{Y}_2} = \sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{153\left(\frac{1}{5}+\frac{1}{7}\right)} = \sqrt{153(0.343)} = \sqrt{52.457} = 7.24 \text{, with 10 df}$$

Completing the two-sample t-test.

1) $\mathrm{H_0}$: $\mu_1 - \mu_2 = \delta$. In this case we could state the null as $\mathrm{H_0}$: $\mu_1 = \mu_2$ since $\delta = 0$.

2) $\mathrm{H_0}$: $\mu_1 - \mu_2 \neq \delta$ or $\mathrm{H_0}$: $\mu_1 \neq \mu_2$

3) Assume d_i ~ NIDr.v. (δ, σ_δ^2). NOTE we have pooled the variances, so obviously we have assumed that all variance is homogeneous and equal to σ_δ^2.

4) α = 0.05 and the critical value is 2.228 (a nondirectional alternative for α=0.05 and 10 df)

5) We have the samples and know that the means are 125.4 and 83.7. The calculated *t* value is:

$$t = \frac{(\bar{Y}_1-\bar{Y}_2)-(\mu_1-\mu_2)}{\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{(\bar{Y}_1-\bar{Y}_2)-0}{S_{\bar{d}}} = \frac{\bar{Y}_1-\bar{Y}_2}{S_{\bar{d}}} = \frac{125.4-83.7}{7.24} = \frac{41.7}{7.24} = 5.76 \text{ with 10 d.f.}$$

6) The calculated value (5.76) clearly exceeds the critical value (2.228) value, so we would reject the null hypothesis.

7) Conclude that the gonadotropin treatment does affect the gonad weight of the fish. We can further state that the treatment increases the weight of gonads.

How about a confidence interval? Could we use a confidence interval here? You betcha!

Confidence interval for the difference between means

The general formula for a two-tailed confidence interval for normally distributed parameters is: "*Some parameter estimate* ± $t_{\alpha/2}$ ** standard error*"

The difference between the means ($\delta = (\mu_1 - \mu_2)$) is another parameter for which we may wish to calculate a confidence interval. For the estimate of the difference between μ_1 and μ_2 we have already determined that for α=0.05 we have $t_{\alpha/2}$ = 2.228 with 10 d.f.. We also found the estimate of the difference $\left(\bar{d} = (\bar{Y}_1 - \bar{Y}_2)\right)$ is 41.7 and the std error of the difference, $\left(S_{\bar{d}} = S_{\bar{Y}_1-\bar{Y}_2}\right)$, is 7.24.

The confidence interval is then $\bar{d} \pm t_{\alpha/2} S_{\bar{Y}}$ or 41.7± 2.228(7.24) and 41.7 ± 16.13. The probability statement is

$$P(\bar{d} - t_{\alpha/2} S_{\bar{d}} \le \mu_1 - \mu_2 \le \bar{d} + t_{\alpha/2} S_{\bar{d}}) = 1 - \alpha$$

$$P(25.57 \le \mu_1 - \mu_2 \le 57.83) = 0.95$$

Note that the interval does not contain zero. This observation is equivalent to doing a test of hypothesis against zero. Some statistical software calculates intervals instead of doing hypothesis tests. This works for hypothesis tests against zero and is advantageous if the hypothesized value of δ is something other than zero. When software automatically tests for differences it almost always test for differences from zero.

Summary

Testing for differences between two means can be done with the two-sample t-test or two sample Z test if variances are known.

For two independently sampled populations the variance will be $\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}$, the variance of a linear combination of the means.

The problem is the d.f. for this expression are not known.

Degrees of freedom are known if the variances can be pooled, so we start our two-sample t-test with an F-test.

Variances are pooled, if not significantly different, by calculating a weighted mean.

$$S_p^2=\frac{\gamma_1 S_1^2+\gamma_2 S_2^2}{\gamma_1+\gamma_2}=\frac{SS_1+SS_2}{\gamma_1+\gamma_2}=\frac{SS_1+SS_2}{(n_1-1)+(n_2-1)}$$

The error variance is given by $S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)$

The standard error is $\sqrt{S_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

If the variances cannot be pooled, the two-sample t-test can still be done, and degrees of freedom are approximated with Satterthwaite’s approximation.

Once the standard error is calculated, the test proceeds as any other t-test.

Confidence intervals can also be calculated in lieu of doing the t-test.

SAS example 4 – PROC TTEST

We would normally do two-sample t-tests with the SAS procedure called PROC TTEST. This procedure has the structure

```
proc ttest data = dataset name;
class group variable;
var variable of interest;
```

The PROC statement functions like any other proc statement.

The VARIABLE or VAR statement works the same as in other procedures we have seen.

The CLASS statement is new. It specifies the variable that will allow SAS to distinguish between observations from the two groups to be tested.

PROC TTEST Example 4a

Example from Steele & Torrie (1980) Table 5.2.

Corn silage was fed to sheep and steers. The objective was to determine if the percent digestibility differed for the two types of animals.

Example 1: Raw data

Obs	Sheep	Steers
1	57.8	64.2
2	56.2	58.7
3	61.9	63.1
4	54.4	62.5
5	53.6	59.8
6	56.4	59.2
7	53.2	

Unfortunately this data is not structured properly for PROC TTEST. It has two variables (sheep and steers) giving the percent digestibility for sheep and steers separately.

We need one variable with percent digestibility for both and a second variable specifying the type of animal.

This can be fixed in the data step.

In program note the following;

Change of the data structure from multivarite to univariate style.

The proc ttest statement

Note intermediate statistics, especially
the confidence intervals for both means and standard deviations

The test the hypothesis for both means and variances are discussed below.