Confidence Interval for $\mu_1 - \mu_2$

We can use the two-sample t statistic to find an approximate confidence interval for the difference in means.

We need to figure out the appropriate degrees of freedom, k, as for the two-sample t test.

We then compute a level C confidence interval for $\mu_1 - \mu_2$ as

$$(\bar{y}_1 - \bar{y}_2) \;\pm\; t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where t^* is the value for which the area under the $t(k)$ density curve between $-t^*$ and t^* is C.

The Pooled Two-Sample t Procedure

When comparing two sample means, we might sometimes we willing to assume that the variances of the two populations are the same.

If both variances are σ^2, then the variance of $\bar{y}_1 - \bar{y}_2$ is

$$\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2} = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)$$

If we don't know σ^2, we can substitute an estimate based on *both* samples:

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

This gives a more accurate estimator of σ^2 than s_1^2 or s_2^2 alone.

We can find P-values and confidence intervals based on a t distribution with $n_1 + n_2 - 2$ df. Unlike the previous two-sample procedure, this is exact if the data are normally-distributed.

When Should We Use the Pooled Variance Procedure?

We should use a pooled two-sample procedure only if we have good reason to think the variances are the same — even though we obviously aren't sure the means are the same!

When might we think this?

Perhaps the numbers vary mostly because of measurement error, and we used the same instrument to measure values in both samples.

Usually, though, it would be hard to be sure that the variances were (at least nearly) the same, so usually you should use the unpooled procedure.

Pooled Two-Sample t Test for the Calcium and Blood Pressure Example

Here are the results of doing a one-sided test of the null hypothesis that the mean change in B.P. is the same for the group given calcium and the group given a placebo, using the two-sample t test based on the pooled estimate of the variance:

```
Two sample T for change

treatmen     N      Mean     StDev   SE Mean
Calcium     10     -5.00      8.74       2.8
Placebo     11      0.64      5.87       1.8

95% CI for mu (Calcium) - mu (Placebo): ( -12.4,  1.1)
T-Test mu (Calcium) = mu (Placebo) (vs <):
   T = -1.75  P = 0.048  DF = 19
Both use Pooled StDev = 7.37
```

The P-value is now slightly smaller than before (0.048 vs. 0.053).

The Real Role of the Null Hypothesis

Sometime, we think a null hypothesis, such as $H_0 : \mu_1 = \mu_2$, might really be true:

> Doctors wonder whether a drug that has an effect when injected has any effect when taken orally, or whether it will be destroyed by acid in the stomach before being absorbed.

Other times, we don't seriously believe the null hypothesis could be true:

> Engineers have devised two quite different ways of assembling cars, and would like to know which one takes less time, on average.

It would be an incredible coincidence if the average times taken by the two methods were *exactly* equal.

We test for $\mu_1 = \mu_2$ anyway, as a way of seeing whether we have enough evidence to say that $\mu_1 < \mu_2$, or that $\mu_2 < \mu_1$.

Statistical Significance Versus Practical Importance

People often say a result is *statistically significant* if the P-value is small (say, < 0.05).

For a test of $H_0 : \mu_1 = \mu_2$ versus $H_a : \mu_1 \neq \mu_2$, we say a result is *practically important* if $\mu_1 - \mu_2$ is big enough to matter for whatever our purpose is.

Can a result be statistically significant but not practically important?

Can a result be practically important but not statistically significant?