

Comparison of two means

Readings: Ch 25, Ch 24 (pooled variance method only)

- Many studies compare two groups, often a treatment and a control.
- Here we consider the case where the outcome is a continuous measure.
- There are two cases of interest.
 1. The treatment and control samples are independent.
 2. The data consists of observations on pairs, with one in each of the treatment and control groups.
- We assume that the subjects are randomly selected from a larger population, or that the subjects are randomly separated into treatment and control groups.

Analysis of two independent samples: two-sample t -test - pooled variance method

- We will assume
 1. Each sample is a random selection from a normal population.
 2. The two populations have the **same variance**, σ^2 .
 3. The two samples are independent, so the values in one are not related to those in the other in any way.
- The data can be summarized using the samples size, sample mean and sample variance

Sample	1	2
size	n_1	n_2
mean	\bar{y}_1	\bar{y}_2
variance	s_1^2	s_2^2

- Example: To assess whether the level of iron in the blood is the same for children with cystic fibrosis as for healthy children, a random sample is selected from each population. The $n_1 = 9$ healthy children have average serum iron level $\bar{y}_1 = 18.9\mu\text{mol/l}$ and standard deviation $s_1 = 5.9\mu\text{mol/l}$. The $n_2 = 13$ children with cystic fibrosis have average iron level $\bar{y}_2 = 11.9\mu\text{mol/l}$ with sample standard deviation $s_2 = 6.3\mu\text{mol/l}$. Is there a true difference in population means?

Hypothesis Test

- In general H_0 is a statement of no difference.
- $H_0 : \mu_1 = \mu_2$ or $H_0 : \mu_1 - \mu_2 = 0$.
- In the example μ_1 would represent the mean iron level of population 1 and μ_2 would represent the mean of population 2.
- In general H_a represents the statement we wish to prove.
- Depending on the situation it can be one-sided or two-sided.
- In the example, the question is whether the two population means are different, so we use a two-sided H_a .
- $H_a : \mu_1 \neq \mu_2$ or $H_a : \mu_1 - \mu_2 \neq 0$.

- The test statistic is

$$T = \frac{\bar{y}_1 - \bar{y}_2}{SE_{pooled}(\bar{y}_1 - \bar{y}_2)}$$

where

$$SE_{pooled}(\bar{y}_1 - \bar{y}_2) = s_{pooled} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

and

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

- The example gives the following summary statistics.

Sample	Healthy(1)	CF(2)
size	$n_1 = 9$	$n_2 = 13$
mean	$\bar{y}_1 = 18.9$	$\bar{y}_2 = 11.9$
variance	$s_1^2 = 5.9^2$	$s_2^2 = 6.3^2$

- Note that the standard deviations are given rather than the variances (so I've squared in the table).
- These standard deviations are quite similar so the assumption of equal population variances is valid.
- The pooled variance estimate is

$$s_{pooled}^2 = \frac{(9 - 1)5.9^2 + (13 - 1)6.3^2}{9 + 13 - 2} = 37.738$$

so the pooled standard deviation estimate is $s_{pooled} = \sqrt{37.738} = 6.14$

- (Note that this is between s_1 and s_2 .)
- The standard error of the mean difference is

$$SE_{pooled}(\bar{y}_1 - \bar{y}_2) = 6.14 \sqrt{\frac{1}{9} + \frac{1}{13}} = 2.6638$$

- The test statistic is

$$T = \frac{18.9 - 11.9}{2.6638} = 2.628$$

- The degrees of freedom are in general $n_1 + n_2 - 2$, the denominator in the expression for s_{pooled}^2 .
- In the example, the degrees of freedom are $9 + 13 - 2 = 20$.

- The significance probability is $P = .016$ using the computer.
- Using tables, we compare our value to the entries in the row corresponding to 20 degrees of freedom to get bounds.
- Because 2.628 is between, 2.528 and 2.845, we find that the probability greater than 2.628 is between .005 and .01.
- With the two-sided alternative we must double this probability, and find that $.01 < P < .02$.
- We conclude that there is strong evidence against the null hypothesis of no difference in mean serum iron levels for healthy and CF patients.
- Our results are statistically significant at the $\alpha = .05$ level but not at the $\alpha = .01$ level.

Confidence interval

- The $100(1 - \alpha)\%$ confidence interval has the form

$$\bar{y}_1 - \bar{y}_2 \pm t_{n_1+n_2-2}^* SE_{pooled}(\bar{y}_1 - \bar{y}_2)$$

where $t_{n_1+n_2-2}^*$ is the value from the t distribution with $n_1 + n_2 - 2$ degrees of freedom which cuts off $\alpha/2$ probability in the right tail.

- In the example, $t_{20}^* = 2.086$ for $\alpha = .05$ and the 95% confidence interval is

$$\begin{aligned} 18.9 - 11.9 &\pm 2.086(2.6638) \\ 7 &\pm 5.5567 \\ (1.443 &, 12.557) \end{aligned}$$

- We are 95% confident that the mean difference in serum iron levels is between 1.443 and 12.557
- Because this interval does not contain zero, our results are statistically significant at the $\alpha = .05$ level.

Analysis of paired data: paired difference t -test

The general approach to comparison of means with paired data is to take the difference, and estimate/test the mean difference, using a paired t -test.

- Paired experiments are used to remove the effect of variation between individuals.
- We analyze the data by using the differences for each pair.
- Example: Ten patients were randomly selected to take part in a nutritional program designed to lower blood cholesterol in children. Two months following the commencement of the program, the pediatrician measured the blood cholesterol levels of the 10 patients again. The results are as follows:

Patient	Before	After	Difference
1	210	212	-2
2	217	210	7
3	208	210	-2
4	215	213	2
5	202	200	2
6	209	208	1
7	207	203	4
8	200	199	1
9	221	218	3
10	218	214	4

- Note that some patients had high readings both before and after the program (e.g. patients 9, 10), others had low readings (e.g. patients 5,6).
- The differences are mostly positive, suggesting the program has a beneficial effect.
- The null hypothesis is $H_0 : \mu_1 - \mu_2 = \mu_d = 0$
- $H_a : \mu_1 > \mu_2$ or $\mu_d > 0$. (the program is designed to lower blood cholesterol, so we use a one-sided H_a)
- The test statistic is

$$T = \frac{\bar{d} - 0}{SE(\bar{d})}$$

where

$$SE(\bar{d}) = s_{\bar{d}} = \frac{s_d}{\sqrt{n}} = \sqrt{\frac{s_d^2}{n}}$$

and n is the number of pairs.

- In the example $\bar{d} = 2$,

$$\begin{aligned}s_d^2 &= \frac{1}{n-1}(\sum d_i^2 - (\sum d_i)^2/n) \\ &= \frac{1}{9}(108 - 20^2/10) \\ &= 7.5556\end{aligned}$$

so $SE(\bar{d}) = \sqrt{7.5556/10} = .8692$.

- The value of the test statistic is therefore

$$T = \frac{2}{.8692} = 2.3009$$

- To evaluate the P value we recognize that T has a t distribution with $n - 1 = 9$ degrees of freedom, if the null hypothesis is correct.
- Note that we must also assume that the differences are a random sample from a normal population.
- Because our alternative hypothesis is one-sided, $P = P(T > 2.3009)$.
- Using the computer, (or an app?) we find $P = .0235$.
- Using tables, we compare our value (its absolute value if necessary) to the values given in the row for 9 df to obtain bounds on P .
- In this case 2.3009 is between 2.262 and 2.821, so $.01 < P < .025$.
- We conclude that there is strong evidence against the null hypothesis of no effect of the nutritional program.

Confidence interval

- The confidence interval has the form

$$(\bar{d} - t_{n-1}^* SE(\bar{d}), \bar{d} + t_{n-1}^* SE(\bar{d}))$$

where t_{n-1}^* is the upper $\alpha/2$ th quantile of the t distribution with $n - 1$ degrees of freedom.

- For a 95% = $(1 - \alpha)100\%$ confidence interval $\alpha = .05$ and t_{n-1}^* is close to 2 when n is large.
- The half-width of the interval is called the **margin of error**. In this case

$$MOE = t_{n-1}^* SE(\bar{d})$$

- In the example $t_9^* = 2.262$,

$$MOE = 2.262(.8692) = 1.9662$$

and the confidence interval is

$$2 \pm 1.9662 \\ (.0338, 3.9662)$$

- We are 95% confident that the true mean difference in cholesterol following the nutritional program is between .0338 and 3.9662.
- Recall the connection between confidence intervals and hypothesis tests: a 95% confidence interval contains all values which are not statistically significant at the $\alpha = .05$ level of significance for a two-sided alternative.
- In the example the interval does not contain zero, so we would have statistical significance ($P < .05$) for a two-sided alternative.
- We double the probability in one tail of the t distribution to get the P value for a two-sided alternative, so based on this confidence interval we can conclude $P < .025$ for the one-sided alternative.
 - this is consistent with our test results above