Comparison of two means

Most studies are **comparative** in that they compare outcomes from one group with outcomes from another, for example the mean blood pressure in reponse to two different treatments.

1 Paired t-test - Deveaux et al Chapter 25

- in the **matched-pairs** design each subject in one group is paired with a similar subject in the other group
- one treatment is randomly assigned to one member of the pair - the other treatment is given to the other
- example: to compare two treatments for a disease, pair subjects who are similarly affected, same sex, age, etc.
- in many cases, the two treatments are given to the same subject in random order - with a 'wash-out' period between.
- eg. population of patients with high blood pressure currently receiving no medication. Blood pressure is measured on a sample of patients at beginning of study, then they are put on a statin. After one year on the treatment, the blood pressure is measured again. Does the statin reduce blood pressure?
- the difference between the two measurements in a pair should only reflect the different
  treatments or experimental conditions
- Individuals act as their own controls, so that the between individual source of variability is removed. When comparing means, the usual approach is to take the difference of means. The direction of the difference is important. For example, if the difference is (after-before), the natural alternative hypothesis in the statin example is that the mean difference is less than 0.

- recall that the mean of the differences equals the difference of the means
- if we can assume the differences are normally distributed, they can be analyzed using the one-sample t test or confidence interval

Example: Suppose we are comparing costs of auto repairs at two locations. We get an estimate at both places for the 6 same cars that have recently been involved in collisions:

Car	Cost at garage 1	Cost at garage 2	Difference
1	760	730	30
2	1020	910	110
3	950	840	110
4	130	150	–20
5	300	270	30
6	630	580	50

Is there evidence that mean costs are different at the two locations?

- repair costs vary considerably between cars
- for each car, the first location tends to be more expensive than the second
- Hypotheses:

$$H_0: \ \mu_d = 0$$
$$H_a: \ \mu_d \neq 0$$

- for the column of differences,

$$n = 6, \quad df = 5, \quad \bar{y} = 51.667, \quad s = 50.761$$

- so test statistic is

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{51.667 - 0}{50.761/\sqrt{6}} = 2.493.$$

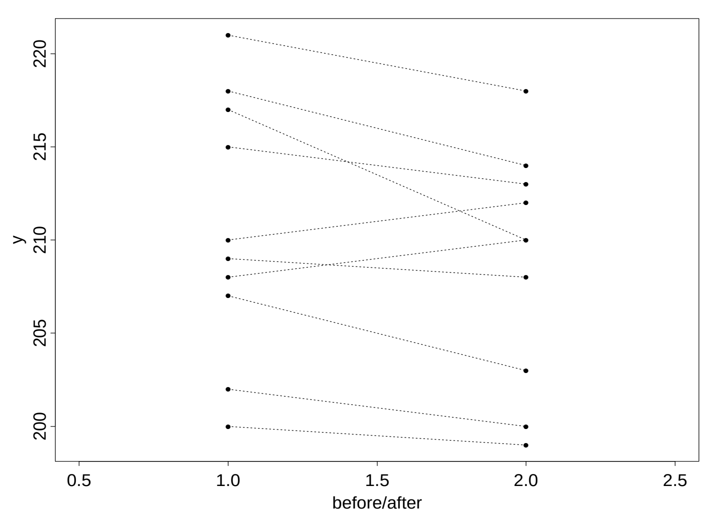
- 2.493 is between 2.015 and 2.571, so $P(T > 2.493)$ is between .025 and .05
- P-value is between 2(0.025) and 2(0.05), that is, between 0.05 and 0.10 (double because two-sided alternative)
- there is only weak evidence of a difference in costs

Example: Ten patients were randomly selected to take part in a nutritional program designed to lower blood cholesterol. 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

Construct a 95% confidence interval for the mean improvement in serum cholestrol.

- a plot of the data shows little apparent difference between before and after



- once the paired points are joined, however, it becomes clear that most values are lower after the nutritional program
- the mean and standard deviation of the differences are $\bar{y} = 2.0$ and $s = 2.749$
- with 9 df, the table value is $t^* = 2.262$
- the 95% confidence interval is

$$2.0 \pm 2.262(2.749)/\sqrt{10}$$

or

$$2.0 \pm 1.966$$

or

$$.034, 3.966$$

- because this interval does not include 0, the difference is significantly different from 0 at the $\alpha = .05$ level
- this may seem surprising considering the individual 95% confidence intervals,
  for before: (205.74, 215.66),
  for after: (204.24, 213.16)
- the paired analysis removes the variation due to the subject, so the difference has a small standard error

2 Comparison of means of two independent samples - pooled t procedure

(DVB Ch. 24, p 668-671)

- matching is not always possible
- however, can divide individuals at random into the two groups to be compared
  - give one group one treatment and the other group the other
- or can take random sample from each of two populations
- because of randomization, groups should be similar in all respects apart from treatment
- any differences are attributable to the treatment
- unlike in the matched pairs experiment, **the two samples are independent**

Notation

	Population		Sample		
Group	Mean	SD	Size	Mean	SD
1	μ_1	σ	n_1	$\bar{y}_1$	s_1
2	μ_2	σ	n_2	$\bar{y}_2$	s_2

- **important assumption**: the population standard deviations are the same in the two groups. (more on this later)
- call this common SD σ
- want to make confidence intervals for $\mu_1 - \mu_2$ or test hypotheses about $\mu_1 - \mu_2$
- idea: base inferences on $\bar{y}_1 - \bar{y}_2$
  - center for confidence interval
  - numerator of test statistic

Theory

- mean of $\bar{y}_1 - \bar{y}_2$ is $\mu_1 - \mu_2$
- SD of $\bar{y}_1 - \bar{y}_2$ is $\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$

- standardized difference is

$$z = \frac{(\bar{y}_1 - \bar{y}_2) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- problem: as before, don't know σ
- use *pooled sample variance*

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

- a weighted average of the two sample variances
- larger sample has larger weight
- now follow usual steps (but with slight changes)
- replace σ by s_p
- replace normal distribution by t distribution with $n_1 + n_2 - 2$ d.f.
- confidence interval for $\mu_1 - \mu_2$ is

$$(\bar{y}_1 - \bar{y}_2) \pm t^*_{n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

- test of $H_0 : \mu_1 = \mu_2$ uses

$$t = \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Example: Nine observations of surface-soil pH were made at each of two different locations.

Location 1	8.53	8.52	8.01	7.99	7.93
	7.89	7.85	7.82	7.80	
Location 2	7.85	7.73	7.58	7.40	7.35
	7.30	7.27	7.27	7.23	

Construct a 99% confidence interval for the difference in mean suface-soil pH at the two locations, using the following summaries.

	n	$\bar{y}$	s
Location 1	9	8.038	.285
Location 2	9	7.442	.224

Minitab

- the pooled two-sample confidence interval and test can be done in Minitab
- output follows for the pH example

```
MTB > set c1
DATA> 8.53 8.52 8.01 7.99 7.93  7.89 7.85 7.82 7.80
DATA> set c2
DATA> 7.85 7.73 7.58 7.40 7.35  7.30 7.27 7.27 7.23
DATA> set c3
DATA> (1)9
DATA> set c4
DATA> (2)9
DATA> mplot c1 c3 c2 c4

       -
       -
  8.50+    2
       -
       -
       -
       -
  8.00+    2
       -   2
       -   3                                            B
       -                                                B
       -                                                B
  7.50+
       -                                                B
       -                                                4
       -                                                B
       -
        ----+---------+---------+---------+---------+---------+--
          1.00      1.20      1.40      1.60      1.80      2.00
      A = C1 vs. C3             B = C2 vs. C4


MTB > twosample .99 c1 c2;
SUBC> pooled.

TWOSAMPLE T FOR C1 VS C2
    N      MEAN     STDEV   SE MEAN
C1  9     8.038     0.285     0.095
C2  9     7.442     0.224     0.075

99 PCT CI FOR MU C1 - MU C2: (0.242, 0.949)

TTEST MU C1 = MU C2 (VS NE): T= 4.92  P=0.0002  DF=16

POOLED STDEV =      0.257
```

- the plot shows
  - the values are higher at location 1
  - the spread of the values is nearly the same at the two locations
  - there is no strong evidence that the values are not from normal population
- note the subcommand ‘pooled” must be used
- the subcommand ‘alternative” can be used to specify a one-sided alternative

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 $\bar{y} = 18.9\mu$mol/l and standard deviation $s_1 = 5.9\mu$mol/l. The $n_2 = 13$ children with cystic fibrosis have average iron level $\bar{y} = 11.9\mu$mol/l with sample standard deviation $s_2 = 6.3\mu$mol/l. Is there a true difference in
population means?

- the hypotheses are, $H_0 : \mu_1 = \mu_2$ and
  $H_a : \mu_1 \neq \mu_2$
- pooling seems appropriate here

$$
\begin{aligned}
s_p^2 &= \frac{(9-1)5.9^2 + (13-1)6.3^2}{9+13-2} \\
&= \frac{754.76}{20} = 37.738
\end{aligned}
$$

  so $s_p = 6.1431$
- the test statistic is

$$
\begin{aligned}
t &= \frac{\bar{y}_1 - \bar{y}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{18.9 - 11.9}{6.1431(.4336)} \\
&= \frac{7}{2.6636} = 2.6280
\end{aligned}
$$

- now, with 20 degrees of freedom,
  $P(t > 2.528) = .01$ and
  $P(t > 2.845) = .005$),
  so $.005 < P(t > 2.628) < .01$
- doubling, because the alternative is two-sided, the P value is between .01 and .02
- there is strong evidence against the null hypothesis of no diffence in iron level.

Two independent samples vs.paired

- can be difficult to tell whether data should be treated as paired or not
- if the two samples are of different sizes, the data cannot be paired
- if the two samples are the same size, the data might be paired, but might not be
- to decide, read the description of the data
- key words for paired problem: paired, matched, before/after
- conclusions can be totally wrong, if wrong analysis is used
- typically, using a two-sample procedure when a paired procedure should be used leads to
  - a larger P value
  - a wider confidence interval because the pooled variance estimate is much larger than the variances of the differences