Comparing means

- Tests with one categorical and one numerical variable
- Goal: to compare the mean of a numerical variable for different groups.

Paired vs. 2 sample comparisons

Paired comparisons allow us to account for a lot of extraneous variation.

2-sample methods are sometimes easier to collect data.

Paired designs

- Data from the two groups are paired
- Each member of the pair shares much in common with the other, except for the tested categorical variable
- There is a one-to-one correspondence between the individuals in the two groups
Paired design: Examples

- Before and after treatment
- Upstream and downstream of a power plant
- Identical twins: one with a treatment and one without
- Earwigs in each ear: how to get them out? Compare tweezers to hot oil

Paired comparisons

- We have many pairs
- In each pair, there is one member that has one treatment and another who has another treatment
  
  (“Treatment” can mean “group”)

Paired comparisons

- To compare two groups, we use the mean of the difference between the two members of each pair

Paired $t$ test

- Compares the mean of the differences to a value given in the null hypothesis
- For each pair, calculate the difference. The paired $t$-test is simply a one-sample $t$-test on the differences.
Example: National No Smoking Day

- Data compares injuries at work on National No Smoking Day (in Britain) to the same day the week before.

- Each data point is a year.


Hypotheses

H₀: Work related injuries do not change during No Smoking Days. (µₜ = 0)

Hₐ: Work related injuries change during No Smoking Days. (µₜ ≠ 0)

Calculate differences

<table>
<thead>
<tr>
<th>Year</th>
<th>Injuries before No Smoking Day</th>
<th>Injuries on No Smoking Day</th>
<th>Difference (d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1987</td>
<td>516</td>
<td>540</td>
<td>24</td>
</tr>
<tr>
<td>1988</td>
<td>610</td>
<td>620</td>
<td>10</td>
</tr>
<tr>
<td>1989</td>
<td>581</td>
<td>599</td>
<td>18</td>
</tr>
<tr>
<td>1990</td>
<td>586</td>
<td>639</td>
<td>53</td>
</tr>
<tr>
<td>1991</td>
<td>554</td>
<td>607</td>
<td>53</td>
</tr>
<tr>
<td>1992</td>
<td>632</td>
<td>603</td>
<td>-29</td>
</tr>
<tr>
<td>1993</td>
<td>479</td>
<td>519</td>
<td>40</td>
</tr>
<tr>
<td>1994</td>
<td>583</td>
<td>560</td>
<td>-23</td>
</tr>
<tr>
<td>1995</td>
<td>445</td>
<td>515</td>
<td>70</td>
</tr>
<tr>
<td>1996</td>
<td>522</td>
<td>556</td>
<td>34</td>
</tr>
</tbody>
</table>
Calculate $t$ using $d$'s

$$
\bar{d} = 25 \\
s_d^2 = 1043.78 \\
n = 10
$$

$$
t = \frac{25 - 0}{\sqrt{1043.78/10}} = 2.45
$$

CAUTION!

- The number of data points in a paired $t$ test is the number of *pairs*. -- *Not* the number of individuals

- Degrees of freedom = Number of pairs - 1

Critical value of $t$

$$
t_{0.05(2),9} = 2.26
$$

$$
t = 2.45 > 2.26
$$

So we can reject the null hypothesis. Stopping smoking increases job-related accidents in the short term.

Assumptions of paired $t$ test

- Pairs are chosen at random

- The differences have a normal distribution

It does *not* assume that the individual values are normally distributed, only the differences.
Comparing the means of two groups

Hypothesis test: 2-sample $t$ test

Estimation: Difference between two means

$$\bar{Y}_1 - \bar{Y}_2$$

Confidence interval:

$$\left( \bar{Y}_1 - \bar{Y}_2 \right) \pm SE_{\bar{Y}_1 - \bar{Y}_2} t_{\alpha(2),df}$$

Standard error of difference in means

$$SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}$$

Pooled variance:

$$s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$$

$$df_1 = n_1 - 1; \ df_2 = n_2 - 1$$

Costs of resistance to aphids

2 genotypes of lettuce: Susceptible and Resistant

Do these genotypes differ in fitness in the absence of aphids?
Data, summarized

<table>
<thead>
<tr>
<th></th>
<th>Susceptible</th>
<th>Resistant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean number of buds</td>
<td>720</td>
<td>582</td>
</tr>
<tr>
<td>SD of number of buds</td>
<td>223.6</td>
<td>277.3</td>
</tr>
<tr>
<td>Sample size</td>
<td>15</td>
<td>16</td>
</tr>
</tbody>
</table>

Both distributions are approximately normal.

Calculating the standard error

\[ df_1 = 15 - 1 = 14; \quad df_2 = 16 - 1 = 15 \]

\[ s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2} = \frac{14(223.6)^2 + 15(277.3)^2}{14 + 15} = 63909.9 \]

\[ SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{63909.9}{15} + \frac{63909.9}{16}} = 90.86 \]

Finding \( t \)

\[ df = df_1 + df_2 = n_1 + n_2 - 2 \]

\[ = 15 + 16 - 2 \]

\[ = 29 \]

\[ t_{0.05(2), 29} = 2.05 \]

The 95% confidence interval of the difference in the means

\[ (\bar{Y}_1 - \bar{Y}_2) \pm SE_{\bar{Y}_1 - \bar{Y}_2} t_{\alpha(2), df} = (720 - 582) \pm 90.86(2.05) \]

\[ = 138 \pm 186 \]
Testing hypotheses about the difference in two means

2-sample t-test

The two sample t-test compares the means of a numerical variable between two populations.

Hypotheses

$H_0$: There is no difference between the number of buds in the susceptible and resistant plants. ($\mu_1 = \mu_2$)

$H_A$: The resistant and the susceptible plants differ in their mean number of buds. ($\mu_1 \neq \mu_2$)

2-sample t-test

$$t = \frac{\bar{Y}_1 - \bar{Y}_2}{SE_{\bar{Y}_1 - \bar{Y}_2}}$$

Calculating $t$

$$t = \frac{(\bar{Y}_1 - \bar{Y}_2)}{SE_{\bar{Y}_1 - \bar{Y}_2}} = \frac{(720 - 582)}{90.86} = 1.52$$
Drawing conclusions...

$t_{0.05(2),29} = 2.05$

$t < 2.05$, so we cannot reject the null hypothesis.

These data are not sufficient to say that there is a cost of resistance.

Assumptions of two-sample $t$ - tests

- Both samples are random samples.
- Both populations have normal distributions.
- The variance of both populations is equal.

The wrong way to make a comparison of two groups

"Group 1 is significantly different from a constant, but Group 2 is not. Therefore Group 1 and Group 2 are different from each other."

A more extreme case...

![Diagram showing differences in resemblance between mother and father groups.](image-url)
Comparing means when variances are not equal

Welch’s t test

*Welch’s approximate t-test* compares the means of two normally distributed populations that have unequal variances.

Dung beetles

Burrowing owls and dung traps

**Experimental design**

- 20 randomly chosen burrowing owl nests
- Randomly divided into two groups of 10 nests
- One group was given extra dung; the other not
- Measured the number of dung beetles on the owls’ diets
Number of beetles caught

• Dung added: \( \bar{Y} = 4.8 \)
  \( s = 3.26 \)

• No dung added: \( \bar{Y} = 0.51 \)
  \( s = 0.89 \)

Hypotheses

\( H_0: \) Owls catch the same number of dung beetles with or without extra dung \((\mu_1 = \mu_2)\)

\( H_A: \) Owls do not catch the same number of dung beetles with or without extra dung \((\mu_1 \neq \mu_2)\)

Welch’s \( t \)

\[
t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}
\]

\[
df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}
\]

Round down \( df \) to nearest integer

Owls and dung beetles

\[
t = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{4.8 - 0.51}{\sqrt{3.26^2/10 + 0.89^2/10}} = 4.01
\]
Degrees of freedom

\[ df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1^2} n_1 - 1\right) + \left(\frac{s_2^2}{n_2^2} n_2 - 1\right)} = \frac{\left(\frac{3.26^2}{10} + \frac{0.89^2}{10}\right)^2}{\frac{3.26^2}{10} / 10 - 1 + \frac{0.89^2}{10} / 10 - 1} = 10.33 \]

Which we round down to \( df = 10 \)

Reaching a conclusion

\[ t_{0.05(2), 10} = 2.23 \]

\[ t = 4.01 > 2.23 \]

So we can reject the null hypothesis with \( P < 0.05 \).

Extra dung near burrowing owl nests increases the number of dung beetles eaten.

Comparing the variance of two groups

\[ H_0 : \sigma_1^2 = \sigma_2^2 \]

\[ H_A : \sigma_1^2 \neq \sigma_2^2 \]

One possible method: the F test

The test statistic \( F \)

\[ F = \frac{s_1^2}{s_2^2} \]

Put the larger \( s^2 \) on top in the numerator.
\( F \ldots \)

- \( F \) has two different degrees of freedom, one for the numerator and one for the denominator. (Both are \( df = n_i - 1 \).) The numerator \( df \) is listed first, then the denominator \( df \).

- The \( F \) test is very sensitive to its assumption that both distributions are normal.

### Example: Variation in insect genitalia

<table>
<thead>
<tr>
<th></th>
<th>Polygamous species</th>
<th>Monogamous species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>-19.3</td>
<td>10.25</td>
</tr>
<tr>
<td>Sample variance</td>
<td>243.9</td>
<td>2.27</td>
</tr>
<tr>
<td>Sample size</td>
<td>7</td>
<td>9</td>
</tr>
</tbody>
</table>

\[ s_1^2 = 243.9 \quad s_2^2 = 2.27 \]

\[ F = \frac{s_1^2}{s_2^2} = \frac{243.9}{2.27} = 107.4 \]

Degrees of freedom

\[ df_1 = 7 - 1 = 6 \]
\[ df_2 = 9 - 1 = 8 \]

\[ F_{0.025,6,8} = 4.7 \]

For a 2-tailed test, we compare to \( F_{\alpha/2, df_1, df_2} \) from Table D.
Why $\alpha/2$ for the critical value?

By putting the larger $s^2$ in the numerator, we are forcing $F$ to be greater than 1.

By the null hypothesis there is a 50:50 chance of either $s^2$ being greater, so we want the higher tail to include just $\alpha/2$.

Conclusion

The $F=107.4$ from the data is greater than $F_{(0.025),6,8}=4.7$, so we can reject the null hypothesis that the variances of the two groups are equal.

The variance in insect genitalia is much greater for polygamous species than monogamous species.

Critical value for $F$

![Critical value for F](image)

The $F$ test is very sensitive to its assumption that both distributions are normal.

A more robust test to compare variances (between 2 or more groups) is:

*Levene’s test*

You should know that Levene’s test exists and why you would use it, but you do not need to know how to do it in this class. You would use a computer to do it, as the calculations are cumbersome.