Chapter 9 Comparison of Paired Samples

Fall 2010
9.1 Introduction
In a paired sample design, we model the data as if there is a single bucket of balls, and each draw from the bucket results in a pair of numbers (that we can distinguish as first and second).

This model applies if there are two measurements on each individual (often before and after), or if a pair of individuals are sampled together (such as twins, siblings, a matched pair design).

In a paired design, the method of analysis is as follows.
- Take individual differences for each pair.
- Treat the differences as a sample from a single population.
The data from a paired design can be tabulated in this form.

<table>
<thead>
<tr>
<th>Individual</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$d_i = Y_1 - Y_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Mean $\bar{y}_1$  $\bar{y}_2$  $d$

SD $s_1$  $s_2$  $s_d$

The important summary statistics are

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n} = \bar{y}_1 - \bar{y}_2 \quad \text{and} \quad s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n - 1}}$$
9.2 Paired-Sample t Test and Confidence Interval
Confidence Intervals

- The population mean difference is represented by $\mu_d$.
- The following confidence interval formula for $\mu_d$ is derived assuming that the population of differences has a normal distribution.

$$
\bar{d} - t_{\alpha/2} \frac{s_d}{\sqrt{n}} < \mu_d < \bar{d} + t_{\alpha/2} \frac{s_d}{\sqrt{n}}
$$

- If there are $n$ pairs, there are $n - 1$ degrees of freedom, and the area between $-t_{\alpha/2}$ and $t_{\alpha/2}$ is the confidence level $1 - \alpha$, often chosen to be 95%.
For hypothesis tests of the null hypothesis $H_0: \mu_d = 0$ versus either a one- or two-sided alternative, the test statistic is

$$t_d = \frac{\bar{d}}{s_d/\sqrt{n}}$$

and p-values are found by computing areas under $t$ distributions with $n - 1$ degrees of freedom.
Cyclic adenosine monophosphate (cAMP) is a substance that can mediate cellular response to hormones. In a study of maturation of egg cells in frogs, oocytes from each of four females were divided into two batches; one batch was exposed to progesterone and the other was not. After two minutes, each batch was assayed for its cAMP content.
Example (cont.) - Summary

- In an experiment, eggs from each of four female frogs are divided into two groups.
- One group of eggs from each frog is treated with progesterone, one is not.
- The cAMP level is measured for each group of eggs.

<table>
<thead>
<tr>
<th>Frog</th>
<th>Control</th>
<th>Progesterone</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>6.01</td>
<td>5.23</td>
<td>0.78</td>
</tr>
<tr>
<td>2</td>
<td>2.28</td>
<td>1.21</td>
<td>1.07</td>
</tr>
<tr>
<td>3</td>
<td>1.51</td>
<td>1.40</td>
<td>0.11</td>
</tr>
<tr>
<td>4</td>
<td>2.12</td>
<td>1.38</td>
<td>0.74</td>
</tr>
<tr>
<td>Mean</td>
<td>2.98</td>
<td>2.31</td>
<td>0.68</td>
</tr>
<tr>
<td>SD</td>
<td>2.05</td>
<td>1.95</td>
<td>0.40</td>
</tr>
</tbody>
</table>

$\mu_d = \text{population mean decrease in cAMP due to progesterone}$
1. Find a 95% confidence interval for $\mu_d$.
2. Test the hypothesis
   \[ H_0: \mu_d = 0 \quad H_A: \mu_d \neq 0. \]
3. Interpret both results in the context of this setting.
Example (cont.) - Chalkboard Calculations

- Do the Calculations on the Board.
We are 95% confident that the mean decrease in cAMP (pmol/oocyte) due to exposure to progesterone for two minutes under the given experimental conditions for eggs sampled from this population of frogs would be between 0.04 and 1.32.
Interpretation — Hypothesis Testing

There is evidence that exposure to progesterone under the experimental conditions causes a change in the mean cAMP levels in this population of frogs (two-sided paired t-test, $p = 0.042$).
9.4 Sign Test
Sign Tests

- The paired t-test assumes that differences are normally distributed.

- If this is not true, the Central Limit Theorem can be used to justify that the t-test is still valid, provided that the sample sizes are large enough. (As usual, how large is large enough depends on the character of the nonnormality in the population.)

- The sign test is a nonparametric test that does not depend on a normal assumption for the population of differences.

- To carry out a sign test, we ignore the magnitude of differences and just record whether each difference is positive or negative.
Sign Tests (cont.)

- P-values are then computed from a binomial distribution with $p = 0.5$.
- Technically, the sign test is not testing equality of population means.
- Instead, a sign test is testing if the differences are equally likely to be positive versus either a nondirectional or directional alternative.
Example

- The compound mCPP is thought to be a hunger suppressant.
- In an experiment, nine obese men had their weight change (kg) recorded after each of two two-week periods, once when taking a placebo and once when taking mCPP.
- There was a two week “washout period” between measurement periods.
Example

- Negative values indicate a weight loss.
- A negative difference indicates more weight was lost with mCPP than with the placebo.

<table>
<thead>
<tr>
<th>Subject</th>
<th>mCPPP</th>
<th>Placebo</th>
<th>difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0</td>
<td>-1.1</td>
<td>1.1</td>
</tr>
<tr>
<td>2</td>
<td>-1.1</td>
<td>0.5</td>
<td>-1.6</td>
</tr>
<tr>
<td>3</td>
<td>-1.6</td>
<td>0.5</td>
<td>-2.1</td>
</tr>
<tr>
<td>4</td>
<td>-0.3</td>
<td>0.0</td>
<td>-0.3</td>
</tr>
<tr>
<td>5</td>
<td>-1.1</td>
<td>-0.5</td>
<td>-0.6</td>
</tr>
<tr>
<td>6</td>
<td>-0.9</td>
<td>1.3</td>
<td>-2.2</td>
</tr>
<tr>
<td>7</td>
<td>-0.5</td>
<td>-1.4</td>
<td>0.9</td>
</tr>
<tr>
<td>8</td>
<td>0.7</td>
<td>0.0</td>
<td>0.7</td>
</tr>
<tr>
<td>9</td>
<td>-1.2</td>
<td>-0.8</td>
<td>-0.4</td>
</tr>
</tbody>
</table>
In the data, six of nine men lost more weight while using mCPP than when using the placebo. Is this difference significant (here consider a one-sided test)?

If mCPP had absolutely no effect, then we would expect the changes in weight to be random and either treatment would be equally likely to appear better for each individual.

With this null assumption, the number of individuals that lose more weight with mCPP than with a placebo is a binomial random variable with $n = 9$ and $p = 0.5$.

If mCPP has an effect, we would expect the proportion of men who lose more weight with the drug than with a placebo to be higher than 0.5.
The test statistic is the number of negative differences, 6.
The p-value is the probability of obtaining 6 or more successes in 9 independent trials with success probability $p = 0.5$.

Here is how to compute this in R.

```r
> p1 = sum(dbinom(6:9, 9, 0.5))
> p1
[1] 0.2539063

or

> p1 = 1 - pbinom(5,9,.5)
> p1
[1] 0.2539063
```
The sign test: test statistic

\[ N_+ = \text{number of positives}, \quad N_- = \text{number of negatives} \]

1. nondirectional \( (\mu_1 \neq \mu_2) \)

\[ B_s = \max(N_+, N_-) \]

2. \( \mu_1 > \mu_2 \)

\[ B_s = N_+ \]

3. \( \mu_1 < \mu_2 \)

\[ B_s = N_- \]
Calculating the p-value (two-sided)

\[ p = 2 \times \Pr \{ Y \geq B_s \}, \]

where \( Y \sim \mathcal{B}(n, 0.5) \)
Calculating the p-value (one-sided)

Two steps:

1. Check directionality—see if the data deviate from $H_0$ in the direction specified by $H_A$:
   1. If not, the p-value is greater than .50
   2. If so, proceed to Step 2.

2. $p = \Pr\{Y \geq B_s\}$, where

$$Y \sim \mathcal{B}(n, 0.5)$$
Example: rat experiment

8 rats were given a drug. Hemoglobin content of blood was measured before and after the drug. Where \( d = y_{\text{before}} - y_{\text{after}} \). Test whether the drug has an impact on hemoglobin content.

Data summary:
\( N_- = 1 \) and \( N_+ = 7 \).
Chalkboard Calculations
The sign test: What if there are ties?

Tie: \( y_1 = y_2 \) for some subject. Then \( d = 0 \) for this subject. No sign!!

Exclude all zeros, and decrease the sample size accordingly.

- idea: each pair whose difference is zero is ignored; such pairs are regarded as providing no evidence against \( H_0 \) in either direction.
Example

If the differences were

\[
\begin{array}{cccccccc}
2.2 & -0.9 & 0.0 & 1.1 & 0.6 & 2.9 & 1.2 & 2.0 \\
+ & - & 0 & + & + & + & + & + \\
\end{array}
\]

\(N_- = 1\) and \(N_+ = 6\) and \(n = 1 + 6 = 7\), not 8.