

Overview

- 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.

Structure of the Data

Notice that the mean of differences is equal to the difference of means. However, the standard error for a paired design does not equal the standard error computed from the same data assuming two independent samples.

Comparison of Paired Samples

Bret Larget

Department of Statistics

University of Wisconsin - Madison

November 12, 2003

Structure of the Data

The data from a paired design can be tabulated in this form.

Individual	Y_1	Y_2	$d_i = Y_1 - Y_2$
1			
2			
⋮			
n			
Mean	\bar{y}_1	\bar{y}_2	\bar{d}
SD	s_1	s_2	s_d

The important summary statistics are

$$\bar{d} = \frac{\sum_{i=1}^n d_i}{n} = \bar{y}_1 - \bar{y}_2$$

and

$$s_d = \sqrt{\frac{\sum_{i=1}^n (d_i - \bar{d})^2}{n - 1}}$$

Hypothesis Tests

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.

Example (cont.)

Find a 95% confidence interval for μ_d .

There are three degrees of freedom, so the t multiplier is 3.182 from the table, or

```
> qt(0.975, 3)
[1] 3.182446
```

Plugging in we have the following.

$$0.68 \pm 3.182 \times \frac{0.40}{\sqrt{4}} \quad \text{or} \quad 0.68 \pm 0.64$$

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.

Confidence Intervals

The population mean difference is represented by the symbol μ_d .

The following confidence interval formula for μ_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%.

Example — Exercise 9.3

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.

Frog	Control	Progesterone	Difference
1	6.01	5.23	0.78
2	2.28	1.21	1.07
3	1.51	1.40	0.11
4	2.12	1.38	0.74
Mean	2.98	2.31	0.68
SD	2.05	1.95	0.40

μ_d = population mean decrease in cAMP due to progesterone

Example — Hypothesis Testing

cAMP levels in this population of frogs (two-sided paired t-test, $p = 0.042$).

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.
- 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 — Hypothesis Testing

Consider a hypothesis test using the same data.

The hypotheses are

$$H_0: \mu_d = 0 \quad H_A: \mu_d \neq 0 .$$

The test statistic is

$$t = \frac{0.68}{0.40/\sqrt{4}} = 3.40 .$$

Using R, we find the p-value as follows.

```
> 2 * pt(-3.4, 3)
[1] 0.04246132
```

Using the t table, a one-sided p-value is between 0.02 and 0.025, so the two-sided p-value is between 0.04 and 0.05.

There is evidence that exposure to progesterone under the experimental conditions causes a change in the mean

Paired Design Settings

Paired designs can arise in a number of settings.

- Randomized block designs.
- Matched-pair designs.
- Repeated Measurements.

Example

Subject	mCPPP	Placebo	difference
1	0.0	-1.1	1.1
2	-1.1	0.5	-1.6
3	-1.6	0.5	-2.1
4	-0.3	0.0	-0.3
5	-1.1	-0.5	-0.6
6	-0.9	1.3	-2.2
7	-0.5	-1.4	0.9
8	0.7	0.0	0.7
9	-1.2	-0.8	-0.4

Calculation

Here are two ways to computer this in R.

```
> p1 = sum(dbinom(6:9, 9, 0.5))
> p2 = 1 - pbinom(5, 9, 0.5)
> c(p1, p2)
[1] 0.2539063 0.2539062
```

In the trial, two thirds of the men lost more weight with mCPPP than with a placebo, a result that is potentially of practical importance. However, there is very little statistical evidence in this small trial that the observed effect is real. The observed data is consistent with what we would expect to see if, in fact, mCPPP had no effect on weight loss ($p = 0.25$, paired one-tailed sign test).

Example

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

Here is the data. Negative values indicate a weight loss. A negative difference indicates more weight was lost with mCPPP than with the placebo.

Calculation

- In the data, six of nine men lost more weight while using mCPPP than when using the placebo. Is this difference significant?
- If mCPPP 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 mCPPP than with a placebo is a binomial random variable with $n = 9$ and $p = 0.5$. If mCPPP 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 simply 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 $p = 0.5$.

Example — Permutation Test

- A simple test statistic for the example data is the mean of the paired differences.
- For the mCPP data example, the mean paired difference is -0.5 .
- There are 9 pairs, so there are $2^9 = 512$ possible permutations of the labels within pairs.
- For a directional alternative hypothesis test that mean weight loss in this population taking mCPP is greater than the mean weight loss taking the placebo, we want to know the proportion of the 512 possible permutations would result in a mean of -0.5 or less.
- This would be tedious to compute by hand.
- Here is an R solution.

Example — Permutation Test with R

For the sake of comparison, we could carry out a t-test as well.

```
> t.test(mcpp, placebo, paired = T, alternative = "less")
Paired t-test

data: mcpp and placebo
t = -1.1915, df = 8, p-value = 0.1338
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf 2.803704
sample estimates:
mean of the differences
                -5
```

This result is very similar to the permutation test because the data is fairly consistent with the normal distribution assumption.

Permutation Tests

- A [permutation test](#) is an alternative way to test for differences in population means in a nonparametric fashion (so we do not assume that the population has a normal distribution, or any particular distribution, and therefore do not make distributional assumptions about the sampling distribution of the test statistic).
- The basic idea of all permutation tests is to consider the sampling distribution of some test statistic under all possible permutations of the labels of the observed data.
- P-values are then the proportion of permutations that lead to test statistics at least as extreme (relative to an alternative hypothesis) as the actual labeling of the data.
- For a paired design, we relabel within each pair which corresponds to considering all possible choices of signs to attach to each difference.

Example — Permutation Test with R

Permutation tests are not in the base R package. You need to load in the library [exactRankTests](#).

For some reason that I cannot fathom, the R function [perm.test](#) only works if the observed values are integers. That is why I multiply the actual data in this example by 10.

```
> mcpp = 10 * c(0, -1.1, -1.6, -0.3, -1.1, -0.9, -0.5, 0.7, -1.2)
> placebo = 10 * c(-1.1, 0.5, 0.5, 0, -0.5, 1.3, -1.4, 0, -0.8)
> library(exactRankTests)
> perm.test(mcpp, placebo, paired = T, alternative = "less")
1-sample Permutation Test

data: mcpp and placebo
T = 27, p-value = 0.1367
alternative hypothesis: true mu is less than 0
```

The p-value is 0.1367, indicating at most very weak evidence against the null hypothesis.