Comparison of Paired Samples

In the present chapter we consider the comparison of two samples that are **not independent** but are paired. In a paired design, the observations (Y_1, Y_2) occur in pairs; the observational units in a pair are linked in some way, so they have more in common with each other than with members of another pair.

One-sample t Tests

The one-sample t-test is a statistical hypothesis test used to determine whether an unknown population mean μ is different from a specific value. We consider the following two-sided hypothesis testing problem:

$$H_0 : \mu = c \text{ v.s. } H_A : \mu \neq c.$$

Similar to two-sample t tests in Chapter 7, for normal population we have

$$T = \frac{\bar{Y} - c}{\text{SE}_{\bar{Y}}} \sim t_{n-1}.$$

We would reject the null hypothesis H_0 at the α level of significance if $|T| > t_{n-1}(\alpha/2)$ or the p-value is less than or equal to α, where the p-value is the area under Student's t curve in the double tails beyond T and $-T$.

H_0	H_A	Test statistic	Rejection region	p-value
$\mu = c$	$\mu \neq c$		$|T| > t_{n-1}(\alpha/2)$	$2 \times P(t_{n-1} > |T|)$
$\mu \geq c$	$\mu < c$	$T = \frac{\bar{Y}-c}{\text{SE}_{\bar{Y}}} \overset{H_0}{\sim} t_{n-1}$	$T < -t_{n-1}(\alpha)$	$P(t_{n-1} < T)$
$\mu \leq c$	$\mu > c$		$T > t_{n-1}(\alpha)$	$P(t_{n-1} > T)$

Example: redwood tree

According to an official statement released by the state of California, the average height of a redwood tree in California is 350 ft. To assess this claim, a scientist collects a sample of 51 randomly chosen redwood trees in California. She finds that the average heights in her sample is 355 ft and the sample standard deviation is 42 ft.

- Conduct a hypothesis test to assess whether the true average height is 350 ft or not. Use $\alpha = 0.1$.
- Find the p-value for the sample.

Since $n = 50 > 30$, the central limit theorem applies and we can conduct a t test. Consider the following hypotheses

$$H_0 : \mu = 350 \text{ v.s. } H_A : \mu \neq 350.$$

The test statistic is

$$T = \frac{\bar{Y} - 350}{\text{SE}_{\bar{Y}}},$$

rejecting H_0 when $|T| < t_{n-1}(\alpha/2)$. Here

$$T = \frac{355 - 350}{42/\sqrt{51}} = 0.8501$$

and $t_{50}(0.1/2) = 1.676$. Because $|0.8501| < 1.676$, we fail to reject H_0 at the 0.01 level of significance.

The p-value for this test (found using a computer) is $2 \times P(t_{50} > 0.8501) = 0.3993$, which is greater than 0.05, so we do not reject H_0. (A quick look at t Table, using $\text{df} = 50$, shows that the p-value is between 0.20 and 0.40.)

The Paired-Sample t Test and Confidence Interval

In this section we discuss the use of Student's t distribution to obtain tests and confidence intervals for paired data.

Example: hunger rating

During a weight loss study, each of nine subjects was given (1) the active drug m-chlorophenylpiperazine (mCPP) for 2 weeks and then a placebo for another 2 weeks, or (2) the placebo for the first 2 weeks and then mCPP for the second 2 weeks. As part of the study, the subjects were asked to rate how hungry there were at the end of each 2-week period. The hunger rating data are shown in the table below, which also shows the difference in hunger ratings (mCPP - placebo).

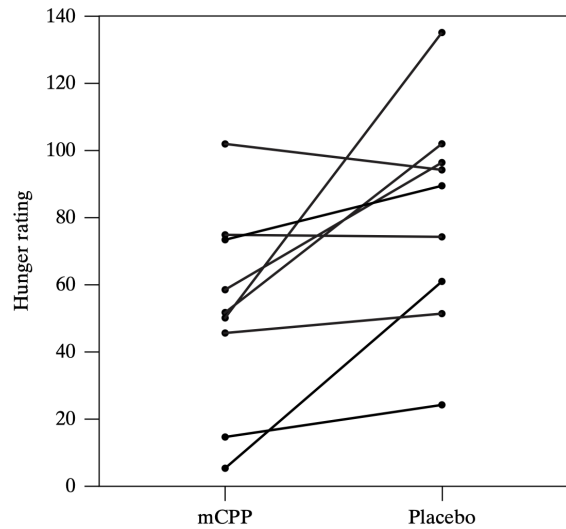| Table 8.1.1 Hunger rating for nine women			
	Hunger rating		Difference
Subject	Drug (mCPP)	Placebo	mCPP – Placebo
1	79	78	1
2	48	54	−6
3	52	142	−90
4	15	25	−10
5	61	101	−40
6	107	99	8
7	77	94	−17
8	54	107	−53
9	5	64	−59
Mean	55.3	84.9	−29.6
SD	31.5	34.1	32.8

Figure 8.1.1 Dotplots of hunger ratings after mCPP and placebo, with line segments connecting readings on each subject

In Chapter 7 we considered how to analyze data from two independent samples. When we have paired data, we make a simple shift of viewpoint: Instead of considering Y_1 and Y_2 separately, we consider the difference D, defined as

$$D = Y_1 - Y_2.$$

Let us denote the mean of sample D's as $\bar{D}$. The quantity $\bar{D}$ is related to the individual sample means as follows:

$$\bar{D} = (\bar{Y}_1 - \bar{Y}_2).$$

The relationship between population means is analogous:

$$\mu_D = \mu_1 - \mu_2.$$

Thus, we may say that the mean of the difference is equal to the difference of the means. Because of this simple relationship, a comparison of two paired means can be carried out by concentrating entirely on the D's. The standard error for D is easy to calculate. Because D is just the mean of a single sample, we can apply the SE formula of Chapter 6 to obtain the following formula:

$$\text{SE}_{\bar{D}} = \frac{s_D}{\sqrt{n_D}},$$

where s_D is the standard deviation of the D's and n_D is the number of D's.

For the hunger rating data, note that the mean of the difference is equal to the difference of the means:

$$\bar{D} = -29.6 = 55.3 - 84.9.$$

While the mean of the difference is the same as the difference of the means, note that the SD of the difference is not the difference of the SDs. Thus the standard error of the mean

difference must be calculated using the SD of the differences. We calculate the standard error of the mean difference as follows:

$$s_D = 32.8, n_D = 9, \text{SE}_{\bar{D}} = \frac{32.8}{\sqrt{9}} = 10.9.$$

Confidence intervals for μ_D

- The two-sided $1 - \alpha$ confidence interval for μ_D is

$$\bar{D} \pm t_{n_D-1}(\alpha/2) \times \frac{s_D}{\sqrt{n_D}}.$$

- The upper one-sided $1 - \alpha$ confidence interval for μ_D is

$$\left(-\infty, \bar{D} + t_{n_D-1}(\alpha) \times \frac{s_D}{\sqrt{n_D}}\right).$$

- The lower one-sided $1 - \alpha$ confidence interval for μ_D is

$$\left(\bar{D} - t_{n_D-1}(\alpha) \times \frac{s_D}{\sqrt{n_D}}, \infty\right).$$

For the hunger rating data, we have $n_D = 9$ and thus the degrees of freedom for the t distribution is $9 - 1 = 8$. From t Table we find that $t_8(0.025) = 2.306$; thus, the (two-sided) 95% confidence interval for μ_D is

$$-29.6 \pm 2.306 \times \frac{32.8}{\sqrt{9}}$$

or $(-54.8, -4.4)$.

Hypothesis tests for paired samples

We can also conduct a t test. To test the hypotheses

$$H_0 : \mu_D = 0 \text{ v.s. } H_A : \mu_D \neq 0,$$

we use the test statistic

$$T = \frac{\bar{D} - 0}{\text{SE}_{\bar{D}}}.$$

Critical values are obtained from Student's t distribution with degrees of freedom being $n_D - 1$.

For the hunger rating data, let us formulate the null hypothesis and alternative hypothesis:

H_0: Mean hunger rating is the same after taking mCPP as it is after taking a placebo.

H_A: Mean hunger rating is different after taking mCPP than after taking a placebo.

Or, in symbols,

$$H_0 : \mu_D = 0 \text{ v.s. } H_A : \mu_D \neq 0.$$

Let us test H_0 against H_A at significance level $\alpha = 0.05$. The test statistic is

$$T = \frac{-29.6 - 0}{32.8/\sqrt{9}} = -2.71.$$

From t Table, $t_8(0.02) = 2.449$ and $t_8(0.01) = 2.896$. We reject H_0 and find that there is sufficient evidence $(0.02 < p\text{-value} < 0.04)$ to conclude that mean hunger rating is reduced more by mCPP than by a placebo. (Using a computer gives the p-value$= 0.027$.)

Conditions for validity of the paired-sample t test and confidence interval

- It must be reasonable to regard the differences (the D's) as a **random sample** from some large population.
- The population distribution of the D's must be **normal**. The methods are approximately valid if the population distribution is approximately normal or if the sample size (n_D) is large.

The preceding conditions are the same as those given in Chapter 6; in the present case, the conditions apply to the D's because the analysis is based on the D's.

Examples of paired designs

Paired designs can arise in a variety of ways, including the following:

- Experiments in which similar experimental units form pairs
    - Often, researchers who wish to compare two treatments will first form pairs of experimental units (pairs of animals, pairs of plots of land, etc.) that are similar (e.g., animals of the same age and sex or plots of land with the same type of soil and exposure to wind, rain, and sun). Then one member of a pair is randomly chosen to receive the first treatment and the other member is given the second treatment.
- Observational studies of identical twins
    - For example, in a study of the effect of "secondhand smoke" it would be ideal to enroll several sets of nonsmoking twins for which, in each pair, one of the twins lived with a smoker and the other twin did not.

- Repeated measurements on the same individual at two different times
    - Many biological investigations involve repeated measurements made on the same individual under different conditions. These include studies of growth and

development, studies of biological processes, and studies in which measurements are made before and after application of a certain treatment.
- Pairing by time
  - In some situations, pairs are formed implicitly when replicate measurements are made at different times.

Two-sample v.s. paired-sample

- In a two-sample test, you compare two **independent** samples. These samples are expected to be uncorrelated due to their independence. The sample sizes for the two groups **do not have to be equal**, although it is often preferable to have roughly equal sizes.

- In a paired-sample test, you work with a single sample of paired observations. This typically consists of paired measurements on individual subjects, such as "before and after" scores on a questionnaire, exam, or lab test. Alternatively, the pairs can represent pairs of subjects, such as married couples or twins, or pairs of objects produced together. The measurements within pairs are expected to be correlated. In the analysis, you would examine the differences between the paired measurements, denoted as $D_i = Y_{1i} - Y_{2i}$.