

Paired vs. Independent Samples

- *Paired samples* arise from a study design in which the units in the samples are carefully matched in pairs so that the units in each pair are as similar as possible, except for the 2 treatments that have been imposed.
- In a paired samples design the units in one data set are directly related to a specific observation in the other data set. Thus the two samples are *dependent*.

Example 1: An educational psychologist studied whether two mathematics achievement tests (Test 1, Test 2) lead to different achievement scores. Eight subjects were randomly selected and each was given the 2 tests. The order of the 2 tests was independently randomized for each subject.

Subject	1	2	3	4	5	6	7	8
Test 1	93	70	79	91	83	106	94	87
Test 2	81	103	88	85	84	100	103	102

- In this example μ_1 is the true mean score on Test 1 and μ_2 is the true mean score on Test 2.
- There are 2 samples of 8 measurements each but the measurements are paired by subject, *i.e.* there are 8 natural pairs of measurements.
- The 2 samples are dependent. If we know that Subject 1 scored high on Test 1 (relative to the other students) then we would expect Subject 1 to score high on Test 2.

Example 2: Another educational psychologist reported the results of a study in which 7 pairs of children reading below grade level were obtained by matching so that within each pair the 2 children were equally deficient in reading ability. One child from each pair received a new experimental training, while the other received the standard training (the assignment of treatments within each pair was random). The psychologist was interested in determining if the new treatment was superior to the old. Improvement scores are shown below.

Pair	1	2	3	4	5	6	7
Experimental	0.5	1.0	0.6	0.1	1.3	0.1	1.0
Standard	0.8	1.1	0.1	0.2	0.2	1.5	0.8

- In this example μ_1 is the true mean improvement score for the experimental treatment and μ_2 is the true mean improvement score for the standard treatment.
- There are 2 samples of 7 measurements each. Even though there are 14 children they have been deliberately matched (or *paired*) with respect to reading ability. Thus there are 7 natural pairs of measurements.
- The 2 samples are dependent. If the child receiving the standard treatment in pair 1 gets a low improvement score (relative to the improvement scores of others receiving the standard treatment) then we expect his/her "partner" to also get a low improvement score for the experimental treatment (relative to the improvement scores of others receiving the experimental treatment).

- *Independent samples* arise from a study design in which statistically independent samples are drawn from each population and comparative information about the 2 populations is derived from a comparison of the independent samples.
- We have 2 independent samples when 2 unrelated sets of units are measured. In this setting the 2 sample sizes do not have to be equal (as they do in the paired design).

Example 3: A market researcher randomly divided 500 households into 2 equal groups of 250 households. Group 1 was interviewed by phone using the current manual procedure. Group 2 received the same interview, but computer-assisted interviewing was used. The researcher was interested in estimating the difference in the mean interview times for the 2 procedures (μ_1 and μ_2).

Example 4: In a small-scale experiment to compare the pain-relieving effectiveness of 2 new medications for arthritis sufferers, 27 volunteers were divided at random into 2 groups of size 14 and 13. Group 1 received Medication 1 and Group 2 received Medication 2. The parameters are μ_1 (the true mean hours of pain relief of those taking Medication 1) and μ_2 (the true mean hours of pain relief of those taking Medication 2).

Which Design is Better?

- When a choice between study designs is available, you want to choose the design that offers for a given cost the smaller sampling error. Usually paired designs are more efficient.

Example 5: This is a totally unrealistic example designed to illustrate why paired designs are usually more efficient. A *population* of 5 banana bunches will be shipped from Central America to the USA. The banana company wants to estimate the weight loss of the bunches during shipment. The pre- and post-shipping weights of the bunches (in kilograms) is shown below.

Bunch	Weight Before Shipment (kg)	Weight After Shipment (kg)
1	34	32
2	29	27
3	31	29
4	28	26
5	23	21

In this idealized example, each of the 5 bunches loses 2 kilograms. Thus the true mean weight loss is 2 kg. If independent samples of size 2 were drawn before and after shipment the uniform weight loss would be hidden by the variation among the weights of bunches. For example, if bunches 1 and 5 were weighed before shipment the sample mean weight would be $(34+23)/2 = 28.5\text{kg}$. If bunches 4 and 5 were weighed after shipment then the sample mean weight would be $(26+21)/2 = 23.5\text{kg}$. The estimated mean weight loss would be 5.0 kg, considerably higher than the true mean weight loss of 2.0 kg. However, the use of paired samples would always reveal exactly the 2.0 kg weight loss per bunch.

- In general, if the paired samples are positively related (as they frequently will be), the paired design will lead to smaller sampling errors than the use of the independent samples design.
- Independent designs are used frequently, in fact they are more common than paired designs. It may be too expensive and/or logistically impractical to carry out a paired design even when theoretically it would be better to do so.

ANALYZING PAIRED SAMPLES

Notation

- Y_{ij} denotes the j th treatment response for the i th pair.
- The basic data layout for paired designs is shown below.

Pair	Treatment 1	Treatment 2	Difference
1	Y_{11}	Y_{21}	$D_1 = Y_{11} - Y_{21}$
2	Y_{12}	Y_{22}	$D_2 = Y_{12} - Y_{22}$
•	•	•	•
•	•	•	•
•	•	•	•
n	Y_{1n}	Y_{2n}	$D_n = Y_{1n} - Y_{2n}$

- The true unknown population means are μ_1 (Treatment 1) and μ_2 (Treatment 2) and we are interested in the mean difference $\mu_D = \mu_1 - \mu_2$.
- Relevant summary statistics are
 - Sample mean difference: $\bar{D} = \frac{\sum D}{n}$. This is the estimator of μ_D .
 - Sample standard deviation of the differences: $s_D = \sqrt{\frac{\sum (D - \bar{D})^2}{n-1}}$.
 - The standard error of \bar{D} :

$$s_{\bar{D}} = \frac{s_D}{\sqrt{n}}.$$

- The sample differences are assumed to be a simple random sample of size n from a population of differences.
- Thus, we find ourselves back in the one-sample setting of Chapter 10. The only difference from here on will be notational.
- We will be using t -tests and t -confidence intervals. These are referred to as *paired t -procedures* when they are applied to data collected from a paired design.

Assumptions of Paired t -Procedures

- The differences D_1, D_2, \dots, D_n are a simple random sample from a population of all possible differences.
- The differences follow a normal distribution with unknown mean μ_D and unknown variance σ_D^2 .

Paired t -Test

- In general we will be interested in determining if there is a difference between the two treatments. Thus the null hypothesis is $H_0 : \mu_D = 0$.
- Testing hypotheses about μ_D will be based on the test statistic

$$t = \frac{\bar{D}}{s_{\bar{D}}} = \frac{\bar{D}}{\frac{s_D}{\sqrt{n}}}.$$

- **IF THE NULL HYPOTHESIS IS TRUE** then t has a t -distribution with $n - 1$ degrees of freedom .
- In terms of a random variable t having a $t(n - 1)$ reference distribution, the P -value for a test of H_0 against

$$\mu_D \geq 0 \quad \text{is} \quad P(t \geq t_o),$$

$$\mu_D \leq 0 \quad \text{is} \quad P(t \leq t_o),$$

$$\mu_D \neq 0 \quad \text{is} \quad 2P(t \geq |t_o|)$$

where t_o is the observed value of the test statistic. Thus, we reject H_0 if the P -value $\leq \alpha$. Otherwise, we fail to reject H_0 .

- The P -values are exact when the population distribution is normal and are approximately correct for large n in other cases.

A $(1 - \alpha)$ 100% Confidence Interval for μ_D

- A $(1 - \alpha)$ 100% confidence interval for μ_D is given by

$$\bar{D} \pm t \left(1 - \frac{\alpha}{2}, n - 1 \right) \left(\frac{s_D}{\sqrt{n}} \right)$$

where $t \left(1 - \frac{\alpha}{2}, n - 1 \right)$ is the appropriate percentile of the $t(n - 1)$ distribution.

- The interval is exact when the population distribution is normal and is approximately correct for large n in other cases.