7: Paired Samples

Data

Paired samples vs. independent sample

This chapter considers the analysis of a **quantitative outcome based on paired samples**. Paired samples (also called **dependent samples**) are samples in which natural or matched couplings occur. This generates a data set in which each data point in one sample is uniquely paired to a data point in the second sample.

Examples of paired samples include:

- *pre-test/post-test* samples in which a factor is measured before and after an intervention,
- *cross-over trials* in which individuals are randomized to two treatments and then the same individuals are crossed-over to the alternative treatment,
- *matched samples,* in which individuals are matched on personal characteristics such as age and sex,
- *duplicate measurements* on the same biological samples, and
- *any* circumstance in which each data point in one sample is uniquely matched to a data point in the second sample.

The “opposite” of paired samples is **independent samples.** Independent samples consider unrelated groups. Independent samples may be achieved by randomly sampling two separate populations or by randomizing an exposure to create two separate treatment groups without first matching subjects.

Illustrative dataset—“oatbran”

A cross-over trial investigated whether eating oat bran lowered serum cholesterol levels. Fourteen (14) individuals were randomly assigned a diet that included either *oat bran* or *corn flakes*. After two weeks on the initial diet, serum cholesterol were measured and the participants were then “crossed-over” to the alternate diet. After two-weeks on the second diet, cholesterol levels were once again recorded.

Data appear below. The variable CORNFLK in the table represents cholesterol level (mmol/L) of the participant on the corn flake diet. The variable OATBRAN represents the participant’s cholesterol on the oat bran diet.

Illustrative data set (OATBRAN)

ID	CORNFLK (mmol/L)	OATBRAN (mmol/L)
1	4.61	3.84
2	6.42	5.57
3	5.40	5.85
4	4.54	4.80
5	3.98	3.68
6	3.82	2.96
7	5.01	4.41
8	4.34	3.72
9	3.80	3.49
10	4.56	3.84
11	5.35	5.26
12	3.89	3.73
13	2.25	1.84
14	4.24	4.14

As background—this is not the main analysis—it helps to calculate summary statistics for each sample separately. Let sample 1 represent CORNFLK values and let sample 2 represent OATBRAN values. Using a calculator or computer, we determine:

$\bar{x}_1 = 4.444$ $s_1 = 0.9688$ $n_1 = 14$
$\bar{x}_2 = 4.081$ $s_2 = 1.0570$ $n_2 = 14$

Difference variable DELTA

Further analysis requires creation of a new variable to hold information about the **difference within pairs**; we call this created variable **DELTA**. When creating DELTA values, it makes little difference whether you subtract sample 1 values from sample 2 values, or vice versa. It is important, however, to keep track of the direction of the difference. For these data, let DELTA = CORNFLK - OATBRAN. Thus, positive DELTA values will reflect higher cholesterol levels on the corn flake diet and negative values will reflect higher cholesterol values on the oat bran diet.

ID	CORNFLK (mmol/L)	OATBRAN (mmol/L)	DELTA
1	4.61	3.84	0.77
2	6.42	5.57	0.85
3	5.40	5.85	-0.45
4	4.54	4.80	-0.26
5	3.98	3.68	0.30
6	3.82	2.96	0.86
7	5.01	4.41	0.60
8	4.34	3.72	0.62
9	3.80	3.49	0.31
10	4.56	3.84	0.72
11	5.35	5.26	0.09
12	3.89	3.73	0.16
13	2.25	1.84	0.41
14	4.24	4.14	0.10

Additional analyses are now directed toward the DELTA variable.

Descriptive and exploratory statistics

It is important to describe and explore the distribution of the within-pair differences (DELTA). Use your calculator or any other computational device to calculate summary statistics for the DELTA value. (Summary statistics were initially covered in Chapter 3). At minimum, report the sample size, mean, and standard deviation. Use the subscript $_d$ to denote that these statistics are for the DELTA variable.

n_d = 14 $\bar{x}_d$= 0.3629 s_d = 0.4060 max$_d$ = 0.86 min$_d$ = -0.45

Narratively, describe your findings, e.g., OATBRAN was associated with 0.36 mmol/L lower cholesterol than CORNFLK (*n* = 14, standard deviation 0.41 mmol/L). That's about an 8% decrease (0.36 / 4.44 = .08).

Then **explore** the distribution of DELTA values via stemplot, boxplot, or whatever graphical method is most informative. A simple stemplot might look like this:

```
-0 | 42
 0 | 011334
 0 | 667788
x1
```

Note the requirement for the negative zero stem value to contain values between –0.49 to –0.01.

Interpretation of stemplot. While providing limited information on the shape of the distribution (because of the small *n*), it is clear that values range from approx −0.4 to +0.8. The median has a depth of (14 + 1) / 2 = 7.5 which puts it between 0.3 and 0.4.

Comment: After some trial and error, I found that quintuple split of the stem provides this plot:

```
-0f | 4
-0t | 2
-0* |
 0* | 011
 0t | 33
 0f | 4
 0s | 6677
 0. | 88
  x 1
```

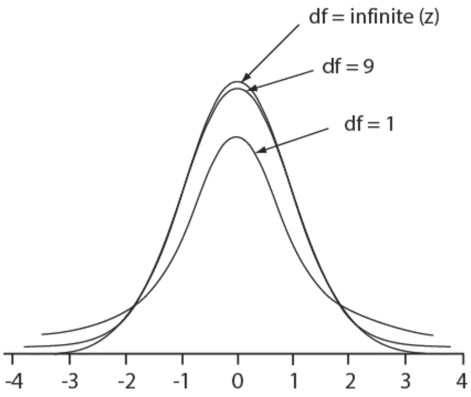
The symbols next to these stem values are reminders of sub-range. For example, the "f" stands for "four" and "five," so "-0f" reserves a space for values between −0.5[9] and −0.4[0].

Inferential statistics

Student's *t* pdf

Inferential methods in this chapter rely on a pdf called **Student's *t***. *t* pdfs are continuous, symmetrical, and centered on 0. They are similar to *a z* pdf but with slightly fatter tails. [Recall that a *z* is a normal pdf with μ = 0 and σ = 1.]

There are many different *t* pdfs, each identified by its **degree of freedom (df)**. The larger the *df*, the more the *t* resembles a *z*. A *t* with infinity *df* is the same as a *Z*!



Estimation

Parameter and point estimate

The parameter we wish to infer is the **expected mean difference** $\mu_{d.}$ The sample mean difference $\bar{x}_d$ is the **point estimator** of μ_d. $\bar{x}_d$ for the illustrative data is 0.363 mmol/L. This is the "maximally likely" *estimate* of the expected effect of the diet change. However, it provides no information about the precision of the estimate.

Interval estimation

The standard point "estimate ± margin of error" approach is used to calculate the confidence interval. The (1 – α)100% CI for μ_d =

$$\bar{x}_d \pm t_{1-\frac{\alpha}{2},n-1} \cdot SEM_d$$

where $t_{1\text{-}\alpha/2,\ n\text{-}1}$ is the *t* percentile with *n* – 1 *df* for (1 – α)100% confidence [from the *t* table] and the **standard error of the mean difference** $SEM_d = \frac{s_d}{\sqrt{n}}$.

Illustration. To determine and interpret the 95% CI for μ_d , *df* = *n* – 1 = 14 – 1 = 13. For 95% confidence, use $t_{.975,13}$= 2.16 [from the *t* table]. Use the n_d and s_d determined earlier in this chapter to calculate $SEM_d = \frac{0.4060}{\sqrt{14}}$ = 0.1085. The 95% CI for μ_d = $\bar{x}_d \pm t_{1-\frac{\alpha}{2},n-1} \cdot SEM_d$ = 0.3629 ± 2.16 · 0.1085 = 0.3629 ± 0.2344 = (0.129, 0.597) mmol/l. Interpretation: This CI is trying to capture μ_d, *not* $\bar{x}_d$. The margin of error is ±0.23. We consider the full extent of the interval from its lower limit (0.129) to its upper limit (0.597).

Required sample size to attain a given margin of error

To limit the margin of error of a (1 – α)100% confidence interval for μ_d to *m*, the sample size should be no less than

$$n = f \times n' \text{ where } f = (df + 3) / (df + 1) \text{ and } n' = \left(z_{1-\frac{\alpha}{2}} \frac{s_d}{m} \right)^2$$

Note that s_d is the sample standard deviation of the within-pair differences, $z_{1-(\alpha/2)}$ is the standard normal deviate for (1 – α)100% confidence, and *m* is the desired margin of error. When *n'* > 30, there is no need to multiply *n'* by *f* as *f* is very close to 1.

Comment: *f* compensate for the additional imprecision in using *s* instead of σ in *t* procedures. When *n'* ≥ 30 there is no need to multiply by *f* because $t_{30+} \approx z$.

Illustration. How large a sample is needed to generate a margin of error of 0.3 mmol/L for the illustrative data?

ANS: $n' = \left(1.96 \frac{0.4060}{0.3} \right)^2$ = 7.03. Since *n'* is less than 30, multiply by correction factor *f* where *f* = (6+3) / (6+1) =1.286. Thus, *n* = *f* × *n'* = 1.286 × 7.03 = 9.04 → resolve to study 10 individuals.

Illustration. How large a sample is needed to cut the margin of error down to 0.1 mmol/L?

ANS: $n' = \left(1.96 \frac{0.4060}{0.1} \right)^2$ = 63.32 → resolve to study 64 individuals. Multiplication by *f* is unnecessary since *n'* exceeded 30.

Null hypothesis significance test

Hypotheses

We are looking for a significant positive or negative mean difference. Under the null hypothesis, we expect no mean difference; H_0: μ_d = 0. Under the alternative hypothesis, we expect a non-zero mean difference; H_1: $\mu_d \neq 0$.

Test statistic and *P* value

This **paired *t* statistic** is needed to determine the *P* value:

$$t_{\text{stat}} = \frac{\text{observed mean difference} \; - \; \text{expected mean difference when } H_0 \text{ true}}{\text{SEM}_d}$$

where the observed mean difference is $\bar{x}_d$, the expected mean difference under H_0 is nearly always set to 0, and $SEM_d = \frac{s_d}{\sqrt{n}}$

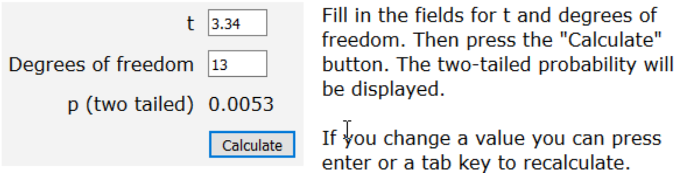
P value corresponds to the AUC in the tails beyond the $-|t_{\text{stat}}|$ and $+|t_{\text{stat}}|$ in the *t* pdf with df = *n* – 1. Use an app (illustrated below) to find the AUC in the tails of the appropriate *t* pdf.[1]

Reporting and interpretation

The results of the test should be reported in plain language and should include a consideration of the observed mean difference and *P* value. The *P* value answers the question "What is the probability of seeing the observed mean difference or a mean difference more extreme assuming H_0 is true?" Small *P* value is evidence against H_0, especially when *P* gets below, say, 10%. The results get more and more "significant" as the *P* value gets lower-and-lower.

Illustration. Test the OATBRAN data for significance.

- H_0: μ_d = 0 vs. H_1: $\mu_d \neq 0$
- $t_{\text{stat}} = \frac{0.3629 - 0}{0.1085}$ = 3.34 with df = 14 - 1 = 13. [We had established that n_d = 14, $\bar{x}_d$= 0.3629, and $SEM_d = \frac{0.4060}{\sqrt{14}}$ = 0.1085 earlier in the chapter.]
- The "two tails of t app"[2] is used to derive AUCs in the tails of t_{13} beyond ±3.34. The input screen for the app should look like this:



t 3.34

Degrees of freedom 13

p (two tailed) 0.0053

Fill in the fields for t and degrees of freedom. Then press the "Calculate" button. The two-tailed probability will be displayed.

If you change a value you can press enter or a tab key to recalculate.

- Interpretation: The oat bran significantly decreased cholesterol by an average of 0.36 mmol/L (*P* = .0053).

[1] Use of *t* tables are discouraged for finding *P* values because they provide only approximate *P* values which are even more apt to be misinterpreted than exact *P* values.

[2] http://onlinestatbook.com/2/calculators/t_dist.html

Conditions for paired *t* procedures

All valid statistical inference require underlying conditions. Paired *t* procedures are no exception. When we assume these conditions are present and they are in fact not, the inferential statistics that follow are unreliable.

Paired *t* procedures require the following conditions:

(1) No selection bias

Selection bias occurs when the process used to identify study subjects tends to create a sample that systematically differs from the population to which inference will be made. In order for inferences to be valid, selection bias must be absent. This is sometimes referred to as “the SRS assumption” because it assumes that the data represent a simple random reflection of the underlying population.

(2) No information bias

Information bias occurs when the data are inaccurate. An analysis is only as good as the quality of its data. Remember the GIGO principal? In order for inferences to be valid, information bias must be absent and measurements must be valid.

(3) The normality assumption

The normality assumption is often misunderstood. This assumption does <u>not</u> require the population to resemble a normal pdf. However, the sampling distribution of the mean (SDM) difference should be *approximately* normal.

Recall that the SDM is hypothetical and really doesn’t exist. Also recall that the central limit theorem will impart normality to the hypothetical distribution when the underlying population is symmetrical and the sample is moderate to large in size. However, the central limit theorem is weak in small samples. Therefore, when analyzing small samples, the underlying population should be approximately normal. It is wise to explore the shape of the distribution to check for major violations in normality (e.g., extreme asymmetry) when the sample is small before using *t* procedures.

Paired difference analysis (summary)

Exploration and description

1. Read the research question, verify sample is paired and outcome is quantitative.
2. If not already given, calculate within-pair differences (DELTAs).
3. Calculate $\bar{x}_d$, s_d, and n_d. If data are asymmetrical, report the 5-point summary.
4. Plot the DELTAs and explore the distribution's location, spread, and shape.

Estimation

1. Read the research question, verify that the data are quantitative and based on paired samples. Confirm that the parameter of interest is μ_d.
2. Calculate the point estimate $\bar{x}_d$ and the $SEM_d = \frac{s_d}{\sqrt{n}}$
3. Calculate the (1 – α)100% CI for μ_d $= \bar{x}_d \pm t_{1-\frac{\alpha}{2},n-1} \cdot SEM_d$
4. Report the point estimate for the mean difference, direction of the difference (increase or decrease), and the CI. Round appropriately (approx. 3 significant digits). Include units of measure.
5. Sample size for limiting the margin of error was also covered in this chapter.

NHST

1. Read the research question, verify that the data are quantitative and based on paired samples. Confirm that the parameter of interest is μ_d. State H_0: $\mu_d = 0$.
2. Calculate $t_{\text{stat}} = \frac{\bar{x}_d - \text{expected mean difference when } H_0 \text{ true}}{\text{SEM}_d}$; df = $n - 1$
3. Determine *P* value (app).
4. Report the point estimate, note direction of the difference (increase or decrease), and the *P* value. Use plain language and round appropriately; be kind to your reader.

Sample size to limit *m* was also considered earlier in this chapter.

Illustration using OATBRAN data

Exploration and description. This cross over trial looked at within-pair differences on oatbran and corn flake diet. There were $n = 14$ paired observations. **Data, procedures, and calculations are shown throughout in this chapter.** Here, we present only the interpretation of key results. Differences (CORNFLK – OATBRAN, mmol/l) are fairly symmetrical (stemplot below) with $\bar{x}_d$= 0.362, median approx. 0.35, and s_d = 0.406.

```
-0 | 42
 0 | 011334
 0 | 667788
x1
```

Estimation. The oat bran diet *decreased* cholesterol by an *average* of 0.362 mmol/l, 95% CI for μ_d = (0.129 - 0.597) mmol/l.

NHST. The decline of 0.362 mmol/l was statistically significant (*P* = .0053).

Advanced Topic – Power and Sample Size

Statistical power of a paired *t* test

The concept of statistical power falls outside the realm of Fisherian NHST but is nonetheless useful in interpreting negative NHST results and calculating the sample size requirements of a test. To understand this concept, we must first accept these definitions.

A type I error occurs when we reject a true H_0. The type I error rate is called alpha (α).

A type II error occurs when we fail to reject a false H_0. The probability of a type II error is called beta (β): β ≡ Pr(retaining a false H_0)

The complement of β, 1 – β, is called "power." Power is the probability of correctly rejecting a false H_0: 1 – β ≡ Pr(rejecting a false H_0)

The power of a paired *t* test can be calculated if we state ahead of time

- the desired type I error rate called α level (usually .05),
- the difference worth detecting (call it Δ),
- the number of pairs tested (n_d), and
- the standard deviation of the within paired differences (call this σ_d)

Then:

$$1-\beta=\Phi\left(-z_{1-\frac{\alpha}{2}}+\frac{|\Delta|\sqrt{n_d}}{\sigma_d}\right)$$

where $\Phi(z)$ is the cumulative probability of the value inside the parentheses on the standard normal pdf.[3] A statistical power of ≥ 80% or ≥ 90% is considered adequate.

Illustration. What was the statistical power of the NHST on prior page for detecting a Δ of 0.3?

ANS: $1-\beta=\Phi\left(-1.96+\frac{|0.3|\sqrt{14}}{0.4060}\right)$ = $\Phi(0.8047)$ = 0.7895 (see figure below).

Note: Mathematically weak student populations may find the above calculation intimidating and should instead focus on the definitions and "inputs" that go into the calculation. Use of the http://www.statstodo.com/SSizPairedDiff_Pgm.php

[3] Determine with cumulative z table or app http://onlinestatbook.com/2/calculators/normal_dist.html

Sample size requirements to achieve a desired power

You can increase the power of the NHST or create conditions by which it can detect smaller "differences worth detecting" more reliably by increasing the sample size of the study. The sample size requirements of a paired *t* NHST depend on:

- The desired power of the study (1 – β)
- The desired "significance level" of the test (α)
- The standard deviation of the differences (σ_d)
- The difference worth detecting (Δ)

According to these presets the required sample size of a paired *t* test is

$$n = \frac{\sigma_d^2 (z_{1-\beta} + z_{1-\alpha/2})^2}{\Delta^2}$$

Note that power and alpha levels are expressed as percentiles on a standard normal curve, denoted $z_{\text{cumulative probability}}$. For example, for 90% power $z_{.9} = 1.28$. For an alpha level of .05, $z_{1-(.05/2)} = z_{.975} = 1.96$.

Illustration. What is the sample size required to achieve 90% power for a significance test with an alpha of .05 for a variable with a standard deviation of 0.40. The difference we wish to detect is 0.1.

$n = \frac{.4^2(1.28+1.96)^2}{0.1^2}$ = 167.96 → resolve to study 168 matched pairs.

If *n* < 30, multiply by *f* = (df + 3) / (df + 1) to compensate for the *t* pdf.