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:

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

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.