Stat 4473 – Data Analysis
Paired Data

Paired data arise when:

(1) we get two measurements from each subject, or
(2) we get one measurement from each of two subjects sharing the same characteristic.

Crest and Paired Samples

In the late 1950's, Proctor & Gamble introduced Crest toothpaste as the first such product with fluoride. To test the effectiveness of Crest in reducing cavities, researchers conducted experiments with several sets of twins. One of the twins in each set was given Crest with fluoride, while the other twin continued to use ordinary toothpaste without fluoride. It was believed that each pair of twins would have similar eating, brushing, and genetic characteristics. Results showed that the twins who used Crest had significantly fewer cavities than those who did not. This use of twins as paired samples allowed the researchers to control many of the different variables affecting cavities. Pairing removes extra variation and allows us to focus on the differences in the response.

Note that pairing didn't have to be done in this experiment. But it made the experiment a better one. Sometimes paired samples are called "matched pairs."

Notation for analyzing paired samples

μ_d = mean value of the differences for the population of paired data

$\overline{X}_d$ = mean of the sample of differences

s_d = standard deviation of the sample of differences

Because it's the *differences* that we care about, we'll treat them as if *they* were the data, ignoring the original two columns. Now that we have only one column of values to consider, we can use the regular t-interval to estimate the true mean difference. Mechanically, making confidence intervals for matched pairs follows exactly the steps for a one-sample t-interval, including checking the nearly normal condition. The formula is given by

$$\overline{X}_d \pm t^* \frac{s_d}{\sqrt{n}}$$

where t* is the value from the t-distribution with n -1 degrees of freedom (n is the number of pairs) for the confidence level chosen.

Example 1

We hear that listening to Mozart improves students' performance on tests. In the EESEE story "Floral Scents and Learning," investigators asked whether pleasant odors have a similar effect. Twenty-one subjects worked a paper-and-pencil maze while wearing a mask. The mask was either unscented or carried a floral scent. The response variable is their average time on three trials. Each subject worked the maze with both masks, in a random order. The randomization is important because subjects tend to improve their times as they work a maze repeatedly. The table below gives the subjects' average times with both masks. Find a 95% confidence interval for μ_d = average difference in time to complete the maze. Interpret your interval estimate.

Subject	Unscented Time (sec)	Scented Time (sec)	Subject	Unscented Time (sec)	Scented Time (sec)
1	30.60	37.97	12	58.93	83.50
2	48.43	51.57	13	54.47	38.30
3	60.77	56.67	14	43.53	51.37
4	36.07	40.47	15	37.93	29.33
5	68.47	49.00	16	43.50	54.27
6	32.43	43.23	17	87.70	62.73
7	43.70	44.57	18	53.53	58.00
8	37.10	28.40	19	64.30	52.40
9	31.17	28.23	20	47.37	53.63
10	51.23	68.47	21	53.67	47.00
11	65.40	51.10			

All the calculations are done on the differences! The differences are calculated as unscented - scented.

$\overline{X}_d = .957,\ s_d = 12.548,\ n = 21$ $\qquad$ $\overline{X}_d \pm t_{.975}\dfrac{s_d}{\sqrt{n}}$

The 95% confidence interval for μ_d is $(-4.755, 6.668)$. Now interpret it.

Example 2

The design of controls and instruments affects how easily people can use them. A student project investigated this effect by asking 25 right-handed students to turn a knob (with their right hands) that moved an indicator by screw action. There were two identical instruments, one with a right-hand thread (the knob turns clockwise) and the other with a left-hand thread (the knob must be turned counterclockwise). The table below gives the times in seconds each subject took to move the indicator a fixed distance. Find a 90% confidence interval for μ_d = average difference in time using right-hand and left-hand threads. Interpret your interval estimate.

Subject	Right thread	Left thread	Subject	Right thread	Left thread
1	113	137	14	107	87
2	105	105	15	118	166
3	130	133	16	103	146
4	101	108	17	111	123
5	138	115	18	104	135
6	118	170	19	111	112
7	87	103	20	89	93
8	116	145	21	78	76
9	75	78	22	100	116
10	96	107	23	89	78
11	122	84	24	85	101
12	103	148	25	88	123
13	116	147			

The differences are calculated as right - left.

$\overline{X}_d = -13.32,\ s_d = 22.936,\ n = 25$ $\qquad$ $\overline{X}_d \pm t_{.95}\dfrac{s_d}{\sqrt{n}}$

The 90% confidence interval for μ_d is $(-21.17, -5.472)$. Now interpret it.