Lesson 39: How do we deal with pared data?

Paired Samples and Blocks

Sometimes the data comes to us paired

Two sets of data are not independent
- data collected before + after a treatment is applied
- races run in pairs - speed skating

Blocking: pairs arise from an experiment

Matching: pairs arise from observational study

We have a problem!

**Two sample t-methods are not valid without independent groups

How do we deal with paired data?

* We look at pairwise differences

treat the differences of paired values as the data

now we have one set of data

use one sample t-test (called paired t-test)

pairs = sample size

Paired t-test

Conditions:

Independence (differences are mutually independent)

Paired data assumption (*you can't just pair data because you have two equal samples)

Randomization

Nearly Normal Condition

THE PAIRED t-TEST

When the conditions are met, we are ready to test whether the mean of paired differences is significantly different from zero. We test the hypothesis

$$H_0\colon \mu_d = \Delta_0,$$

where the d's are the pairwise differences and Δ_0 is almost always 0.

We use the statistic

$$t_{n-1} = \frac{\bar{d} - \Delta_0}{SE(\bar{d})},$$

where $\bar{d}$ is the mean of the pairwise differences, n is the number of *pairs*, and

$$SE(\bar{d}) = \frac{s_d}{\sqrt{n}}.$$

$SE(\bar{d})$ is the ordinary standard error for the mean, applied to the differences.

When the conditions are met and the null hypothesis is true, we can model the sampling distribution of this statistic with a Student's t-model with $n-1$ degrees of freedom, and use that model to obtain a P-value.

PAIRED t-INTERVAL

When the conditions are met, we are ready to find the confidence interval for the mean of the paired differences. The confidence interval is

$$\bar{d} \pm t^*_{n-1} \times SE(\bar{d}),$$

where the standard error of the mean difference is $SE(\bar{d}) = \frac{s_d}{\sqrt{n}}$.

The critical value t^* from the Student's t-model depends on the particular confidence level, C, that you specify and on the degrees of freedom, $n-1$, which is based on the number of pairs, n.

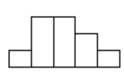
One indicator of physical fitness is resting pulse rate. Ten men volunteered to test an exercise device advertised on television by using it three times a week for 20 minutes. Their resting pulse rates (beats per minute) were measured before the test began, and then again after six weeks. Results are shown in the table. Is there evidence that this kind of exercise can reduce resting pulse rates? How much?

	Pulse rates (beats/min)	
Subject	**Before**	**After**
Allen	73	73
Brandon	83	79
Carlos	85	81
David	87	86
Edwin	91	87
Franco	99	91
Graeme	87	84
Hans	85	83
Ivan	83	84
Jorge	79	76

Before and after measurements are not independent

Hypotheses. $H_0: \mu_d = 0 \quad H_A: \mu_d < 0$ (because we're interested in a *decrease*)

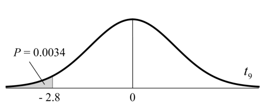
Model. Although the results are independent from one subject to the next, the measurements are paired by subject. A histogram of the paired differences is approximately unimodal and symmetric. Okay to do a matched pairs t-test.



Mechanics. $n = 10$, $df = 9$, $\bar{d} = -2.8$, $s_d = 2.53$

$$t = \frac{(-2.8) - 0}{2.53/\sqrt{10}} = -3.5$$

$$P = P(t_9 < -3.5) = 0.0034$$



Conclusion. With a P-value this low, we reject the null hypothesis. There is strong evidence that this form of exercise can reduce resting pulse rates.

Confidence Interval. The interval looks just like those we saw in Chapter 23; only the notation changes to represent the differences. We choose 95% as our confidence level.

$$\bar{d} \pm t^*_9 \cdot SE(\bar{d}) = -2.8 \pm 2.262 \frac{2.53}{\sqrt{10}} = (-4.61, -0.99)$$

We are 95% confident that six weeks of this exercise program can produce a mean decrease in resting pulse rate of between 1 and 4.6 beats per minute.