STAT 22000 Lecture Slides
Analysis of Paired Data

Yibi Huang
Department of Statistics
University of Chicago

Coverage

This set of slides covers Section 5.2 in the 3rd edition of *OpenIntro Statistics* (or Section 7.2 in the 4th edition).

Example: Coffee & Blood Flow During Exercise

Doctors studying healthy men measured myocardial blood flow (MBF)[1] during bicycle exercise after giving the subjects a placebo or a dose of 200 mg of caffeine that was equivalent to drinking two cups of coffee[2].

There were 8 subjects, each was tested twice, 4 of them were randomly selected to receive caffeine in the first test and placebo in the second test; the other 4 received placebo first and caffeine second.

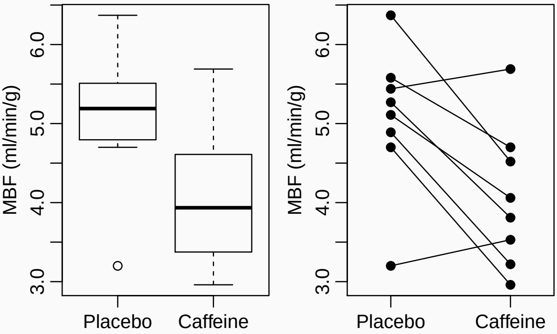
There was a 24-hour gap between the two tests (washout period).

[1] MBF was measured by taking positron emission tomography (PET) images after oxygen-15 labeled water was infused in the patients.

[2] Namdar et. al (2006). Caffeine decreases exercise-induced myocardial flow reserve. *Journal of the American College of Cardiology* **47**, 405-410.

Data for the Coffee & Blood Flow Experiment

	MBF (ml/min/g)	
Subject	Placebo	Caffeine
1	6.37	4.52
2	5.44	5.69
3	5.58	4.70
4	5.27	3.81
5	5.11	4.06
6	4.89	3.22
7	4.70	2.96
8	3.20	3.53
Mean	5.07	4.06
SD	0.91	0.89



Discussion

- Why did 4 subjects caffeine first and placebo second and the other 4 received placebo first and caffeine second?
- Why do we need a washout period (the 24 hour gap) between the two tests?
- Can we analyze the data of the experiment like two independent samples?

Hypothesis Tests for Paired Data

- Paired data cannot be analyzed like 2-sample data since the 2 measurements on the same subject are *dependent*.
- Nonetheless, if measurements on different pairs can be reasonably assumed independent, we can take differences of the two measurements within each pair and analyze the differences like **one-sample data**.

To test H_0: $\mu_1 = \mu_2$, the test statistic is

$$t = \frac{\bar{d}}{s_d / \sqrt{n}} \sim t_{n-1}$$

where

$$\bar{d} = \text{sample mean of the diffs}$$
$$s_d = \text{sample SD of the diffs}$$
$$n = \#\ \text{of } \textbf{pairs.}$$

	MBF (ml/min/g)		
Subject	Placebo	Caffeine	Diff
1	6.37	4.52	1.85
2	5.44	5.69	−0.25
3	5.58	4.70	0.88
4	5.27	3.81	1.46
5	5.11	4.06	1.05
6	4.89	3.22	1.67
7	4.70	2.96	1.74
8	3.20	3.53	−0.33
Mean	5.07	4.06	*1.01*
SD	0.91	0.89	*0.87*

Example: Coffee & Blood Flow During Exercise

	MBF (ml/min/g)		
Subject	Placebo	Caffeine	Diff.
1	6.37	4.52	1.85
2	5.44	5.69	−0.25
3	5.58	4.70	0.88
4	5.27	3.81	1.46
5	5.11	4.06	1.05
6	4.89	3.22	1.67
7	4.70	2.96	1.74
8	3.20	3.53	−0.33
Mean	5.07	4.06	*1.01*
SD	0.91	0.89	*0.87*

In this example, $\bar{d} = 1.01$, $s_d = 0.87$. Please note that

$$\bar{d} = \bar{x}_{\text{placebo}} - \bar{x}_{\text{caffeine}}$$
$$1.01 = 5.07 - 4.06$$

but

$$s_d \neq s_{\text{placebo}} - s_{\text{caffeine}}$$
$$0.87 \neq 0.91 - 0.89$$

s_d is the sample SD of the 8 differences:

$$1.85, -0.25, 0.88, 1.46, 1.05, 1.67, 1.74, -0.33$$

Example: Coffee & Blood Flow During Exercise

	MBF (ml/min/g)		
Subject	Placebo	Caffeine	Difference
1	6.37	4.52	1.85
2	5.44	5.69	−0.25
3	5.58	4.70	0.88
4	5.27	3.81	1.46
5	5.11	4.06	1.05
6	4.89	3.22	1.67
7	4.70	2.96	1.74
8	3.20	3.53	−0.33
Mean	5.07	4.06	*1.01*
SD	0.91	0.89	*0.87*

In this example,

$\bar{d} = 1.01$, $s_d = 0.87$,

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = \frac{1.01}{0.87/\sqrt{8}} \approx 3.28$$

with $8 - 1 = 7$ degrees of freedom.

The two-sided P-value can be found in R to be ≈ 0.0278.

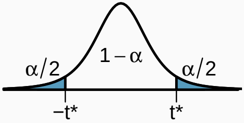
```
> 2*pt(3.28, df=7, lower.tail=F)
[1] 0.01348706
```

Confidence Intervals for the Mean Difference in Paired Data

The $100(1-\alpha)\%$ confidence interval for the difference is

$$\bar{d} \pm t^* s_d / \sqrt{n}$$

where t^* is the value for the t distribution with $n-1$ degrees of

freedom such that 

For the coffee experiment, the t^* for a 95% CI is $t^* \approx 2.3646$.

```
> qt(0.05/2, df=7, lower.tail=F)
[1] 2.364624
```

So the 95% CI for the mean difference is

$$\bar{d} \pm t^* \frac{s_d}{\sqrt{n}} = 1.01 \pm 2.3646 \times \frac{0.87}{\sqrt{8}} \approx 1.01 \pm 0.73 = (0.28, 1.74).$$

Tests/CIs for Paired Data in R

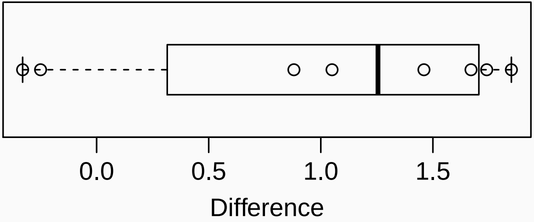
```
> caffeine = c(4.52, 5.69, 4.70, 3.81, 4.06, 3.22, 2.96, 3.53)
> placebo = c(6.37, 5.44, 5.58, 5.27, 5.11, 4.89, 4.70, 3.20)
> t.test(placebo,caffeine, paired=T, conf.level=0.95)

Paired t-test

data:  placebo and caffeine
t = 3.2857, df = 7, p-value = 0.01338
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2827867 1.7347133
sample estimates:
mean of the differences
                1.00875
```

Checking Conditions for Paired Data

As the inference problem for paired data is simply one-sample problem on the difference within each pair, we need to make sure that

- the differences are independent
- the distribution (histogram) of the differences is not too skewed and has no outlier

Whether the distributions of the two groups are skewed or have outlier(s) do not matter.



Exercise 5.18. Paired or Not

In each of the following scenarios, determine if the data are paired?

1. We would like to know if Intel's stock and Southwest Airlines' stock have similar rates of return. To find out, we take a random sample of 50 days, and record Intel's and Southwest's stock on those same days. $\Rightarrow$ paired
2. We randomly sample 50 items from Target stores and note the price for each. Then we visit Walmart and collect the price for each of those same 50 items. $\Rightarrow$ paired
3. A school board would like to determine whether there is a difference in average SAT scores for students at one high school versus another high school in the district. To check, they take a simple random sample of 100 students from each high school. $\Rightarrow$ not paired

Benefits of Paired Designs

Experimenters have come up with all kinds of clever ways to use pairing to cut down on variability:

- Crossover studies (same subjects are reused)
- Twin studies
- If subjects cannot be reused and twins are not available, some studies try to pair subjects with similar age, sex, weight or other important risk factors

As variability (noise) goes down,

- confidence intervals become shorter
- hypothesis tests become more powerful (smaller p values)

If paired data were analyzed like 2-sample data

	MBF (ml/min/g)		
subject	placebo	caffeine	diff.
1	6.37	4.52	1.85
2	5.44	5.69	−0.25
3	5.58	4.70	0.88
4	5.27	3.81	1.46
5	5.11	4.06	1.05
6	4.89	3.22	1.67
7	4.70	2.96	1.74
8	3.20	3.53	−0.33
Mean	5.07	4.06	1.01
SD	0.91	0.89	0.87

If we ignore pairing, and analyze the caffeine data as two-sample data, the two-sample t-statistic

$$\frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{n}}} = \frac{5.07 - 4.06}{\sqrt{\frac{0.91^2}{8} + \frac{0.89^2}{8}}} \approx 2.244$$

would be less than the paired t-statistic

$$\frac{\bar{d}}{s_d/\sqrt{n}} = \frac{1.01}{0.87/\sqrt{8}} \approx 3.28,$$

The p-value (6%) given by a two-sample t test is larger than the one given by a paired t-test (1.3%), less significant.

95% two-sample CI: $5.07 - 4.06 \pm 2.36\sqrt{\frac{0.91^2}{8} + \frac{0.89^2}{8}} \approx 1.01 \pm 1.06$

95% paired CI: $1.01 \pm 2.36 \times 0.87/\sqrt{8} \approx 1.01 \pm 0.73$ (shorter)