STAT 22000 Lecture Slides
Analysis of Paired Data

Yibi Huang
Department of Statistics
University of Chicago

Outline

This set of slides covers Section 5.2 in the text.

- analysis of paired data (5.2)

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.

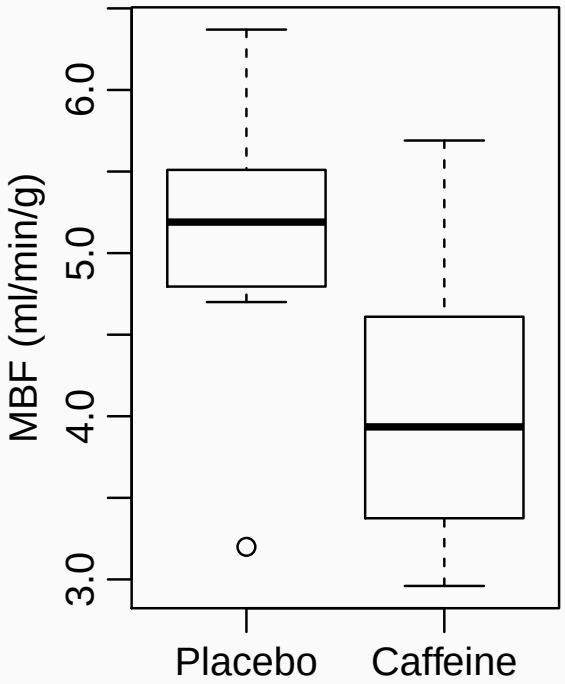
There was a 24-hour gap between the two tests (washout period).

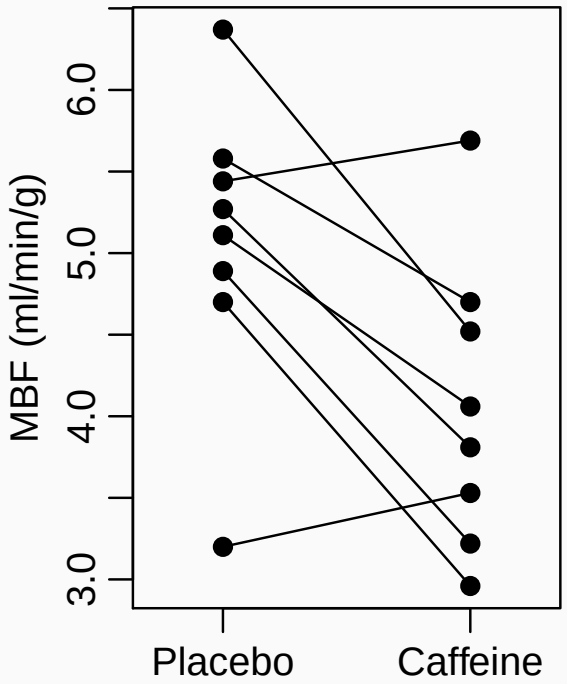
[1] MBF was measured by taking positron emission tomography (PET) images after oxygen-15 labeled water was infused in the patients.

[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





- 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?

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 we can safely assume measurements on different pairs to be 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

  - $\bar{d}$ is the sample mean of the differences
  - s_d is the sample SD of the differences, and
  - n is the number of **pairs.**

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.

From the *t*-table, we can see the 2-sided *P*-value is between 0.01 and 0.02.

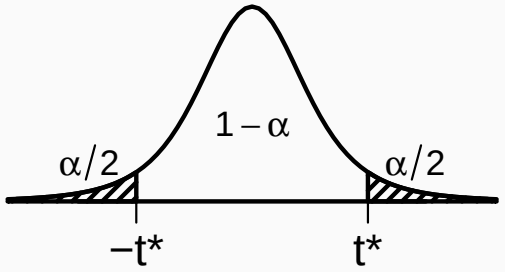
one tail	0.1	0.05	0.025	0.01	0.005
two tails	0.2	0.10	0.050	*0.02*	*0.010*
df 7	1.41	1.89	2.36	*3.00*	*3.50*

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 $t^* = 2.36$.

one tail	0.1	0.05	0.025	0.01	0.005
two tails	0.2	0.10	*0.050*	0.02	0.010
df 7	1.41	1.89	*2.36*	3.00	3.50

So the 95% CI for the mean difference is

$$\bar{d} \pm t^* \frac{s_d}{\sqrt{n}} = 1.01 \pm 2.36 \times \frac{0.87}{\sqrt{8}} \approx 1.01 \pm 0.73 = (0.28, 1.74).$$

Tests/CIs for Paired Data in R

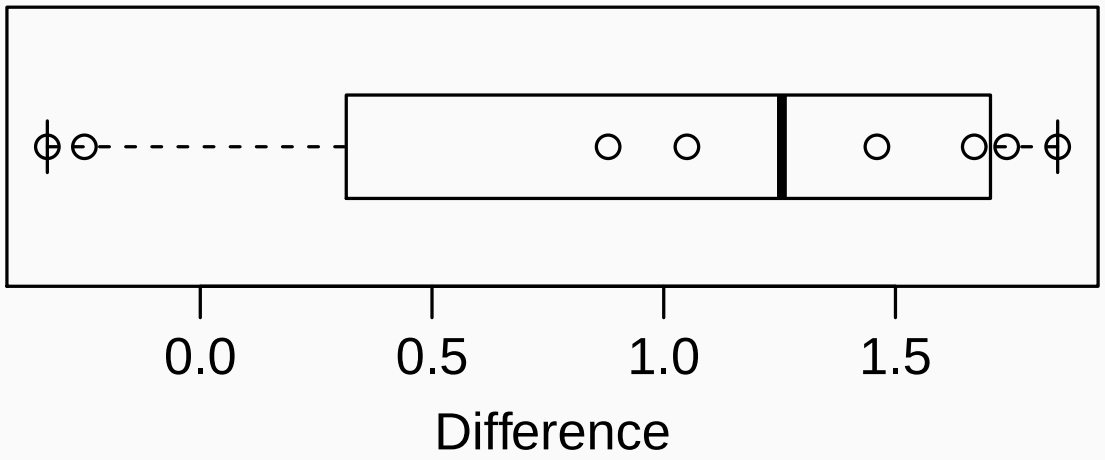
```
> caffeine = c(4.52, 5.69, 4.70, 3.81, 4.06, 3.22, 2.96, 3.53)
> placebo = c(6.37, 5.44, 5.58, 5.27, 5.11, 4.89, 4.70, 3.20)
> t.test(placebo,caffeine, paired=T, conf.level=0.95)

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.



Two-Sample Design v.s. Paired Design

- Suppose we are conducting a study investigating whether sunblock A is better than sunblock B at preventing sunburns
- The first design that comes to mind is probably to randomly assign sunblock A to one group and sunblock B to a different group
- There is nothing wrong with this design, but we can do better

Signal and Noise

$$\text{Response} = \begin{cases} \mu_A + \text{noise} & \text{if using sunblock A} \\ \mu_B + \text{noise} & \text{if using sunblock B} \end{cases}$$

- *Signal* is the magnitude of the difference between the two groups — which is $\mu_A - \mu_B$ in the present context, how much better one sunblock is than the other
- *Noise* is the variability present in the outcome from all other sources besides the one you are interested in — in the sunblock experiment, this would include factors like how sunny the day was, how much time the person spent outside, how easily the person burns, etc.
- Hypothesis tests depend on the ratio of *signal* to *noise* — how easily we can distinguish the treatment effect $\mu_A - \mu_B$ from all other sources of variability

Signal-to-Noise Ratio

- To get a larger signal-to-noise ratio, we must either increase the signal or reduce the variability
- The signal is usually determined by nature and out of our control
- Instead, we are going to have to reduce the variability/noise
- If our sunblock experiment were controlled, we could attempt such steps as forcing all participants to spend an equal amount of time outside, on the same day, in an equally sunny area, etc.

Subject-to-Subject Variability

$$\begin{aligned}\text{Response} &= \text{treatment effect} + \text{noise} \\ &= \text{treatment effect} + \text{subject effect} + \text{other noise}\end{aligned}$$

- But we cannot control subject-to-subject variability (how easily certain people burn)?
- A powerful technique for reducing subject-to-subject variability is *pairing*
- For each person, we can apply sunblock A (at random) to one of their arms, and sunblock B to the other arm, and as an outcome, look at the difference between the two arms
- In this experiment, the thing that we randomize is which arm receiving A or B.

Removing Subject-to-Subject Variability by Pairing

In a paired design, we have two measurements ($i = A, B$) on every subject ($j = 1, 2, \ldots, n$), one for each treatment

$$\begin{aligned} \text{Response}_{ij} &= \mu_i + \text{noise}_{ij} \\ &= \mu_i + \text{subject effect}_j + \text{other noise}_{ij} \end{aligned}$$

The subject effect can be removed by taking the difference between the two measurements on the same individual.

$$\text{Response}_{Aj} - \text{Response}_{Bj} = \mu_A - \mu_B + \text{other noise}_{Aj} - \text{other noise}_{Bj}$$

When person-to-person variation (subject effect) is considerable, paired designs can substantially reduce the size of noise.

Benefits of Paired Designs

As variability (noise) goes down,

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

Pairing in Observational Studies

- Experimenters have come up with all kinds of clever ways to use pairing to cut down on variability:
  - Crossover studies
  - Twin studies
- Pairing is also widely used in observational studies, e.g.,
  - case-control study: each subject having a certain disease (case) is paired with a subject w/o the disease but similar in some variables such as age, sex, or race (control), then analyze the difference between the pairs
- In addition to increasing power, pairing in observational studies also eliminates (some of the) potential confounding variables

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

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

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)