Chapter 8 Paired observations

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 205: Elementary Statistics for the Biological and Life Sciences

Book review of two-sample t-test ingredients

t Test

$$H_0\colon \mu_1 = \mu_2$$
$$H_A\colon \mu_1 \neq \mu_2 \text{ (nondirectional)}$$
$$H_A\colon \mu_1 < \mu_2 \text{ (directional)}$$
$$H_A\colon \mu_1 > \mu_2 \text{ (directional)}$$

$$\text{Test statistic: } t_s = \frac{(\bar{y}_1 - \bar{y}_2) - 0}{\mathrm{SE}_{(\bar{Y}_1 - \bar{Y}_2)}}$$

$$P\text{-value} = \text{tail area under Student's } t \text{ curve with}$$

$$\mathrm{df} = \frac{(\mathrm{SE}_1^2 + \mathrm{SE}_2^2)^2}{\mathrm{SE}_1^4/(n_1 - 1) + \mathrm{SE}_2^4/(n_2 - 1)}$$

Nondirectional H_A: P-value $=$ two-tailed area beyond t_s and $-t_s$

Directional H_A: Step 1. Check directionality.

Step 2. P-value $=$ single-tail area beyond t_s

Decision: Significant evidence for H_A if P-value $\leq \alpha$

Paired designs

- Paired data arise when two of the same measurements are taken from the same subject, but under different experimental conditions.
- Subjects often receive both a treatment Y_1 and a control Y_2.
- Pairing observations reduces the subject-to-subject variability in the response.
- The analysis focuses on *the difference* in response from treatment to control. Let μ_D be the mean difference for the entire population.
- We want a confidence interval for μ_D and will want to test $H_0 : \mu_D = 0$ vs. one of (a) $H_A : \mu_D \neq 0$, (b) $H_A : \mu_D < 0$, or (c) $H_A : \mu_D > 0$.

Example 8.1.1 Coffee and blood flow

- Doctors studying healthy subjects measured myocardial blood flow (MBF) (ml/min/g) during bicycle exercise before and after giving the subjects the equivalent of two cups of coffee (200 mg of caffeine).
- Some people have high blood flow both before and after caffeine. Others have low blood flow before and after.
- By focusing on *the differences* from the same individual before and after, we *adjust* for individuals tendancy to be high or low regardless.
- How does this analyis differ from those in Chapters 6 and 7? Observations are collected *on the same person*.

Example 8.1.1 blood flow data

Table 8.2.1 Myocardial blood flow (ml/min/g) for eight subjects

	MBF		
Subject	Baseline y_1	Caffeine y_2	Difference $d = y_1 - y_2$
1	6.37	4.52	1.85
2	5.69	5.44	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	3.53	3.20	0.33
Mean	5.14	3.99	1.15
SD	0.83	0.86	0.63

Example 8.1.1 blood flow data

Each subject has a connected line (control and treatment). What does caffeine do to bloodflow?

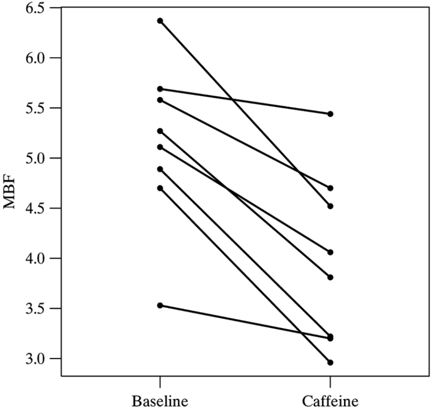


Figure 8.1.1 Dotplots of MBF readings before and after caffeine consumption, with line segments connecting readings on each subject

Paired analysis in R

- Null is $H_0 : \mu_D = 0$.
- t.test(sample1,sample2,paired=TRUE) gives P-value for $H_A : \mu_D \neq 0$.
- t.test(sample1,sample2,paired=TRUE,alternative="less") gives P-value for $H_A : \mu_D < 0$.
- t.test(sample1,sample2,paired=TRUE,alternative="greater") gives P-value for $H_A : \mu_D > 0$.

R code for bloodflow data

```
> baseline=c(6.37,5.69,5.58,5.27,5.11,4.89,4.70,3.53)
> caffeine=c(4.52,5.44,4.70,3.81,4.06,3.22,2.96,3.20)
> t.test(baseline,caffeine,paired=TRUE)

        Paired t-test

data:  baseline and caffeine
t = 5.1878, df = 7, p-value = 0.00127
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.6278643 1.6796357
sample estimates:
mean of the differences
                1.15375
```

We estimate μ_D as 1.15 ml/min/g. We are 95% confident that the true mean bloodflow is between 0.63 and 1.68 ml/min/g greater in the control group. We reject $H_0 : \mu_D = 0$ at the 5% level because P-value $= 0.0013 < 0.05$. Caffeine significantly reduces bloodflow.

Validity of paired t-test (p. 306)

- Let n be the number of paired observations.
- The paired sample t-test and confidence interval are valid if (a) The sample size is large enough, $n > 30$, say, or (b) the *differences* are approximately normal.
- Normality can be checked with a normal probability plot.
- If the two samples are `sample1` and `sample2`, type `qqnorm(sample1-sample2)` in R.

Example 8.2.4 Hunger rating

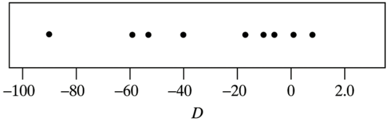
- During a weight loss study each of $n = 9$ subjects was given either the active drug m-chlorophenylpiperazine (mCPP) for two weeks and then a placebo for another two weeks, or else was given the placebo for the first two weeks and then mCPP for the second two weeks.
- As part of the study the subjects were asked to rate how hungry they were at the end of each two-week period.

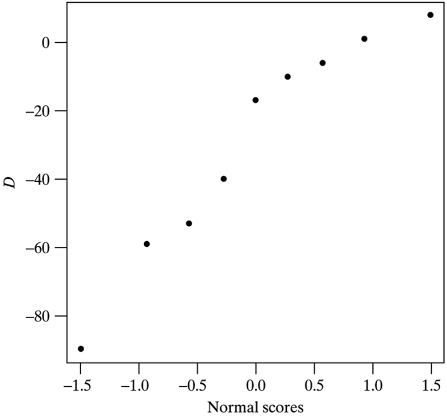
Hunger rating data

Table 8.2.2 Hunger Rating for Nine Women

Subject	Hunger rating: Drug (mCPP) y_1	Placebo y_2	Difference $d = y_1 - y_2$
1	79	78	1
2	48	54	−6
3	52	142	−90
4	15	25	−10
5	61	101	−40
6	107	99	8
7	77	94	−17
8	54	107	−53
9	5	64	−59
Mean	55	85	−30
SD	32	34	33

Hunger rating dotplot & normal probability plot





R code for hunger rating

```
> drug=c(79,48,52,15,61,107,77,54,5)
> placebo=c(78,54,142,25,101,99,94,107,64)
> qqnorm(drug-placebo)
> t.test(drug,placebo,paired=TRUE)

        Paired t-test

data:  drug and placebo
t = -2.7014, df = 8, p-value = 0.02701
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -54.784709  -4.326402
sample estimates:
mean of the differences
              -29.55556
```

We estimate μ_D as -30. We are 95% confident that the drug reduces hunger between 4 and 55 points. We reject $H_0 : \mu_D = 0$ at the 5% level because P-value $= 0.027 < 0.05$. The drug significantly reduces hunger.

8.3 Paired designs

Paired analyses reduce variability and make it easier to reject $H_0 : \mu_D = 0$. Need to have the paired observations come from very similar experimental units.

Examples:

- Ex. 8.3.1 Two plants grown in the same container.
- Ex. 8.3.2 Case-control data from people matched on gender, age.
- Ex. 8.3.3 Tryglycerides measured before and after exercise.

Example 8.3.4 Triglycerides and exercise

Triglycerides play a role in coronary artery disease. Researchers measured blood triglycerides in seven men before and after a 10-week exercise program.

Subject	Before	After
1	0.87	0.57
2	1.13	1.03
3	3.14	1.47
4	2.14	1.43
5	2.98	1.20
6	1.18	1.09
7	1.60	1.51

8.4 The sign test

- The paired t-test assumes that differences follow a normal distribution.
- If the data aren't normal and the sample size is small, e.g. $n < 30$, then you can use the *sign test*.
- The sign test focuses on the median difference η_D rather than the mean μ_D.
- This test looks at the number of differences $D = Y_1 - Y_2$ that are positive N_+ and the number that are negative N_-. These numbers should be similar if $H_0 : \eta_D = 0$ is true.
- A P-value is based on the binomial distribution. Under $H_0 : \eta_D = 0$, $N_+ \sim \text{bin}(n, 0.5)$.

Sign test in R

- In R, binom.test(N_+,n) tests $H_0 : \eta_D = 0$ vs. $H_A : \eta_D \neq 0$.
- Need to count the number of $+$'s and put that as first number, second number is sample size.
- For $H_A : \eta_D < 0$ use binom.test(N_+,n,alternative="less").
- For $H_A : \eta_D > 0$ use binom.test(N_+,n,alternative="greater").
- Ignore all output except the P-value.

Example 8.3.4 Triglycerides and exercise

Subject	Before	After	Sign
1	0.87	0.57	$+$
2	1.13	1.03	$+$
3	3.14	1.47	$+$
4	2.14	1.43	$+$
5	2.98	1.20	$+$
6	1.18	1.09	$+$
7	1.60	1.51	$+$

$N_+ = 7$ and $N_- = 0$; P-value should be small.

```
> binom.test(7,7)
number of successes = 7, number of trials = 7, p-value = 0.01563
```

Two more examples

Hunger rating

```
> binom.test(2,9)
number of successes = 2, number of trials = 9, p-value = 0.1797
```

P-value from t-test is 0.02701; not close at all. The t-test has greater power to reject H_0 when data are really normal.

Caffeine and blood flow

```
> binom.test(8,8)
number of successes = 8, number of trials = 8, p-value = 0.007812
```

P-value from t-test is 0.00127; fairly similar but t-test has smaller P-value (more power if differences really are normal).