Hypothesis Testing: Two-Sample Inference

Chapter 8

Outline

1. Introduction
2. Matched data/ paired sample
3. Two independent groups
   a) Two sample t test with equal variances
   b) Testing for the equality of two variances
   c) Two sample t test with unequal variances
4. Power analysis and Sample-size determination

Introduction

- In a two-sample hypothesis-testing problem, the underlying parameters of two different populations, neither of whose values is assumed known, are compared.

- Let's say we are interested in the relationship between oral contraceptive (OC) use and blood pressure in women.

- Two different experimental designs can be used to assess this relationship.

Longitudinal Study

1. Identify a group of nonpregnant, premenopausal women of childbearing age (16–49 years) who are not currently OC users, and measure their blood pressure, which will be called the *baseline blood pressure*.

2. Rescreen these women 1 year later to ascertain a subgroup who have remained nonpregnant throughout the year and have become OC users. This subgroup is the study population.

3. Measure the blood pressure of the study population at the follow-up visit.

4. Compare the baseline and follow-up blood pressure of the women in the study population to determine the difference between blood pressure levels of women when they *were* using the pill at follow-up and when they *were not* using the pill at baseline.

Cross-Sectional Study

1. Identify both a group of OC users and a group of non-OC users among nonpregnant, premenopausal women of childbearing age (16–49 years), and measure their blood pressure.
2. Compare the blood pressure level between the OC users and nonusers.

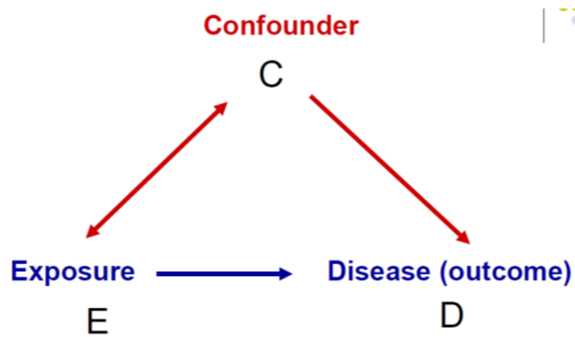
- Two different experimental designs
  - longitudinal Study (or follow-up study): the same group of people is followed over time.

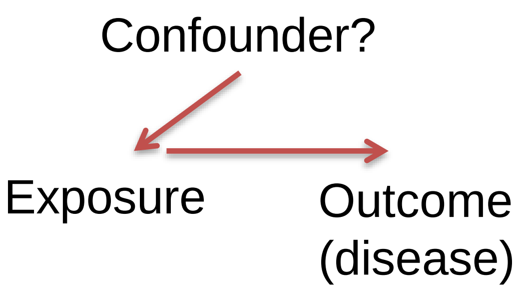
    $\Longrightarrow$ paired sample
  - cross-sectional study: the participants are seen at only one point in time.

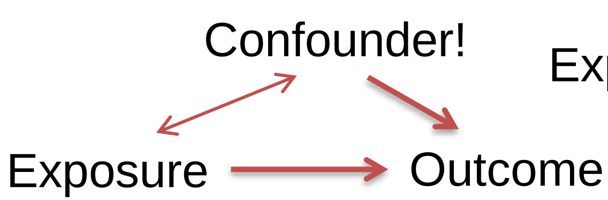
    $\Longrightarrow$ independent sample

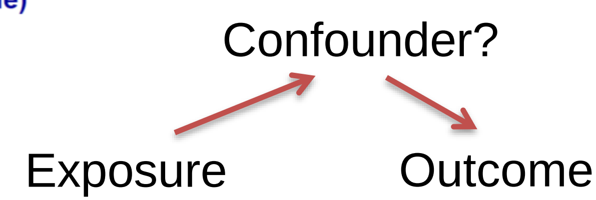
Comparison

- Confounders
  - For example, OC users are known to weigh less than non-OC users. Low weight tends to be associated with low BP, so the blood-pressure levels of OC users as a group would appear lower than the levels of non-OC users.









- finance
  - a follow-up study is more expensive than a cross-sectional study

Matched data/ paired sample

The paired t test

Example recalled

- Suppose in the previous example the paired-sample study design is adopted and the sample data are obtained.

SBP levels (mm Hg) in 10 women while not using (baseline) and while using (follow-up) OCs

i	SBP level while not using OCs (x_{i1})	SBP level while using OCs (x_{i2})	d_i^*
1	115	128	13
2	112	115	3
3	107	106	–1
4	119	128	9
5	115	122	7
6	138	145	7
7	126	132	6
8	105	109	4
9	104	102	–2
10	115	117	2

The paired t test

- When dealing a single set of paired data, one strategy is to take the difference between the paired observation

$$d = X_1 - X_2$$

and do a one-sample t test of

$$H_0: \Delta = 0 \text{ vs. } H_1: \Delta \neq 0 \text{ (or >0, <0)}$$

- Under the assumption $Z \sim N(\Delta, \sigma^2)$, the test statistic

$$t = \frac{\bar{d}}{s/\sqrt{n}}$$

where $\bar{d}$ is the mean difference and s is the sample SD.

example

- Is there any evidence that blood pressure in women has relationship with oral contraceptive (OC) use?

```
> sbp.y <- c(128,115,106,128,122,145,132,109,102,117)
> sbp.n <- c(115,112,107,119,115,138,126,105,104,115)
> t.test(sbp.y,sbp.n, paired = T)

Paired t-test

data: sbp.n and sbp.y
t = 3.3247, df = 9, p-value = 0.008874
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
  1.533987  8.066013
sample estimates:
mean of the differences
                4.8
```

Two independent groups

Example

- Suppose a sample of eight 35- to 39-year-old non-prenant, premenopausal OC users who work in a company and have a mean systolic blood pressure (SBP) of 132.86 mm Hg and sample standard deviation of 15.34 mm Hg are identified.

- A sample of 21 non-pregnant, premenopausal, non-OC users in the same age group are similarly identified who have mean SBP of 127.44 mm Hg and sample standard deviation of 18.23 mm Hg.

- What can be said about the underlying mean difference in blood pressure between the two groups?

Two-Sample t Test

- H_0: $\mu_1 = \mu_2$, versus H_1: $\mu_1 \neq \mu_2$ (or >, <)
- Assuming a common error variance, we have

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S\sqrt{1/n_1 + 1/n_2}}$$

where S^2 is the pooled estimate of the variance,

$$S^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- The statistic t follows a $t_{n_1+n_2-2}$ distribution under the null hypothesis

Example

- The common variance is first estimated:

$$s^2 = \frac{7(15.34)^2 + 20(18.23)^2}{27} = \frac{8293.9}{27} = 307.18$$

- The following test statistic is then computed:

$$t = \frac{132.86 - 127.44}{17.527\sqrt{1/8 + 1/21}} = 0.74$$

- The p-value is .46.

- If the SBP example given before is cross-sectional design study rather than a paired-sample study, the same data gives different results.
- See next slide:

Paired-data example

```
> t.test(sbp.y,sbp.n, alt = "two.sided",var.equal = T)

Two Sample t-test

data: sbp.y and sbp.n
t = 0.9052, df = 18, p-value = 0.3773
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.340831 15.940831
sample estimates:
mean of x mean of y
        120.4 115.6
```

- significant differences could be detected in the pair t test, in contrast to the non-significant results that were obtained using the cross-sectional design here.
- Thus the longitudinal design is usually more efficient because it uses people as their own controls.

Testing for the equality of two variances

- What about when the assumption that the underlying variances of the two samples were the same doesn't hold? $(\text{i.e.}, s_1^2/s_2^2 \ > 5)$

- Testing for the equality of two variances

- Test statistic

$$F = s_1^2/s_2^2$$

follows a F_{n_1-1,n_2-1} distribution under the null hypothesis

```
> var.test(sbp.y,sbp.n)
```

Notes

- These tests, employing the F distribution, rely heavily on normality assumptions.

- Alternatively,
  - for moderate sample sizes, consider creating a bootstrap confidence interval for the ratio of two variances.
  - For smaller sample sizes, consider exploratory data analysis

Two-Sample *t* Test

- If assuming a common error variance is questionable, then use the statistic
$$\frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 \; + S_2^2/n_2}}$$
- Which follows a standard normal distribution for large n_1 and n_2.
- It follows a t distribution if the $X_{1,i}$ and $X_{2,i}$ are normally distributed
- The approximate degrees of freedom are (Satterthwaite)
$$\frac{(S_1^2/n_1 \; + S_2^2/n_2)^2}{\left(S_1^2/n_1\right)^2/(n_1 - 1) + \left(S_2^2/n_2\right)^2/(n_2 - 1)}$$

```
>t.test(sbp.y,sbp.n, alt = "two.sided")
```

Power analysis and sample-size determination

Power analysis

- For paired-data, the power calculation in two sample case is similar to those in one-sample case.
  - Based on t test

```
> power.t.test(d=d,power=0.8,type="paired",alt = "two.sided")
Paired t test power calculation
              n = 153.5979
          delta = 0.2274885
             sd = 1
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

  - Based on normal test (approximate by z statistic)

```
> pwr.norm.test(d=d,power=0.8,alt="two.sided")
Mean power calculation for normal distribution with known variance
              d = 0.2274885
              n = 151.6658
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```

Sample-Size Determination

- For two-sample test, suppose we know σ_1^2 and σ_2^2 and equal size of samples will be recruited.

- To conduct a two-sided test with significance level α and power of 1 – β
  - Equal sample size for *each* group :

$$n_1 = n_2 = \frac{\left(\sigma_1^2 + \sigma_2^2\right)\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\Delta^2}$$

where $\Delta = |\mu_2 - \mu_1|$.

  - Unequal sample size, i.e., $n_2 = kn_1$:

$$n_2 = \frac{\left(k\sigma_1^2 + \sigma_2^2\right)\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\Delta^2}$$

- Replacing z percentiles with t percentiles leads to implicit expressions of sample size n.

R codes

```
> pwr.norm.test(d=d,n=100,alt="two.sided")

Mean power calculation for normal distribution with known variance

d = 0.2274885
n = 100
sig.level = 0.05
power = 0.6236008
alternative = two.sided

> pwr.t.test(d=d,n=100,alt="two.sided")

Two-sample t test power calculation

           n = 100
           d = 0.2274885
   sig.level = 0.05
       power = 0.3599158
 alternative = two.sided
NOTE: n is number in *each* group
```