Elementary Statistics Lecture 8

Comparing Two Groups

Chong Ma

Department of Statistics
University of South Carolina

Outline

Confidence Interval for $\mu_1 - \mu_2$

Given two populations $X_1 \sim (\mu_1, \sigma_1)$ and $X_2 \sim (\mu_2, \sigma_2)$, we are interested in making inference about $\mu_1 - \mu_2$.

Confidence Interval for Difference Between Population Means
Assumptions

- A quantitative response variable observed in each of two groups.
- Independent random samples, either from random sampling or a randomized experiment.
- An approximately normal population distribution for each group.(mainly important for small sample sizes and even then the method is robust to violations of this assumption)

Confidence Interval($\sigma_1^2 \neq \sigma_2^2$)

Confidence Interval for Difference Between Population Means
For two samples with sizes n_1 and n_2 and standard deviations s_1 and s_2, a 95% confidence interval for the difference $\mu_1 - \mu_2$ between the two population means is

$$[(\bar{x}_1 - \bar{x}_2) \pm t_{.025}(se)], \text{with } se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Where $t_{.025}$ is the t-score with right-tail probability 0.025 with degrees of freedom df^*. And

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

Remark: Use unpooled se if $\frac{s_1}{s_2} > 2$ or $\frac{s_1}{s_2} < 0.5$.

Confidence Interval($\sigma_1^2 = \sigma_2^2$)

Confidence Interval for Difference Between Population Means
For two samples with sizes n_1 and n_2 and standard deviations s_1 and s_2, a 95% confidence interval for the difference $\mu_1 - \mu_2$ between the two population means is

$$[(\bar{x}_1 - \bar{x}_2) \pm t_{.025}(se)], \text{with } se = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Where $t_{.025}$ is the t-score with right-tail probability 0.025 with degrees of freedom $df = n_1 + n_2 - 2$. And

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

Remark: Use pooled se if $0.5 < \frac{s_1}{s_2} < 2$

Significance Test for $\mu_1 - \mu_2$ ($\sigma_1^2 \neq \sigma_2^2$)

1. **Assumption** (refer to Confidence Interval Assumption)
2. **Hypothesis** $H_0 : \mu_1 = \mu_2$ VS. $H_a : \mu_1 \neq \mu_2$(two-sided)
   (One-sided $H_a : \mu_1 > \mu_2$ or $H_a : \mu_1 < \mu_2$)
3. **Test Statistic**

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{se} \text{ where } se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

4. **P-value** Two-tail probability from t distribution of values even more extreme than observed t test statistic, presuming H_0 is true with degrees of freedom df^*(refer to slides 4).
5. **Conclusion** Reject H_0 if P-value $\leq \alpha$(significance level, usually be 0.05). Smaller P-values give stronger evidence against H_0 and supporting H_a. Interpret the P-value in context.

Significance Test for $\mu_1 - \mu_2$ ($\sigma_1^2 = \sigma_2^2$)

1. **Assumption** (refer to Confidence Interval Assumption)
2. **Hypothesis** $H_0 : \mu_1 = \mu_2$ VS. $H_a : \mu_1 \neq \mu_2$(two-sided)
   (One-sided $H_a : \mu_1 > \mu_2$ or $H_a : \mu_1 < \mu_2$)
3. **Test Statistic**

   $$T = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{se} \text{ where } se = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

   Where $s = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}}$
4. **P-value** Two-tail probability from t distribution of values even more extreme than observed t test statistic, presuming H_0 is true with degrees of freedom $df = n_1 + n_2 - 2$.
5. **Conclusion** Reject H_0 if P-value $\leq \alpha$(significance level, usually be 0.05). Smaller P-values give stronger evidence against H_0 and supporting H_a. Interpret the P-value in context.

Housework for women and men

Do women tend to spend more time on housework than men? If so, how much more? Based on data from the National Survey of Families and Households, one study reported the results in the table for the number of hours spend in housework per week.

Housework Hours			
Gender	**Sample Size**	**Mean**	**Standard Deviation**
Women	476	33.0	21.9
Men	496	19.9	14.6

a. Calculate the population mean more hours that women spend on housework than men.
b. Calculate the standard error for comparing the means.
c. Calculate the 95% confidence interval comparing the population means for women and men.
d. State the assumptions upon which the interval in part c is based.

Housework for women and men

Housework Hours			
Gender	**Sample Size**	**Mean**	**Standard Deviation**
Women	476	33.0	21.9
Men	496	19.9	14.6

a. $\bar{x}_1 - \bar{x}_2 = 33.0 - 19.9 = 13.1$

b. Note $s_1/s_2 = 21.9/14.6 = 1.5 < 2$, consider pooled standard errors. Thus,

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = 18.54, se = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 1.19$$

c. $[(\bar{x}_1 - \bar{x}_2) \pm 1.96se] = [13.1 \pm 1.96 * 1.19] = [10.77, 15.43]$

d. Random, Independent, approximately normal.

Continue...

Housework Hours			
Gender	**Sample Size**	**Mean**	**Standard Deviation**
Women	476	33.0	21.9
Men	496	19.9	14.6

Interpretation: We have 95% confidence that on the average, women spend on housework at least 10.77 hours and at most 15.43 hours more than men. Since 0 is not contained in the 95% confidence interval, we have sufficient evidence to conclude that women spend more hours on housework than men.

Nicotine dependence

A study on nicotine dependence for teenage smokers obtained a random sample of seventh graders. The response variable was constructed from a questionnaire called the Hooked on Nicotine Checklist(HONC). The higher HONC, the more hooked that students is on nicotine. One explanatory variable considered in the study was whether a subject reported inhaling when smoking.

	HONC Score		
Group	**Sample Size**	**Mean**	**Standard Deviation**
Inhalers	237	2.9	3.6
Noninhalers	95	0.1	0.5

a. Calculate the standard error for comparing the means.
b. Calculate the 95% confidence interval comparing the population means for inhalers and noninhalers.
c. Interpret the 95% confidence interval.

Nicotine dependence

HONC Score			
Group	**Sample Size**	**Mean**	**Standard Deviation**
Inhalers	237	2.9	3.6
Noninhalers	95	0.1	0.5

a. Note $s_1/s_2 = 3.6/0.5 = 7.2 > 2$, consider unpooled standard errors. Thus,
$$se = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = 0.24$$
b. $[(\bar{x}_1 - \bar{x}_2) \pm 1.96se] = [2.8 \pm 1.96 * 0.24] = [2.33, 3.27]$

c. With 95% confidence, the inhalers' HONC is at least 2.33 and at most 3.27 more than noninhalers on the average.

Females or males more nicotine dependent?

Refer to previous example about studying nicotine dependence using a random sample of teenagers. Of those seventh graders in the study who had tried tobacco, the mean HONC score was 2.8(s=3.6) for the 150 females and 1.6(s=2.9) for the 182 males.

a. Calculate the standard error for comparing the means.
b. Find the test statistic and the P-value for $H_0 : \mu_1 = \mu_2$ Vs. $H_a : \mu_1 \neq \mu_2$. Interpret and explain what(if any) effect gender has on the mean HONC score.

Females or males more nicotine dependent?

Refer to previous example about studying nicotine dependence using a random sample of teenagers. Of those seventh graders in the study who had tried tobacco, the mean HONC score was 2.8(s=3.6) for the 150 females and 1.6(s=2.9) for the 182 males.

a. Note $s_1/s_2 = 3.6/2.9 = 1.24 < 2$, consider pooled se.

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} = 3.23, se = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = 0.36$$

b.

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{se} = \frac{1.2}{0.36} = 3.36, P-value = P(|T| > 3.36) = 0.0009$$

Reject H_0 since P-value¡ 0.05. We have sufficient evidence to conclude that the population HONC scores are different between genders.

CI for Dependent Samples Comparison

To compare means with dependent samples, conduct confidence interval and hypothesis tests using the single sample of differences.

The 95% confidence interval for the population mean difference $\mu_1 - \mu_2$ (μ_d) is

$$[\bar{x}_d \pm t_{.025}(se)] \,\text{with}\, se = s_d/\sqrt{n}$$

where $\bar{x}_d$ and s_d are sample mean and standard deviation of the differences between two matched-pairs, respectively. Again, $t_{0.025}$ is the t-score having 0.025 right-tail probability with degrees of freedom $n-1$ and n is the number of paired-observations.

1. **Assumption**
   - The sample of differences is a random sample.
   - The differences have an approximately normal distribution.
2. **Hypothesis**
   $H_0 : \mu_1 - \mu_2 = 0$ Vs. $H_a : \mu_1 - \mu_2 \neq 0$(two-sided)
   Or ($H_0 : \mu_d = 0$ Vs. $H_a : \mu_d \neq 0$)
3. **Test Statistic**
   $$T = \frac{\bar{x}_d - 0}{s_d/\sqrt{n}}$$
4. **P-value** Two-tail probability from t distribution of values even more extreme than observed t test statistic, presuming H_0 is true with degrees of freedom $df = n - 1$.
5. **Conclusion** Reject H_0 if P-value $\leq \alpha$(significance level, usually be 0.05). Smaller P-values give stronger evidence against H_0 and supporting H_a. Interpret the P-value in context.

Cell Phone while Driving Reaction Times

An experiment investigated whether cell phone use impairs drivers' reaction times, using a sample of 64 students from the university of Utah. Students were randomly assigned to a cell phone group or to a control group, 32 to each. On a simulation of driving situations, a target flashed red or green at irregular periods. Participants pressed a brake button as soon as they detected a red light. Reaction times are measured when subjects performed the driving task without using cell phones and then again while the same subjects used cell phones.

	Using Cell Phone?		
Student	**No**	**Yes**	**Difference**
1	604	636	32
2	556	623	67
...	...	...	...
31	521	527	6
32	543	536	-7

Cell Phone while Driving Reaction Times

```
Summary statistics:
   Column     n    Mean    Std. dev.   Std. err.
Differences 32 51.59375 117.10868 20.702085
```

95% confidence interval

$$[\bar{x}_d \pm t_{.025,31} se] = [51.59 \pm 2.04 * 20.7] = [9.37, 93.8]$$

Hypothesis Test $H_0 : \mu_d = 0$ Vs. $H_a : \mu_d \neq 0$

$$T = \frac{\bar{x}_d - 0}{se} = \frac{51.59}{20.7} = 2.49, pvalue = P(|T| > 2.49) = 0.018$$

Cell Phone while Driving Reaction Times

```
Paired T hypothesis test:
D = 1 - 2 : Mean of the difference between Cell phone and Cont
H0 : D = 0
HA : D != 0
Hypothesis test results:
Difference              Mean Std. Err. DF T-Stat P-value
Cell phone - Control 51.59375 20.702085 31 2.4922006 0.0182
```

Interpret: At the significance level $\alpha = 0.05$, reject H_0 because P-value ¡ 0.05. We have sufficient evidence to conclude that the population mean reaction times are different for using and not using cellphones while driving.

Outline

1. Comparing Two Means
   - Two Independent Samples Comparison
   - Two Dependent Samples Comparison
2. Comparing Two Proportions
   - Confidence Interval for Comparing Proportions
   - Hypothesis Test for Comparing Proportions
3. Types of Errors in Significance Tests

Comparing Two Population Proportions

A recent **meta-analysis** combined the results of eight related studies with a minimum duration of treatment of four years to determine the effects of aspirin on the risk of cancer depth. All experimental trials used were randomized and double-blind.

	Death from Cancer		
Group	**Yes**	**No**	**Total**
Placebo	347	11,188	11,535
Aspirin	327	13,708	14,035
Total	674	24,896	25,570

Question. Is there of any difference in cancer-death proportion between two populations of intaking placebo and Aspirin? If there is, how much magnitude of values between the two population proportions in terms of cancer death?

Confidence Interval for Comparing Proportions

Assume that we have two populations, with success probabilities p_1 and p_2, respectively. A confidence interval for the difference $p_1 - p_2$ between two population proportions is

$$(\hat{p}_1 - \hat{p}_2) \pm z(se), \; where \; se = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

z-score depends on the confidence level, such as $z = 1.96$ for 95% confidence. The method is constructed based on the following assumptions

- A categorical response variable observed in each of two groups.
- Independent random samples for the two groups, either from random sampling or a randomized experiment.
- Large enough sample sizes n_1 and n_2, so that in each sample there are at least 10 successes and 10 failures.

Confidence Interval for Comparing Proportions

- $\hat{p}_1 = \frac{347}{11{,}535} = 0.0301, \hat{p}_2 = \frac{327}{14{,}035} = 0.0233$
- $$\begin{aligned} se &= \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}} \\ &= \sqrt{\frac{0.0301 * (1-0.0301)}{11,535} + \frac{0.0233 * (1-0.0233)}{14,035}} \\ &= 0.002 \end{aligned}$$
- 95% confidence interval
  $[(\hat{p}_1 - \hat{p}_2) \pm 1.96 * se] = [0.0068 \pm 0.004] = [0.0028, 0.0108]$
- **Interpretation:** With 95% confidence, the cancer-death proportion for placebo-intake population is at least 0.28% and at most 1.08% higher than that for Aspirin-intake population on the average.

Hypothesis Test for Comparing Proportions

1. **Assumptions** (refer to confidence interval)
2. **Hypothesis**
   $H_0 : p_1 = p_2 (p_1 - p_2 = 0)$
   $H_a : p_1 \neq p_2$ (or $p_1 > p_2$ or $p_1 < p_2$)
3. **Test Statistics**

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{se_0} \; with \, se_0 = \sqrt{\hat{p}(1 - \hat{p})(\frac{1}{n_1} + \frac{1}{n_2})}$$

   where $\hat{p}$ is the pooled estimate.
4. **P-value**
   P-value=Two-tail probability from standard normal distribution of values even more extreme than observed z test statistic presuming H_0 is true.
5. **Conclusion**
   Smaller P-values give stronger evidence against H_0 and supporting H_a. Reject H_0 if P-value$\leq \alpha$(significance level, such as 0.05).

Hypothesis Test for Comparing Proportions

1. **Assumptions**
   Categorical, independent random, large enough successes(failures)
2. **Hypothesis** $H_0 : p_1 = p_2$ vs. $H_a : p_1 \neq p_2$
3. **Test Statistic**

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{se_0} = \frac{0.0068}{0.002} = 3.38$$

   where $\hat{p} = \frac{674}{25570} = 0.0264, se_0 = \sqrt{\hat{p}(1-\hat{p})(\frac{1}{n_1} + \frac{1}{n_2})} = 0.002$
4. **P-value** P-value$=P(|Z| > 3.38) = 0.0007$
5. **Conclusion** Reject H_0 since P-value¡0.05. At the level of significance 0.05, we have sufficient evidence to conclude that there is of difference in cancer-death proportion between the placebo-intake population and the Aspirin-intake population.

Outline

1. Comparing Two Means
   - Two Independent Samples Comparison
   - Two Dependent Samples Comparison
2. Comparing Two Proportions
   - Confidence Interval for Comparing Proportions
   - Hypothesis Test for Comparing Proportions
3. Types of Errors in Significance Tests

Types of Errors

Significance tests can have two types of potential errors, called **Type I** and **Type II errors**. Type I and Type II errors are actually two probabilities, and they are inversely related. In other words, as type I error goes down, type II error goes up.

- when H_0 is true, a **Type I error** occurs when H_0 is rejected.
- when H_0 is false, a **Type II error** occurs when H_0 is not rejected.

	Decisions	
Reality About H_0	**Do not reject H_0**	**Reject H_0**
H_0 true	Correct decision	Type I error
H_0 false	Type II error	Correct decision

Table 1: The Four Possible Results of a Decision in a Significance Test

Remark. P(Type I error)=Significance level α. i.e., Suppose H_0 is true. The probability of rejecting H_0, thereby making a Type I error, equals the significance level for the test.

Types of Errors

Defendant	Legal Decision: Acquit	Legal Decision: Convict
Innocent(H_0)	Correct decision	Type I error
Guilty(H_a)	Type II error	Correct decision

Table 2: The Four Possible Results of a legal trial

Consequences for different types of errors

- A potential consequence of a Type I error is sending an innocent person to jail.
- A potential consequence of a Type II error is setting free a guilty person.

Practice Problems

- **Page 436-439** 9.62, 9.63, 9.67, 9.68, 9.77
- **Page 458-459** 10.5, 10.10, 10.11, 10.12
- **Page 470-472** 10.19, 10.20, 10.27, 10.29
- **Page 506-508** 10.79, 10.85, 10.89, 10.94, 10.98