Chapter 10 - Lecture 1
Tests and Confidence Intervals for a Difference between two population means

Yuan Huang

March 18, 2013

1. Case II: large sample with unknown variance for general population
    - Statistical Setting
    - Confidence Interval of $\mu_1 - \mu_2$
    - Testing $H_o : \mu_1 - \mu_2 = c_0$
    - Type II error

Case II: large sample with unknown variance for general population

Statistical Settings:

1. $X_1, \ldots, X_m$ i.i.d , σ_1^2 unknown, $m > 40$;
2. $Y_1, \ldots, Y_n$ i.i.d , σ_2^2 unknown, $n > 40$;
3. The two samples are independent.

Then we have the probability distribution:

$$\frac{\bar{X}_m - \bar{Y}_n - (\mu_1 - \mu_2)}{\sqrt{S_1^2/m + S_2^2/n}} \to N(0, 1).$$

Confidence Interval of $\mu_1 - \mu_2$

In this case, an approximate $100(1-\alpha)\%$ Confidence Interval of $\mu_1 - \mu_2$ is:

$$(\bar{x}_m - \bar{y}_n - z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}, \bar{x}_m - \bar{y}_n + z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}).$$

Similarly we can get one-sided intervals.

Confidence Interval of $\mu_1 - \mu_2$

In this case, an approximate $100(1-\alpha)\%$ Confidence Interval of $\mu_1 - \mu_2$ is:

$$(\bar{x}_m - \bar{y}_n - z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}, \bar{x}_m - \bar{y}_n + z_{\alpha/2}\sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}).$$

Similarly we can get one-sided intervals.

Note: conditions $m > 40$ and $n > 40$.

Example 10.5 For many calculus instructors it seems that students taking Calculus I in fall semester are better prepared than tare the students taking it in the spring. If so, it would be nice to have some measure of the difference. We use data from a study of the influence of various predictor on calculus performance.

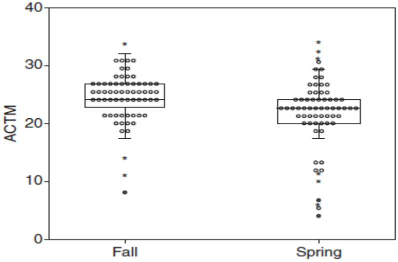

Figure 10.2 Boxplot/dotplot for fall and spring ACT mathematics scores

Table 10.3 Summary results for Example 10.5

Group	Sample Size	Sample Mean	Sample SD
Fall	80	28.25	3.25
Spring	74	25.88	4.59

Let's now, use a confidence level of 95% ($z_{0.025} = 1.96$)to calculate a confidence interval for the difference between true average fall ACT score and true average spring ACT score:

$$\begin{aligned}&(28.25 - 25.88) \pm (1.96)\sqrt{\frac{3.25^2}{80} + \frac{4.59^2}{74}}\\ = \quad & 2.37 \pm (1.96)(0.6456) = (1.10, 3.64)\end{aligned}$$

Interpretation: That is, with 95% confidence, $1.10 < \mu_1 - \mu_2 < 3.64$. We can therefore be highly confident that the true fall average exceeds the true spring average by between 1.10 and 3.64.

Testing $H_o : \mu_1 - \mu_2 = c_0$

In this case the test statistic $H_o : \mu_1 - \mu_2 = c_0$ is:

$$Z = \frac{\bar{X}_m - \bar{Y}_n - c_0}{\sqrt{s_1^2/m + s_2^2/n}}$$

1. For $H_1 : \mu_1 - \mu_2 > c_0$, we reject H_o when $z > z_\alpha$;
2. For $H_1 : \mu_1 - \mu_2 < c_0$, we reject H_o when $z < -z_\alpha$;
3. For $H_1 : \mu_1 - \mu_2 \neq c_0$, we reject H_o when $|z| > z_{\alpha/2}$.

Recall: p-value approach

Testing $H_o : \mu_1 - \mu_2 = c_0$

In this case the test statistic $H_o : \mu_1 - \mu_2 = c_0$ is:

$$Z = \frac{\bar{X}_m - \bar{Y}_n - c_0}{\sqrt{s_1^2/m + s_2^2/n}}$$

1. For $H_1 : \mu_1 - \mu_2 > c_0$, we reject H_o when $z > z_\alpha$;
2. For $H_1 : \mu_1 - \mu_2 < c_0$, we reject H_o when $z < -z_\alpha$;
3. For $H_1 : \mu_1 - \mu_2 \neq c_0$, we reject H_o when $|z| > z_{\alpha/2}$.

Recall: p-value approach

Note: conditions $m > 40$ and $n > 40$.

Relationship between the confidence interval and the hypothesis test (for any number of populations) based on the same probability distribution:

1. We reject H_o in two-tailed test if the null value is not included in the two-tailed C.I;
2. We reject H_o in upper-tailed test if the null value is not included in the lower-tailed C.I;
3. We reject H_o in lower-tailed test if the null value is not included in the upper-tailed C.I.

Calculating Type II error probabilities

Let

$$s = \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}}$$

1. $H_A : \mu_1 - \mu_2 > \Delta_0$,

$$\beta(\Delta') = \Phi\left(z_\alpha - \frac{\Delta' - \Delta_0}{s}\right)$$

2. $H_A : \mu_1 - \mu_2 < \Delta_0$

$$\beta(\Delta') = 1 - \Phi\left(-z_\alpha - \frac{\Delta' - \Delta_0}{s}\right)$$

3. $H_A : \mu_1 - \mu_2 \neq \Delta_0$

$$\beta(\Delta') = \Phi\left(z_{\alpha/2} - \frac{\Delta' - \Delta_0}{s}\right) - \Phi\left(-z_{\alpha/2} - \frac{\Delta' - \Delta_0}{s}\right)$$

Example 10.4

Background: A study was carried out in an attempt to improve student performance in a lowlevel university mathematics course. Experience had shown that many students had fallen by the wayside, meaning that they had dropped out or completed the course with minimal effort and low grades.

Method: The study involved assigning the students to sections based on odd or even Social Security number. It is important that the assignment to sections not be on the basis of student choice, because then the differences in performance might be attributable to differences in student attitude or ability.

1. Half of the sections were taught traditionally
2. The other half were taught in a way that hopefully would keep the students involved. They were given frequent assignments that were collected and graded, they had frequent quizzes, and they were allowed retakes on exams.

Data: Lotus Hershberger conducted the experiment and he supplied the data. Here are the final exam scores for the 79 students taught traditionally (the control group) and for the 85 students taught with more involvement (the experimental group):

Table 10.1 Summary results for Example 10.4

Group	Sample Size	Sample Mean	Sample SD
Control	79	23.87	11.60
Experimental	85	27.34	8.85

Test:

1. Let μ_1 and μ_2 denote the true mean scores for the control condition and the experimental condition, respectively.
   $H_0 : \mu_1 - \mu_2 = 0$ versus $H_0 : \mu_1 - \mu_2 < 0$
2. $m > 40$ and $n > 40$, hence z test could be applied.
3. Test statistics value: $\frac{(23.87-27.34)}{\sqrt{\frac{11.60^2}{79}+\frac{8.85^2}{85}}} = -2.14$
4. Rejection region: $z \leq -z_{0.05} = -1.645$ or
   P-value$= \Phi(z) = \Phi(-2.14) = 0.016$
5. Based on P-value< 0.05, or based on $z \in$ Rejection region.
   Reject H_0 at the 0.05 critical level.

We have shown fairly conclusively that the experimental method of instruction is an improvement. However, it is important to view the data graphically to see if there is anything strange.

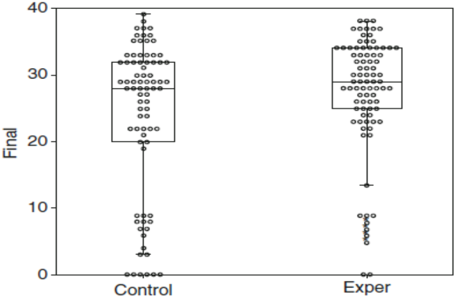

Figure 10.1 Boxplot/dotplot for the teaching experiment

What happens if we compare the groups while ignoring the low performers whose scores are below 10?

Table 10.2 Summary results without poor performers

Group	Sample Size	Sample Mean	Sample SD
Control	61	29.59	5.005
Experimental	76	29.88	4.950

If we perform the z test again, the z statistic gives -0.34, which provide no evidence to reject the H_0.

Using a Comparison to Identify Causality

A **randomized controlled experiment** results when investigators assign subjects to the two treatments in a random fashion. When statistical significance is observed in such an experiment, the investigator and other interested parties will have more confidence in the conclusion that the difference in response has been caused by a difference in treatments.

In contrast, in an **observational study**, the researchers simply "observe" a group of subjects without actually doing anything to the subjects.

- *A prospective cohort* study is a cohort study that follows over time a group of similar individuals (cohorts) who differ with respect to certain factors under study, to determine how these factors affect rates of a certain outcome. (by Wikipedia)
- *A retrospective cohort* study generally means to take a look back at events that already have taken place.(by Wikipedia)