5/10/12

- Warm up by finishing example 2 from 9.2
- Continue Ch.9

Tuesday 5/15

- Finish Ch.9
- Extra credit SLO assessment → NO Make-up

* Project DUE

Thursday 5/17

Review for Final

Tuesday 5/22

- Final exam from 5-7PM
- Ch.9, 10 homework due

Example 2: Among 2739 female atom bomb survivors, 1397 developed thyroid diseases. Among 1352 male atom bomb survivors, 436 developed thyroid diseases (based on data from "Radiation Dose-Response Relationships for Thyroid Nodules and Autoimmune Thyroid Diseases in Hiroshima and Nagasaki Atomic Bomb Survivors 55-58 Years After Radiation Exposure," by Imaizumi, et al., *Journal of the American Medical Association,* Vol. 295, No. 9).

a. Use a 0.01 significance level to test the claim that the female survivors and male survivors have different rates of thyroid diseases.

$\hat{p}_1 = \frac{1397}{2739} \approx 0.510$, $n_1 = 2739$ $\quad \bar{p} = \frac{1397+436}{2739+1352} \approx 0.448$

$\hat{p}_2 = \frac{436}{1352} \approx 0.322$, $n_2 = 1352$ $\quad \bar{q} = 0.552$

① $H_0: p_1 = p_2$

$H_A: p_1 \neq p_2$

② $z = \dfrac{0.510 - 0.322 - 0}{\sqrt{\frac{(.448)(.552)}{2739} + \frac{(.448)(.552)}{1352}}} \approx 11.37$

③ P-value is:

$2P(z > 11.37) = 2[1 - P(z < 11.37)] = 2[0.0001] = 0.0002$

and $0.0002 < 0.01$ so reject H_0

④ There is sufficient evidence to support the claim that female and male survivors have different rates of thyroid disease.

b. Construct the confidence interval corresponding to the hypothesis test conducted with a 0.01 significance level.

$E = 2.575\sqrt{\frac{(0.510)(0.490)}{2739} + \frac{(0.322)(0.678)}{1352}}$

$E \approx 0.0409$

$(0.510 - 0.322) - 0.0409 \leq p_1 - p_2 \leq (0.510 - 0.322) + 0.0409$

$\boxed{0.147 \leq p_1 - p_2 \leq 0.229}$

c. What conclusion does the confidence interval suggest?

The CI suggests that the rates are different since zero is not included.

9.3 INFERENCES ABOUT TWO MEANS: INDEPENDENT SAMPLES

Key Concept…

In this section, we present methods for using sample data from 2 independent samples to test hypotheses made about 2 population means or to construct confidence interval estimates of the difference between 2 population means.

PART I: INDEPENDENT SAMPLES WITH σ_1 AND σ_2 UNKNOWN AND NOT ASSUMED EQUAL

DEFINITION

Two samples are **independent** if the sample values from one population are not related to or somehow naturally paired or matched with the sample values from the other population.

Two samples are **dependent** if the sample values are paired.

Inferences about Means of Two Independent Populations, With σ_1 and σ_2 Unknown and Not Assumed to be Equal

OBJECTIVES

Test a claim about 2 independent pop. means or construct a CI estimate of the difference between 2 independent pop. means.

NOTATION

Population 1:

μ_1 = population mean

s_1 = sample standard deviation

σ_1 = population standard deviation

$\bar{x}_1$ = sample mean

n_1 = size of the first sample

The corresponding notations for μ_2, σ_2, $\bar{x}_2$, s_2, and n_2 apply to population 2.

REQUIREMENTS

1. σ_1 and σ_2 are unknown and it is not assumed that σ_1 and σ_2 are equal.
2. The 2 samples are independent.
3. Both samples are simple random samples.
4. Either or both of these conditions are satisfied: The two sample sizes are both large (with $n_1 > 30$ and $n_2 > 30$) or both

samples come from populations having normal distributions.

HYPOTHESIS TEST STATISTIC FOR TWO MEANS: INDEPENDENT SAMPLES

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $\mu_1 - \mu_2$ are assumed to be zero under the null

Degrees of Freedom: When finding critical values or P-values, use the following for determining the number of degrees of freedom.

1. In this book we use the conservative estimate: df = Smaller of $n_1 - 1$ and $n_2 - 1$.

2. Statistical software packages typically use the more accurate but more difficult estimate given below:

$$\text{df} = \frac{(A+B)^2}{\frac{A^2}{n_1 - 1} + \frac{B^2}{n_2 - 1}}, \quad A = \frac{s_1^2}{n_1}, \quad B = \frac{s_2^2}{n_2}$$

***P*-values and critical values:** Use Table A-3.

CONFIDENCE INTERVAL ESTIMATE OF $\mu_1 - \mu_2$: INDEPENDENT SAMPLES

The confidence interval estimate of the difference $\mu_1 - \mu_2$ is

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E \quad \text{where } E = t_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

and the number of degrees of freedom df is as described above for hypothesis tests.

EQUIVALENCE OF METHODS

The P-value method of hypothesis testing, the traditional method of hypothesis testing, and confidence intervals all use the same distribution and standard error, so they are equivalent in the sense that they result in the same conclusions. A null hypothesis of $\mu_1 = \mu_2$ or $\mu_1 - \mu_2 = 0$ can be tested using the P-value method, the traditional method, or by determining whether the confidence interval includes zero.

Example 1: Determine whether the samples are independent or dependent.

a. To test the effectiveness of Lipitor, cholesterol levels are measured in 250 subjects before and after Lipitor treatments.

Dependent

b. On each of 40 different days, the author measured the voltage supplied to his home and he also measured the voltage produced by his gasoline powered generator.

Independent

Example 2: Assume that the two samples are independent simple random samples selected from normally distributed populations. Do not assume that the population standard deviations are equal. A simple random sample of 13 four-cylinder cars is obtained, and the braking distances are measured. The mean braking distance is 137.5 feet and the standard deviation is 5.8 feet. A SRS of 12 six-cylinder cars is obtained and the braking distances have a mean of 136.3 feet with a standard deviation of 9.7 feet (based on Data Set 16 in Appendix B).

a. Construct a 90% CI estimate of the difference between the mean braking distance of four-cylinder cars and six-cylinder cars.

$\bar{x}_1 = 137.5,\ s_1 = 5.8,\ n_1 = 13$

$\bar{x}_2 = 136.3,\ s_2 = 9.7,\ n_2 = 12$

d.f.: $12 - 1 = 11$

$t_{.05} = 1.796$ (if you look at area in 1 tail)

$t_{.10} = 1.796$ (if you look at area in 2 tails)

$$E = 1.796\sqrt{\frac{5.8^2}{13} + \frac{9.7^2}{12}}$$

$E \approx 5.8$

$$(137.5 - 136.3) - 5.8 < \mu_1 - \mu_2 < (137.5 - 136.3) + 5.8$$

$$\boxed{-4.6 < \mu_1 - \mu_2 < 7.0}$$

b. Does there appear to be a difference between the two means?

$\boxed{\text{No} \rightarrow \text{the CI includes zero.}}$

c. Use a 0.05 significance level to test the claim that the mean braking distance of four-cylinder cars is greater than the mean braking distance of six-cylinder cars.

① $H_0: \mu_1 = \mu_2$

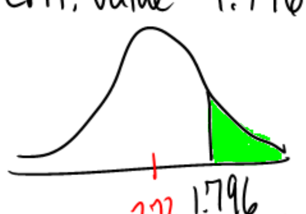
$H_A: \mu_1 > \mu_2$

② $$t = \frac{137.5 - 136.3 - 0}{\sqrt{\frac{5.8^2}{13} + \frac{9.7^2}{12}}}$$

$t \approx 0.372$

③ Critical value test (traditional)

Crit. value = 1.796



not in rejection region

Fail to reject Ho

④ There is not sufficient evidence to support the claim that the mean braking distance of four-cylinder cars is greater than six-cylinder cars

PART 2: ALTERNATIVE METHODS

Part 1 in this section dealt with situations in which the two population standard deviations are unknown and not assumed to be equal. In Part 2 we address two other situations: (1) The two population standard deviations are both known; (2) the two population standard deviations are unknown but assumed to be equal.

ALTERNATIVE METHOD WHEN σ_1 AND σ_2 ARE KNOWN

In reality, the population standard deviations are almost never known, but if they are known, the test statistic and confidence interval are based on the normal distribution instead of the t distribution

Inferences about Means of Two Independent Populations, With σ_1 and σ_2 Known

REQUIREMENTS

1. The two population standard deviations σ_1 and σ_2 are both known.
2. The two samples are independent.
3. Both samples are simple random samples.

4. Either or both of these conditions is satisfied: The two sample sizes are both large, with $n_1 > 30$ and $n_2 > 30$ or both samples come from populations having normal distributions.

HYPOTHESIS TEST

Test statistic: $$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$, where $\mu_1 - \mu_2 = 0$ under the null

P-values and critical values: Refer to Table A2.

CONFIDENCE INTERVAL ESTIMATE OF $\mu_1 - \mu_2$

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$

$$E = z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

ALTERNATIVE METHOD: ASSUME THAT $\sigma_1 = \sigma_2$ AND POOL THE SAMPLE VARIANCES

Even when the specific values of σ_1 and σ_2 are not known, if it can be assumed that they have the same value, the sample variances s_1^2 and s_2^2 can be pooled to obtain an estimate of the common population variance σ^2. The pooled estimate of σ^2 is denoted s_p^2 and is a weighted average of s_1^2 and s_2^2, which is shown below.

Inferences about Means of Two Independent Populations, Assuming that $\sigma_1 = \sigma_2$

REQUIREMENTS

1. The two population standard deviations are not known, but they are assumed to be equal. That is, $\sigma_1 = \sigma_2$.
2. The two samples are independent.
3. Both samples are simple random samples.
4. Either or both of these conditions is satisfied: The two sample sizes are both large, with $n_1 > 30$ and $n_2 > 30$ or both samples come from populations having normal distributions.

HYPOTHESIS TEST

Test statistic: $$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$$, where $\mu_1 - \mu_2 = 0$ under the null

where

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

and the number of degrees of freedom is given by $n_1 + n_2 - 2$.

CONFIDENCE INTERVAL ESTIMATE OF $\mu_1 - \mu_2$

Confidence interval:

$$(\bar{x}_1 - \bar{x}_2) - E < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + E$$

where $E = t_{\alpha/2}\sqrt{\dfrac{s_p^2}{n_1} + \dfrac{s_p^2}{n_2}}$

and s_p^2 is as given in the above test statistic and the number of degrees of freedom is $n_1 + n_2 - 2$.

TI-83/84 PLUS

STAT

TESTS

2-SAMPTtest (hypothesis test)

or

2-SAMPTint (CI)

Example 3: Assume that the two samples are independent simple random samples selected from normally distributed populations. Also assume that the population standard deviations are equal. The mean tar content of a simple random sample of 25 unfiltered king size cigarettes is 21.1 mg, with a standard deviation of 3.2 mg. The mean tar content of a SRS of 25 filtered 100 mm cigarettes is 13.2 mg with a standard deviation of 3.7 mg (based on Data Set 4 in Appendix B).

a. Construct a 90% CI estimate of the difference between mean tar content of unfiltered king size cigarettes and the mean tar content of filtered 100 mm cigarettes.

$\bar{x}_1 = 21.1$, $s_1 = 3.2$, $n_1 = 25$

$\bar{x}_2 = 13.2$, $s_2 = 3.7$, $n_2 = 25$

d.f.: $25 + 25 - 2 = 48$

$$s_p^2 = \frac{24(3.2)^2 + 24(3.7)^2}{24 + 24}$$

$$s_p^2 = 11.965$$

$$E = 1.676\sqrt{\frac{11.965}{25} + \frac{11.965}{25}}$$

$$E \approx 1.6$$

$$(21.1 - 13.2) - 1.6 < \mu_1 - \mu_2 < (21.1 - 13.2) + 1.6$$

$$\boxed{6.3 < \mu_1 - \mu_2 < 9.5}$$

b. Does the result suggest that 100 mm filtered cigarettes have less tar than unfiltered king size cigarettes?

Yes – the CI estimate of the mean difference is positive.

c. Use a 0.05 significance level to test the claim that unfiltered king size cigarettes have a mean tar content greater than that of filtered 100 mm cigarettes. What does the reslt suggest about the effectiveness of cigarette filters?

① $H_0: \mu_1 = \mu_2$

$H_A: x_1 > x_2$

② $$t = \frac{(21.1 - 13.2) - 0}{\sqrt{\frac{11.965}{25} + \frac{11.965}{25}}} \approx 8.075$$

③ P-value $< 0.005 \rightarrow 0.005 < 0.05$ so reject H_0

④ There's sufficient evidence to support the claim that unfiltered king size cigarettes have a mean tar content greater than that of filtered 100mm cigarettes.

9.4 INFERENCES FROM DEPENDENT SAMPLES

Key Concept…

In this section we present methods for testing hypotheses and constructing confidence intervals involving the _______________ of the _________________ of the ______________ of two ______________ ___________________. With _______________ samples, there is some ___________________ whereby each value in one sample is _____________ with a _________________ value in the other sample. Here are two typical examples of dependent samples:

- Each pair of sample values consists of two measurements from the _____________ subject
- Each pair of sample values consists of a _____________ __________.

Because the hypothesis test and CI use the same _________________ and ________________ ______________, they are _______________ in the sense that they result in the ______________ ___________________.

Consequently, the ______________ hypothesis that the _____________ _______________ _______________ _____ can be tested by determining whether the _________________ _________________ includes ______.

There are no exact procedures for dealing with __________________ samples, but the ______ ______________________ serves as a reasonably good approximation, so the following methods are commonly used.

Inferences about Means of Two Dependent Populations

OBJECTIVES

NOTATION

$d =$ $s_d =$

μ_d

$\bar{d} =$ $n =$

REQUIREMENTS

1. The _____________ data are ______________________.
2. The samples are _________________ ____________________ __________________.
3. Either or both of these conditions are satisfied: The number of ____________ of ________________ is ________________ (___________) or the pairs of values have _____________________ that are from a population that is approximately _____________________.

HYPOTHESIS TEST FOR DEPENDENT SAMPLES

$t = \underline{\qquad\qquad\qquad\qquad\qquad\qquad}$

Degrees of Freedom: ____________

***P*-values and critical values:** Use Table A-3.

CONFIDENCE INTERVALS FOR DEPENDENT SAMPLES

where

and

Example 1: Assume that the paired sample data are SRSs and that the differences have a distribution that is approximately normal.

a. Listed below are BMIs of college students.

April BMI	20.15	19.24	20.77	23.85	21.32
September BMI	20.68	19.48	19.59	24.57	20.96

i. Use a 0.05 significance level to test the claim that the mean change in BMI for all students is equal to 0.

ii. Construct a 95% CI estimate of the change in BMI during freshman year.

iii. Does the CI include zero, and what does that suggest about BMI during freshman year?

b. Listed below are systolic blood pressure measurements (mm Hg) taken from the right and left arms of the same woman. Use a 0.05 significance level to test for a difference between the measurements from the two arms. What do you conclude?

Right arm	102	101	94	79	79
Left arm	175	169	182	146	144