Warm Up

Pick up the handout and look at the front side.

EXPERIMENTAL DESIGN 8

Which of the following is most useful in establishing cause-and-effect relationships?

(A) A complete census

(B) A least squares regression line showing high correlation

(C) A simple random sample (SRS)

(D) A well-designed, well-conducted survey incorporating chance to ensure a representative sample

(E) A controlled experiment

Regression lines show association, not causation, Surveys suggest relationships, which controlled experiments can help show to be cause and effect.

Answer: **E**

EXPERIMENTAL DESIGN 12

Sampling error occurs

(A) when interviewers make mistakes resulting in bias.

(B) when interviewers use judgment instead of random choice in picking the sample.

(C) when samples are too small.

(D) because a sample statistic is used to estimate a population parameter.

(E) in all of the above cases.

Different samples give different statistics, all of which are estimates for the same population parameter, and so error, called sampling error, is naturally present.

Answer: **D**

Ch.8 Confidence Intervals σ μ

Ch.9 Hypothesis Test σ μ

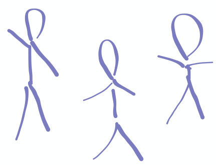
Ch.10 Confidence Intervals Hypotheses Tests $p - p$ } 10.1

Confidence Intervals Hypotheses Tests $\mu - \mu$ } 10.2

10.2 Day 1 10.2 Day 2

TWO GROUPS
(heights)

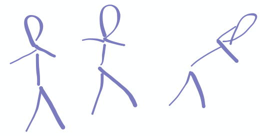
both groups
Population
Normally distributed



$\mu_G = 56.4$ in

$\sigma_G = 2.7$ in

10 year old Girls



$\mu_B = 55.7$ in

$\sigma_B = 3.8$ in

10 year old boys

Suppose we take
independent SRS's
of 12 girls and 8 boys

and measure their heights

but, what can we say about the difference in sample means $\bar{X}_G - X_B$

The Sampling Distribution of $\bar{x}_1 - \bar{x}_2$

- The sampling distribution of $\bar{x}_1 - \bar{x}_2$ is **Normal** if both population distributions are Normal. It is **approximately Normal** if both sample sizes are large ($n_1 \geq 30$ and $n_2 \geq 30$) or if one population is Normally distributed and the other sample size is large.

Shape

The Sampling Distribution of $\bar{x}_1 - \bar{x}_2$

- The sampling distribution of $\bar{x}_1 - \bar{x}_2$ is **Normal** if both population distributions are Normal. It is **approximately Normal** if both sample sizes are large ($n_1 \geq 30$ and $n_2 \geq 30$) or if one population is Normally distributed and the other sample size is large.
- **The mean of the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is**

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$$

Center

The Sampling Distribution of $\bar{x}_1 - \bar{x}_2$

- The sampling distribution of $\bar{x}_1 - \bar{x}_2$ is **Normal** if both population distributions are Normal. It is **approximately Normal** if both sample sizes are large ($n_1 \geq 30$ and $n_2 \geq 30$) or if one population is Normally distributed and the other sample size is large.
- The **mean** of the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$$

- The **standard deviation** of the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

as long as the *10% condition* is met for both samples: $n_1 < 0.10N_1$ and $n_2 < 0.10N_2$.

Variability

So.......

When creating C.I.'s

for $\mu_1 - \mu_2$

Conditions for Constructing a Confidence Interval About a Difference in Means

Random: The data come from two independent random samples or from two groups in a randomized experiment.

10%: When sampling without replacement, $n_1 < 0.10N_1$ and $n_2 < 0.10N_2$.

Large Counts: Normal/Large Sample: For each sample, the corresponding population distribution (or the true distribution of response to the treatment) is Normal or the sample size is large ($n \geq 30$). For each sample, if the population (treatment) distribution has unknown shape and $n < 30$, a graph of the sample data shows no strong skewness or outliers.

HW
LOTTERY

Is there a difference between the number of chocolate chips in a Chips Ahoy cookies versus Farmhouse cookies?

Which cookie has the most chips?

VS **~~Treasure Chips~~** Farmhouse

Is there a difference in the number of chocolate chips in **Chips Ahoy** cookies versus the number of chocolate chips in **Treasure Chips** cookies? Each pair of students will count the number of chocolate chips in 1 Chips Ahoy cookie and 1 Treasure Chips cookie. Due to the factories processes, we can assume the population distributions of # of chips are approximately normal and that the samples are random.

1. Record the number of chocolate chips in each cookie. Write them on the board.

in Chips Ahoy = _______ # in Treasure Chips = _______

It doesn't matter how the chips are counted as long as you're consistent with both cookies.

Disect if you want.

# of chips in your cookie

13 20 27 28

~~Treasure Chips~~ Farmhouse

8 11 10 11

2. Find the class totals find the mean number of chocolate chips for each type of cookie, the standard deviation and the difference.

Chips Ahoy: $\bar{x}_1 =$ ~~Treasure Chips~~ Farmhouse Chips: $\bar{x}_2 =$ Difference: $\bar{x}_1 - \bar{x}_2 =$

$s_1 =$ $s_2 =$

3. If we repeated this process many times and created a dotplot, we would have the sampling distribution of $\bar{x}_1 - \bar{x}_2$. Describe the shape, center and variabity of the sampling distribution.

Shape: Approx normal since the populations are approx. Normal

Center: $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

Variability: $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Random "samples are random"

10% $n_1 < \frac{1}{10}$(All Chips Ahoy Cookies)

$n_2 < \frac{1}{10}$(All Farmhouse Cookies)

Normal already established in #3

4. With at least one other person in the class, discuss and confirm tthat the conditions for constructing a confidence interval been met? (don't have to write details)

5. Construct a 95% confidence interval for the true difference in the mean number of chocolate chips in Chips Ahoy and Treasure Chips. (you don't need to write the general or specific formulas in this particular unless you want) but do show the "work".

Formula use

$$\left(\bar{X}_1 - \bar{X}_2\right) \pm t^* \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

4. With at least one other person in the class, discuss and confirm tthat the conditions for constructing a confidence interval been met? (don't have to write details)

5. Construct a 95% confidence interval for the true difference in the mean number of chocolate chips in Chips Ahoy and Treasure Chips. (you don't need to write the general or specific formulas in this particular unless you want) but do show the "work".

Formula use

$$\left(\bar{X}_1 - \bar{X}_2\right) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$12 \pm 3.18 \sqrt{\frac{6.98^2}{4} + \frac{1.41^2}{4}}$$

Need degrees of freedom

$$12 \pm 11.32 \longrightarrow (0.68, 23.32)$$

11.3208

Degrees of Freedom

It's complicated

Option 1 (technology): Just let technology compute the df. This will most likely not be a whole number.

Option 2 (conservative): smaller of $n_1 - 1$ and $n_2 - \cancel{2}$ 1

df = 4-1
= 3

df = 4-1
= 3

It's a cinch

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

Just kidding

degrees of freedom

It's complicated

Best Option

Option 1 (technology): Just let technology compute the df. This will most likely not be a whole number.

Option 2 (conservative): smaller of $n_1 - 1$ and $n_2 - 2$

we'll use this for now

where t^* is the critical value with $C\%$ of the area between $-t^*$ and t^* for the t distribution with degrees of freedom using either

- Option 1 (let technology calculate the degrees of freedom) or
- Option 2 (the smaller of $n_1 - 1$ and $n_2 - 1$)

smaller df would yield larger which is more conservative

$n_1 =$ $n_2 =$

TABLE B

C 95%

Tail Prob .025

$df = n-1$

$df =$

$t^* =$

4. With at least one other person in the class, discuss and confirm tthat the conditions for constructing a confidence interval been met? (don't have to write details)

5. Construct a 95% confidence interval for the true difference in the mean number of chocolate chips in Chips Ahoy and Treasure Chips. (you don't need to write the general or specific formulas in this particular unless you want) but do show the "work".

$$\bar{x}_1 - \bar{x}_2 \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$\pm \sqrt{}$$

4. With at least one other person in the class, discuss and confirm tthat the conditions for constructing a confidence interval been met? (don't have to write details)

5. Construct a 95% confidence interval for the true difference in the mean number of chocolate chips in Chips Ahoy and Treasure Chips. (you don't need to write the general or specific formulas in this particular unless you want) but do show the "work".

Option 1 (Tech)

If using Technology for the "DO" step, you must:

- -have already named the procedure in the in the "Plan" step
- -Give the interval (3.9632, 17.724) for example
- -and report the degrees of freedom.

4. With at least one other person in the class, discuss and confirm tthat the conditions for constructing a confidence interval been met? (don't have to write details)

5. Construct a 95% confidence interval for the true difference in the mean number of chocolate chips in Chips Ahoy and Treasure Chips. (you don't need to write the general or specific formulas in this particular unless you want) but do show the "work".

$\bar{x}_1 = 22$

$s_1 = 6.98$

$n_1 = 4$

$\bar{x}_2 = 10$

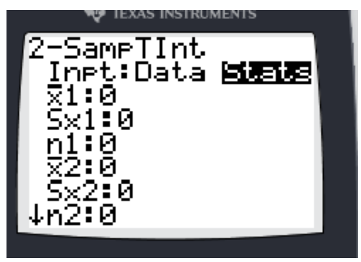
$s_2 = 1.41$

$n_2 = 4$

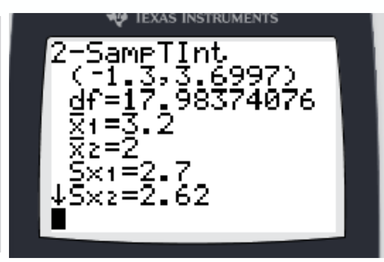
2-SampTInt gives

(1.137 , 22.863)

using df = 3.24



for 2 sample t procedures No to pooling



example

Should I Tech or Table?

↑ ↑ ↑
recommended
on FR

The two methods will not produce the same answer

If you write out the formula with numbers substituted in •
a) leave t^* • in the formula instead of using conservative value •
b) Then use 2 SampTInt and report the calculator's interval and df.

6. Do we have evidence that there is a difference in the average number of chocolate chips in a Chips Ahoy and a ~~Treasure~~ cookie?

$(-,-) \rightarrow$ ~~Treasure~~ Farmhous C has more —

$(+,+) \rightarrow$ Chips Ahoy has more •

$(-,+) \rightarrow$ No difference

What if we swithed the order of subtraction to $(\hat{p}_2 - \hat{p}_1)$?

example: $(\overset{\hat{p}_1}{34.53} - \overset{\hat{p}_2}{23.70})$ becomes $(\overset{\hat{p}_2}{23.70} - \overset{\hat{p}_1}{34.53})$

10.83 becomes −10.83

confid. interval (3.83, 17.83) gets flipped (−17.83, −3.83)

In answers in the back of your textbook:

It's OK if the endpoints of the endpoints of the interval in your answer to have opposite signs.

Either subtraction order is correct as long as you clearly identify the order you are using.

What's expected on the AP exam for CI for diff. of means

State
- Defind both paramters
- State Confidence level

Plan Plan
- Idetify the procedure
- State and check conditions

confidence intervals for diff. of means

Two Sample t Interval for $\mu_1 - \mu_2$

Assuming the conditions have been met

Important ideas:

Sampling Distribution

shape

center

Variability

Key Points

confidence intervals for diff of means

Two Sample t Interval for $\mu_1 - \mu_2$

Assuming the conditions have been met

Important ideas:

Sampling Distribution

shape • Normal
- Pop is normal
- $n \geq 30$ (CLT)
- No strong skew or outliers

Center

Variability

Key Points

confidence intervals for diff of means

Two Sample t Interval for $\mu_1 - \mu_2$

Assuming the conditions have been met

Important ideas:

Sampling Distribution

shape • Normal
- Pop is normal
- $n \geq 30$ (CLT)
- No strong skew or outliers

Center

$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

Variability

Key Points

confidence intervals for diff of means

Two Sample t Interval for $\mu_1 - \mu_2$

Assuming the conditions have been met

Important ideas:

Sampling Distribution

shape: Normal
- Pop is normal
- $n \geq 30$ (CLT)
- No strong Skew or outliers

Center $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

Variability

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Key Points

confidence intervals for diff of means

Two Sample t Interval for $\mu_1 - \mu_2$

Assuming the conditions have been met

Calculator App

Important ideas:

Sampling Distribution

shape: Normal
- Pop is normal
- $n \geq 30$ (CLT)
- No strong Skew or outliers

Center $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

Variability

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Key Points

confidence intervals for diff of means

Two Sample t Interval for $\mu_1 - \mu_2$

Assuming the conditions have been met

Calculator App

Important ideas:

Sampling Distribution

shape: Normal
- Pop is normal
- $n \geq 30$ (CLT)
- No strong Skew or outliers

center: $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

Variability: $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Key Points

Conditions: Random, 10%, Large Sample

confidence intervals for diff of means

Two Sample t Interval for $\mu_1 - \mu_2$

Assuming the conditions have been met

Calculator App

Important ideas:

Sampling Distribution

shape: Normal
- Pop is normal
- $n \geq 30$ (CLT)
- No strong Skew or outliers

center: $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

Variability: $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Key Points

Conditions: Random, 10%, Large Sample

CI

statistic ± (critical value)(Stand. error)

Confidence intervals for diff of means

Two Sample t Interval for $\mu_1 - \mu_2$

Assuming the conditions have been met

Calculator App

Important ideas:

Sampling Distribution

shape: Normal
- Pop is normal
- $n \geq 30$ (CLT)
- No strong Skew or outliers

Center: $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$

Variability: $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Key Points

Conditions: Random, 10%, Large Sample

CI

statistic ± (critical value)(Stand. error)

$(\bar{x}_1 - \bar{x}_2) \pm$

---

statistic ± (critical value) · (standard deviation of statistic)

$(\bar{x}_1 - \bar{x}_2) \pm$ (critical value) · (standard deviation of statistic)

don't know population std. deviations most of the time

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$SE_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\boxed{s_1^2}}{n_1}+\frac{\boxed{s_2^2}}{n_2}}$$

We use s_1 and s_2 to estimate σ_1 and σ_2. We call this the **standard error** of $\bar{x}_1 - \bar{x}_2$.

confidence intervals for diff of means

Two Sample *t* Interval for $\mu_1 - \mu_2$

Assuming the conditions have been met

Important ideas:

Sampling Distribution

shape: Normal
- Pop is normal
- $n \geq 30$ (CLT)
- No strong Skew or outliers

center $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$

Variability $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$

Key Points

Conditions Random, 10%, Large Sample

CI

statistic ± (critical value)(Stand. error)

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

2 Samp T Int

	rs of Sampling Distribution	Standard Error* of Sample Statistic
	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
$-\mu_2$	$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$

Absolute/Must Haves for Confidence Intervals for a difference of Means

State
- Define both parameters
- State Confidence level

Plan
- Identify the procedure
- State and check conditions (in experiments, we are not sampling w/o replacement so don't check 10% cond.)

Do

Calculate the Confidence Interval

1. State sample $\bar{x}$'s and s_x's
2. Calculate the interval with option 1 or 2 (but not both)

Option 1 – include *for example:* $2-SampTInt$ gives $(3.9632, 17.724)$
using df = 13

Option 2 - include *for example:* $df = 15$ $t^* = 1.26$ and

$$CI = (43.24 - 38.26) \pm 1.26\sqrt{\frac{14.26^2}{30} + \frac{17.50^2}{34}} = \ldots\ldots$$

It's not a must to show the general & specific formula but the longer you continue the better you will know it for M/C (and ch. 12)

Conclude

- Report Confidence level (*We are.....*)
- Report *computed interval*
- Give interpretation in context ← like you normally do

Now Put it all together

Pulse Rates

Pulse Rates:

Mr. Cedarlund's class performed an experiment to investigate whether drinking a caffeinated beverage would increase pulse rates. Twenty students in the class volunteered to take part in the experiment. All of the students measured their initial pulse rates (in beats per minute). Then Mr. Wilcox randomly assigned the students into two groups of 10. Each student in the first group drank 12 ounces of cola with caffeine. Each student in the second group drank 12 ounces of caffeine-free cola. All students then measured their pulse rates again. The table displays the change in pulse rate for the students in both groups.

	Change in pulse rate (Final pulse rate − Initial pulse rate)										Mean change
Caffeine	8	3	5	1	4	0	6	1	4	0	3.2
No caffeine	3	−2	4	−1	5	5	1	2	−1	4	2.0

Construct and interpret a 95% confidence interval for the difference in true mean change in pulse rate for subjects like these who drink caffeine versus who drink no caffeine.

Pulse Rates:

Mr. Cedarlund's class performed an experiment to investigate whether drinking a caffeinated beverage would increase pulse rates. Twenty students in the class volunteered to take part in the experiment. All of the students measured their initial pulse rates (in beats per minute). Then Mr. Cedarlund randomly assigned the students into two groups of 10. Each student in the first group drank 12 ounces of cola with caffeine. Each student in the second group drank 12 ounces of caffeine-free cola. All students then measured their pulse rates again. The table displays the change in pulse rate for the students in both groups.

	Change in pulse rate (Final pulse rate − Initial pulse rate)										Mean change	
Caffeine	8	3	5	1	4	0	6	1	4	0	3.2	$s_1 = 2.70$
No caffeine	3	−2	4	−1	5	5	1	2	−1	4	2.0	$s_2 = 2.62$

Construct and interpret a 95% confidence interval for the difference in true mean change in pulse rate for subjects like these who drink caffeine versus who drink no caffeine.

$\boxed{\text{STATE}}$ 95% CI for $\mu_1 - \mu_2$ where μ_1 = the true mean change in pulse rate for students like these after drinking 12 oz. of cola w/caffeine

and μ_2 = the true mean change in pulse rate for students like these after drinking 12 oz of caffeine free cola.

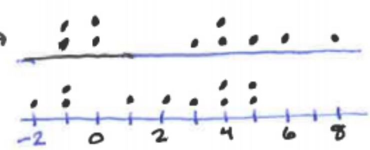
$\boxed{\text{PLAN}}$ Two-Sample t interval for $\mu_1 - \mu_2$

✓ Random
Randomly assigned to groups

✓ Large Counts :

caffeine →

No caff.



No strong skew / No outliers

$\boxed{\text{DO}}$ $\bar{x}_1 = 3.2$ $\quad \bar{x}_2 = 2$

$s_1 = 2.70$ $\quad s_2 = 2.62$

$n = 10$ $\quad n = 10$

2 - Samp T Int gives $(-1.302, 3.702)$ using $df = 17.986$

Stolen from calculator

Option 2

df = 9 $t^* = 2.262$

$$(3.2 - 2) \pm 2.262\sqrt{\frac{2.70^2}{10} + \frac{2.62^2}{10}}$$

$$1.2 \pm 2.691$$

$$(-1.491, 3.891)$$

CONCLUDE We are 95% confident that the interval from −1.302 to 3.702 beats per minute captures $\mu_1 - \mu_2$, the difference in the true mean change in pulse rate for for all students like these after drinking 12 oz of cola with caffeine versus w/o caffeine.

10.2.....~~37, 39, 41~~, 42, 45, 49

study pp. 645 - 654

don't forget that the take home LCQ 10.1 is due tomorrow at the beginning of class.