10.2 Comparing Two Means

(Day 1)

vocab

examples

Objectives

Polyester Demo

How quickly do synthetic fabrics such as polyester decay in landfills? A researcher buried polyester strips in the soil for different lengths of time, then dug up the strips and measured the force required to break them. Breaking strength is easy to measure and is a good indicator of decay. Lower strength means the fabric has decayed.

The researcher buried 10 strips of polyester fabric in well-drained soil in the summer. The strips were randomly assigned to two groups: 5 of them were buried for 2 weeks and the other 5 were buried for 16 weeks. Here are the breaking strengths in pounds

Group 1 (2 weeks):	118	126	126	120	129
Group 2 (16 weeks):	124	98	110	140	110

Polyester Demo

Group 1 (2 weeks):	118	126	126	120	129
Group 2 (16 weeks):	124	98	110	140	110

What is the mean breaking strength of group 1?

What is the mean breaking strength of group 2?

What is the difference in breaking strength (group 1 - group 2)?

Is it plausible that the difference is due to chance in random assignment rather than the treatments themselves?

Polyester Demo

Group 1 (2 weeks):	118	126	126	120	129
Group 2 (16 weeks):	124	98	110	140	110

* 10 cards - copy down each breaking strength onto a card
* Assign one partner to be group 1 and one partner to be group 2
* shuffle and give 5 cards to each partner
* find the means of each group
* find the difference of the means of the groups (group 1 - group 2)
* repeat for a total of 5 trials & dotplot results as a class

*how often did we exceed a mean difference of 7.4?

10.2 Comparing Two Means

(Day 1)

vocab

examples

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

SHAPE: When the population distributions are Normal, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ is Normal.

Or, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ will be approximately Normal if the sample sizes are large enough ($n_1 \geq 30$ and $n_2 \geq 30$).

CENTER: The mean of the sampling distribution is $\mu_1 - \mu_2$

SPREAD: standard deviation = $\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ as long as the 10% condition is met

Standard Error of Sample Means

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Two Sample z Statistic

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Two Sample t Statistic

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

Degrees of Freedom (df)

The formula is somewhat messy, so it is recommended to use a t-distribution on your calculator.

The other option is a conservative route. Take the smaller of n_1-1 and n_2-1.

Two-Sample t Interval for a Difference Between Means

Check 3 conditions:

Random

Normal: n ≥ 30, or plot data

Independent

If the conditions are met, then a C-Level confidence interval is constructed by:

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

Plastic or Plastic?

Do plastic bags from Target or plastic bags from Bashas hold more weight? A group of AP Statistics students decided to investigate by filling a random sample of 5 bags from each store with common grocery items until the bags ripped. Then they weighed the contents of items in each bag to determine its capacity. Here are their results, in grams:

Target:	12,572	13,999	11,215	15,447	10,896
Bashas:	9552	10,896	6983	8767	9972

Construct and interpret a 99% confidence interval for the difference in mean capacity of plastic grocery bags from Target and Bashas.

Plastic or Plastic?

Do plastic bags from Target or plastic bags from Bashas hold more weight? A group of AP Statistics students decided to investigate by filling a random sample of 5 bags from each store with common grocery items until the bags ripped. Then they weighed the contents of items in each bag to determine its capacity. Here are their results, in grams:

Target:	12,572	13,999	11,215	15,447	10,896
Bashas:	9552	10,896	6983	8767	9972

Construct and interpret a 99% confidence interval for the difference in mean capacity of plastic grocery bags from Target and Bashas.

STATE: We want to estimate $\mu_1 - \mu_2$ at a 99% C-level. μ_1 is mean capacity of Target Bags & μ_2 is mean capacity of Bashas bags.

Plastic or Plastic?

Do plastic bags from Target or plastic bags from Bashas hold more weight? A group of AP Statistics students decided to investigate by filling a random sample of 5 bags from each store with common grocery items until the bags ripped. Then they weighed the contents of items in each bag to determine its capacity. Here are their results, in grams:

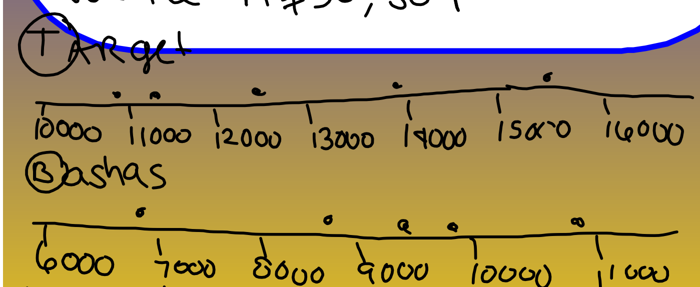
Target: 12,572 13,999 11,215 15,447 10,896
Bashas: 9552 10,896 6983 8767 9972

Construct and interpret a 99% confidence interval for the difference in mean capacity of plastic grocery bags from Target and Bashas.

PLAN: Random - yes

Independent - assuming at least 50 bags from each store

Normal - $n \not\geq 30$, so plot data



No outliers or strong skew so we use 2 sample T interval for $\mu_1 - \mu_2$

Plastic or Plastic?

Do plastic bags from Target or plastic bags from Bashas hold more weight? A group of AP Statistics students decided to investigate by filling a random sample of 5 bags from each store with common grocery items until the bags ripped. Then they weighed the contents of items in each bag to determine its capacity. Here are their results, in grams:

Target:	12,572	13,999	11,215	15,447	10,896
Bashas:	9552	10,896	6983	8767	9972

Construct and interpret a 99% confidence interval for the difference in mean capacity of plastic grocery bags from Target and Bashas.

DO: 2-SampTInt

df = 7.51 $\bar{x}_1 = 12025.8$ $S_{x1}: 1912.48$

$\bar{x}_2 = 9234$ $S_{x2}: 1474.2$

(-100.9, 7284.5)

Plastic or Plastic?

Do plastic bags from Target or plastic bags from Bashas hold more weight? A group of AP Statistics students decided to investigate by filling a random sample of 5 bags from each store with common grocery items until the bags ripped. Then they weighed the contents of items in each bag to determine its capacity. Here are their results, in grams:

Target:	12,572	13,999	11,215	15,447	10,896
Bashas:	9552	10,896	6983	8767	9972

Construct and interpret a 99% confidence interval for the difference in mean capacity of plastic grocery bags from Target and Bashas.

CONCLUDE: We are 99% confident that the true mean difference in capacity of target & bashas bags falls between -100.9 g and 7284.5 g.

Bellwork

High levels of cholesterol in the blood are associated with higher risk of heart attacks. Will using a drug to lower blood cholesterol reduce heart attacks? The Helsinki Heart Study recruited middle-aged men with high cholesterol but no history of other serious medical problems to investigate this question. The volunteer subjects were assigned at random to one of two treatments: 2051 men took the drug gemfibrozil to reduce their cholesterol levels, and a control group of 2030 men took a placebo. During the next five years, 56 men in the gemfibrozil group and 84 men in the placebo group had heart attacks. Is the apparent benefit of gemfibrozil statistically significant? Perform an appropriate test to find out.

Use $H_0: p_1 - p_2 = 0$ and $H_a: p_1 - p_2 < 0$
where p_1 is the actual heart attack rate for middle-aged men like the ones in this study who take gemfibrozil, and p_2 is the actual heart attack rate for middle-aged men like the ones in this study who take only a placebo.
(Show DO & CONCLUDE steps)

Con: $.0068 \leq .05 \rightarrow$ We reject the null $\rightarrow$ We can conclude that the difference in prop. of those who took gem. & the placebo is less than 0. Therefore, the apparent benefit is stat. sig.

10.2 Comparing Two Means

(Day 2)

vocab

examples

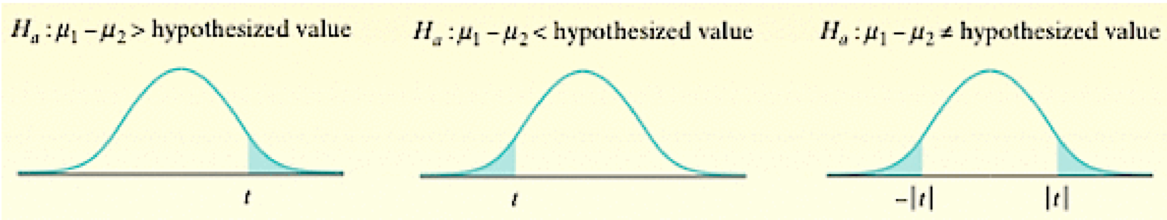
Two Sample t Test for the Difference between two means

Suppose the Random, Normal, and Independent conditions are met. To test the hypothesis H_0: $\mu_1 - \mu_2$ = hypothesized value, compute the two-sample t statistic:

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

Two Sample t Test for the Difference between two means

Check the Random, Normal, and Independent conditions



Use the 4 step process

Stronger Picker Upper

In commercials for Bounty paper towels, the manufacturer claims that they are the "quicker picker-upper." But are they also the stronger picker upper? Two AP Statistics students, decided to find out. They selected a random sample of 30 Bounty paper towels and a random sample of 30 generic paper towels and measured their strength when wet. To do this, they uniformly soaked each paper towel with 4 ounces of water, held two opposite edges of the paper towel, and counted how many quarters each paper towel could hold until ripping, alternating brands. For these data, $\bar{x}_B = 117.6$, $s_B = 6.64$, $\bar{x}_G = 88.1$, and $s_G = 6.30$. Is there convincing evidence that wet Bounty paper towels can hold more weight, on average, than wet generic paper towels?

State: $H_0: \mu_B - \mu_G = 0$ $\alpha = .05$
$H_a: \mu_B - \mu_G > 0$
μ_B is the true avg. # of quarters the Bounty can hold. μ_G is the true avg. # of quarters the generic can hold.

Stronger Picker Upper

In commercials for Bounty paper towels, the manufacturer claims that they are the "quicker picker-upper." But are they also the stronger picker upper? Two AP Statistics students, decided to find out. They selected a random sample of 30 Bounty paper towels and a random sample of 30 generic paper towels and measured their strength when wet. To do this, they uniformly soaked each paper towel with 4 ounces of water, held two opposite edges of the paper towel, and counted how many quarters each paper towel could hold until ripping, alternating brands. For these data, $x_B = 117.6$, $s_B = 6.64$, $x_G = 88.1$, and $s_G = 6.30$. Is there convincing evidence that wet Bounty paper towels can hold more weight, on average, than wet generic paper towels?

PLAN: Random: yes
Independent: Assuming at least ~300 sheets of each brand ✓
Normal: $n_B \geq 30$ $n_G \geq 30$
We will use a 2-sample t-test for $\mu_B - \mu_G$

Stronger Picker Upper

In commercials for Bounty paper towels, the manufacturer claims that they are the "quicker picker-upper." But are they also the stronger picker upper? Two AP Statistics students, decided to find out. They selected a random sample of 30 Bounty paper towels and a random sample of 30 generic paper towels and measured their strength when wet. To do this, they uniformly soaked each paper towel with 4 ounces of water, held two opposite edges of the paper towel, and counted how many quarters each paper towel could hold until ripping, alternating brands. For these data, $\bar{x}_B = 117.6$, $s_B = 6.64$, $\bar{x}_G = 88.1$, and $s_G = 6.30$. Is there convincing evidence that wet Bounty paper towels can hold more weight, on average, than wet generic paper towels?

DO: 2-SampTTest

$\bar{x}_B$: 117.6 Sx_B: 6.64 n_B: 30 df = 57.9

$\bar{x}_G$: 88.1 Sx_G: 6.3 n_G: 30

$\mu_1 > \mu_2$ t = 17.65 p-value ≈ 0

Stronger Picker Upper

In commercials for Bounty paper towels, the manufacturer claims that they are the "quicker picker-upper." But are they also the stronger picker upper? Two AP Statistics students, decided to find out. They selected a random sample of 30 Bounty paper towels and a random sample of 30 generic paper towels and measured their strength when wet. To do this, they uniformly soaked each paper towel with 4 ounces of water, held two opposite edges of the paper towel, and counted how many quarters each paper towel could hold until ripping, alternating brands. For these data, $\bar{x}_B = 117.6$, $s_B = 6.64$, $\bar{x}_G = 88.1$, and $s_G = 6.30$. Is there convincing evidence that wet Bounty paper towels can hold more weight, on average, than wet generic paper towels?

Conclude: p-value ≈ 0 < .05 → Reject Ho → We can conclude that the true avg. # of quarters that a bounty paper towel can hold is greater than the true avg. # of quarters that a generic p.t. can hold. The results are stat. sig. @ 5% level.

Calcium and Blood Pressure

Does increasing the amount of calcium in our diet reduce blood pressure? Examination of a large sample of people revealed a relationship between calcium intake and blood pressure. The relationship was strongest for men. Such observational studies do not establish causation. Researchers therefore designed a randomized comparative experiment.

The subjects were 21 healthy men who volunteered to take part in the experiment. They were randomly assigned to two groups: 10 of the men received a calcium supplement for 12 weeks, while the control group of 11 men received a placebo pill that looked identical. The experiment was double-blind. The response variable is the decrease in systolic (top number) blood pressure for a subject after 12 weeks, in millimeters of mercury. An increase appears as a negative response. Here are the data:

Group 1 (calcium):	7	−4	18	17	−3	−5	1	10	11	−2	
Group 2 (placebo):	−1	12	−1	−3	3	−5	5	2	−11	−1	−3

Calcium and Blood Pressure

Group 1 (calcium):	7	−4	18	17	−3	−5	1	10	11	−2	
Group 2 (placebo):	−1	12	−1	−3	3	−5	5	2	−11	−1	−3

STATE: H_0: $\mu_1 - \mu_2 = 0$ H_a: $\mu_1 - \mu_2 > 0$ $\alpha = .05$

μ_1 is the true average decrease in systolic blood pressure of mean blood pressure of men taking calcium. μ_2 is the true average decrease in systolic blood pressure of men taking placebo

Calcium and Blood Pressure

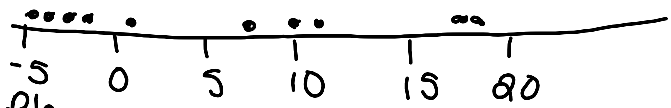
Group 1 (calcium):	7	−4	18	17	−3	−5	1	10	11	−2	
Group 2 (placebo):	−1	12	−1	−3	3	−5	5	2	−11	−1	−3

PLAN: Random - random assignment

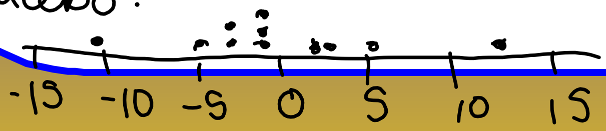
Independent - assuming at least 210 men

Normal - plot data since n_1 & n_2 are not at least 30

calcium:



Placebo:



Since no strong skew or outliers, we use a two-sample t test for $\mu_1 - \mu_2$

Calcium and Blood Pressure

Group 1 (calcium):	7	−4	18	17	−3	−5	1	10	11	−2	
Group 2 (placebo):	−1	12	−1	−3	3	−5	5	2	−11	−1	−3

DO: 2 samp T Test

$\mu 1 > \mu 2$

df = 15.59

x1: 5 sx1: 8.74 n1: 10

x2: -.27 sx2: 5.9 n2: 11

t = 1.6

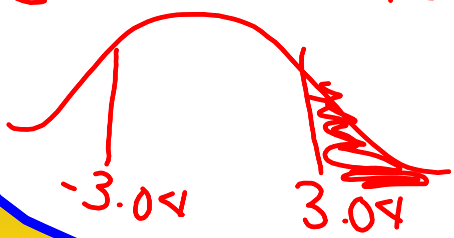
p-value = .0644

Calcium and Blood Pressure

Group 1 (calcium):	7	−4	18	17	−3	−5	1	10	11	−2	
Group 2 (placebo):	−1	12	−1	−3	3	−5	5	2	−11	−1	−3

CONCLUDE: .0644 ≥ .05 --> We fail to reject H_0 --> We can NOT conclude that the difference between the true mean calcium b.p. and the true mean placebo b.p. is greater than 0. There is not convincing evidence that calcium helps lower blood pressure. The results are NOT statistically significant at the 5% level.

$t = 3.04$ tcdf(3.04, 1E99, df)



10.2 Comparing Two Means

(Day 3)

vocab

examples

Pooled Estimator of σ²

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

benefits: gives larger df, which will increase power

drawbacks: may give you a p-value that is off depending on the data set

Two Sample t Procedures

The two-sample t procedures are quite robust against departures from Normality, especially when both sample/group sizes are large.

Two Sample vs Paired

Paired must be matched with same person, or like subject. One way to think of this is that when you have your data in lists, the paired subjects must be matched side by side.

Two sample has 2 unique samples and it does not matter what order you collect the data in. For example, we can scramble the data in a list without making a difference in our calculations.

Paired or Two-Sample?

In each of the following settings, decide whether you should use paired t procedures or two-sample t procedures to perform inference. Explain your choice.

(a) To test the wear characteristics of two tire brands, A and B, one brand A tire is mounted on one side of each car in the rear, while a Brand B tire is mounted on the other side. Which side gets which brand is determined by flipping a coin. The same procedure is used on the front.

(b) Can listening to music while working increase productivity? Twenty factory workers agree to take part in a study to investigate this question. Researchers randomly assign 10 workers to do a repetitive task while listening to music and the other 10 workers to do the task in silence.

Testing w/ Distractions

Suppose you are designing an experiment to determine if students perform better on tests when there are no distractions, such as a teacher talking on the phone. You have access to two classrooms and 30 volunteers who are willing to participate in your experiment.

(a) Design an experiment so that a two-sample t test would be the appropriate inference method.

(b) Design an experiment so that a paired t test would be the appropriate inference method.

(c) Which experimental design is better? Explain.

Which is Which?

	One Sample Z Interval/Test	One Sample T Interval/Test	Two Sample Z Interval/Test	Two Sample T Interval/Test	Paired T or Z Test/Interval
Means		*Use when you have 1 set of data*		*Use when you have 2 sets of data (not paired)*	*Subtract to create one list of data, then use one sample t test/interval*
Proportions	*Use when you have 1 set of data*		*Use when you have 2 sets of data (not paired)*		*Subtract to create one list of data, then use one sample z test/interval*