I am giving tests
Per 2, 3, and 4
- that is why the desks are in rows.

- Sit wherever you like today.

preferably not toward the back.

Today

What happens when we have paired data?

Construct Confidence Intervals

Perform a significance test

10.3 Day 1

When to use differences of two means (or p) and when to use procedures for paired data?

10.3

Day 2

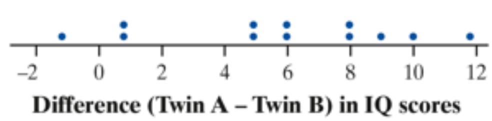
A researcher studied a random sample of identical twins who had been separated and adopted at birth. In each case, one twin (Twin A) was adopted by a high-income family and the other (Twin B) by a low-income family. Both twins were given an IQ test as adults.

Paired data result from recording two values of the same quantitative variable for each individual or for each pair of similar individuals.

Pair	1	2	3	4	5	6	7	8	9	10	11	12
Twin A's IQ (high-income family)	128	104	108	100	116	105	100	100	103	124	114	112
Twin B's IQ (low-income family)	120	99	99	94	111	97	99	94	104	114	113	100
Difference (A − B)	8	5	9	6	5	8	1	6	−1	10	1	12

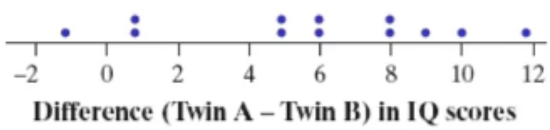
The *mean difference* is $\bar{x}_{diff}$ = 5.833 points

s_{diff} = 3.93 points



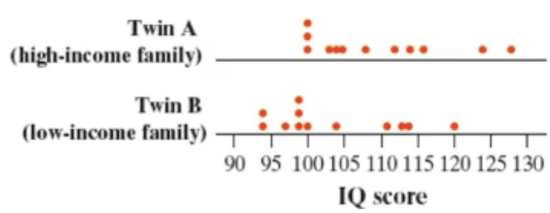
Proper analysis looks at the differences
(not as two separate populations)

YES!



$$\bar{x}_{\text{diff}} = \bar{x}_{\text{A}-\text{B}} = \frac{8+5+9+\cdots+1+12}{12} = \frac{70}{12} = 5.833 \text{ points}$$

NO!



YES!

Difference (Twin A – Twin B) in IQ scores

$$\bar{x}_{\text{diff}} = \bar{x}_{A-B} = \frac{8+5+9+\cdots+1+12}{12} = \frac{70}{12} = 5.833 \text{ points}$$

therefore conditions for inference have to focus on the difference

i.e. μ_{diff}

Difference (Twin A – Twin B) in IQ scores

$$\bar{x}_{\text{diff}} = \bar{x}_{A-B} = \frac{8+5+9+\cdots+1+12}{12} = \frac{70}{12} = 5.833 \text{ points}$$

therefore conditions for inference have to focus on the difference

→ for Confidence Intervals

Calculations are just a one-sample t interval for a mean (also called a paired t interval)

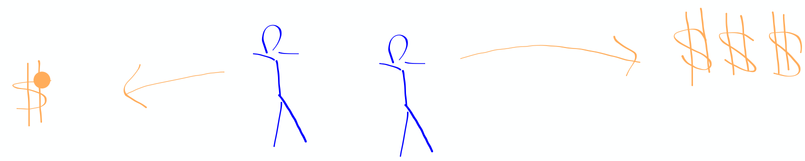
i.e. μ_{diff}

Interesting to Note:

Because a random sample of twins was selected, the results can be generalized to all identical twins

but..

Researches could not infer a cause and effect relationship between family income and twins' IQ because random assignment was not passsible



There are two ways that a statistical study involving a single quantitative variable can yield paired data:

1. Researchers can record two values of the variable for each individual. (experiment investigating whether music helps or hinders learning)

2. The researcher can form pairs of similar individuals and record the value of the variable once for each individual. (observational study of identical twins' IQ scores)

Analyzing Paired Data

To analyze paired data, start by computing the difference for each pair.

Then make a graph of the differences. Use the mean difference $\bar{x}_{diff}$ and thestandard deviation of the differences s_{diff} as summary statistics.

Lesson 10.3 (Day 1) Comparing Two Means: Paired Data

A teacher hoped to determine if the two versions of her AP® Statistics final exam were equally difficult. Last year, she set up a randomized comparative experiment, and the data showed no statistically significant difference in the scores between the two versions. This year, she selected a random sample of 20 students from the large district and used a **matched pairs** design, where each student was given both versions of the test. For each student, the test order was randomized. Here are the scores, along with the differences in scores for each student:

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Version A	90	85	77	96	88	83	93	81	84	71	65	98	93	83	85	91	82	79	74	72
Version B	86	84	80	91	83	80	87	83	80	70	60	99	96	75	78	88	81	74	65	66
Difference (A − B)	4	1	−3	5	5	3	6	−2	4	1	5	−1	−3	8	7	3	1	5	9	6

(a) Make a dotplot of the difference (Version A – Version B) in scores for each student.

handout

In case you forgot

Matched Pairs design is a common experimental design for comparing two treatments that uses blocks of size 2

In some matched pairs designs, two very similar experimental units are paired and the two treatments are randomly assigned within each pair.

In others, each experimental unit (students in this case) receives both treatments (both tests in this case) in a random order.

Lesson 10.3 (Day 1) Comparing Two Means: Paired Data

A teacher hoped to determine if the two versions of her AP® Statistics final exam were equally difficult. Last year, she set up a randomized comparative experiment, and the data showed no statistically significant difference in the scores between the two versions. This year, she selected a random sample of 20 students from the large district and used a **matched pairs** design, where each student was given both versions of the test. For each student, the test order was randomized. Here are the scores, along with the differences in scores for each student:

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Version A	90	85	77	96	88	83	93	81	84	71	65	98	93	83	85	91	82	79	74	72
Version B	86	84	80	91	83	80	87	83	80	70	60	99	96	75	78	88	81	74	65	66
Difference (A – B)	4	1	–3	5	5	3	6	–2	4	1	5	–1	–3	8	7	3	1	5	9	6

(a) Make a dotplot of the difference (Version A – Version B) in scores for each student.

one-sample scenario

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Version A	90	85	77	96	88	83	93	81	84	71	65	98	93	83	85	91	82	79	74	72
Version B	86	84	80	91	83	80	87	83	80	70	60	99	96	75	78	88	81	74	65	66
Difference (A – B)	4	1	–3	5	5	3	6	–2	4	1	5	–1	–3	8	7	3	1	5	9	6

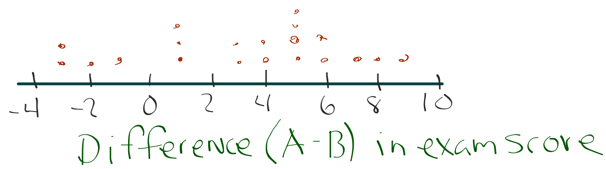
(a) Make a dotplot of the difference (Version A – Version B) in scores for each student.

(b) Describe what the graph reveals about whether Version A is harder or easier than Version B.

(c) Calculate the mean difference and the standard deviation of the differences. Interpret the mean difference.

Student	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Version A	90	85	77	96	88	83	93	81	84	71	65	98	93	83	85	91	82	79	74	72
Version B	86	84	80	91	83	80	87	83	80	70	60	99	96	75	78	88	81	74	65	66
Difference (A − B)	4	1	−3	5	5	3	6	−2	4	1	5	−1	−3	8	7	3	1	5	9	6

(a) Make a dotplot of the difference (Version A – Version B) in scores for each student.



(b) Describe what the graph reveals about whether Version A is harder or easier than Version B.

There is some evidence that Version A was easier than Version B. 16 out of the 20 did better on A.

(c) Calculate the mean difference and the standard deviation of the differences. Interpret the mean difference.

$\bar{x}_{diff} = \bar{x}_{A-B} = 3.2$

$S_{diff} = 3.533$

Confidence Intervals for μ_{diff}

Conditions for Constructing a Confidence Interval About a Mean Difference

Random: Paired data come from a random sample from the population of interest or from a randomized experiment.

why?

10%: When sampling without replacement, $n_{\text{diff}} < 0.10N_{\text{diff}}$.

Normal/Large Sample: The population distribution of differences (or the true distribution of differences in response to the treatments) is Normal **or the number of differences in the sample is large ($n_{\text{diff}} \geq 30$).** If the population (true) distribution of differences has unknown shape and the number of differences in the sample is less than 30, a graph of the sample differences shows no strong skewness or outliers.

So we can generalize to the population (of all pairs)

or

so we can show causation in an experiment (random assignment)

Conditions for Constructing a Confidence Interval About a Mean Difference
Random: Paired data come from a random sample from the population of interest or from a randomized experiment.
10%: When sampling without replacement, $n_{diff} < 0.10N_{diff}$.
Normal/Large Sample: The population distribution of differences (or the true distribution of differences in response to the treatments) is Normal **or the number of differences in the sample is large ($n_{diff} \geq 30$).** If the population (true) distribution of differences has unknown shape and the number of differences in the sample is less than 30, a graph of the sample differences shows no strong skewness or outliers.

So that sampling without replacement is ok

(not needed in many experiments)

Conditions for Constructing a Confidence Interval About a Mean Difference
Random: Paired data come from a random sample from the population of interest or from a randomized experiment.
10%: When sampling without replacement, $n_{diff} < 0.10N_{diff}$.
Normal/Large Sample: The population distribution of differences (or the true distribution of differences in response to the treatments) is Normal **or the number of differences in the sample is large ($n_{diff} \geq 30$).** If the population (true) distribution of differences has unknown shape and the number of differences in the sample is less than 30, a graph of the sample differences shows no strong skewness or outliers.

so the sampling distribution of $\bar{x}_{diff}$ will be approximately Normal, and we can use t^* to do calculations.

Conditions for Constructing a Confidence Interval About a Mean Difference

Random: Paired data come from a random sample from the population of interest or from a randomized experiment.

10%: When sampling without replacement, $n_{diff} < 0.10N_{diff}$.

Normal/Large Sample: The population distribution of differences (or the true distribution of differences in response to the treatments) is Normal **or the number of differences in the sample is large ($n_{diff} \geq 30$).** If the population (true) distribution of differences has unknown shape and the number of differences in the sample is less than 30, a graph of the sample differences shows no strong skewness or outliers.

↑ Same for Significance Tests

One-Sample *t* Interval for a Mean Difference (Paired *t* Interval for a Mean Difference)

When the conditions are met, a *C*% confidence interval for μ_{diff} is

$$\bar{x}_{diff} \pm t^* \frac{s_{diff}}{\sqrt{n_{diff}}}$$

where t^* is the critical value with *C*% of the area between $-t^*$ and t^* for the *t* distribution with $n_{diff} - 1$ degrees of freedom.

Same as in Ch.9

FINAL EXAM REVISITED

Final exam revisited *Confidence interval for a mean difference*
In the preceding alternate example, a teacher used a matched pairs design to collect data on scores for two different versions of her final exam for 20 students. Recall that $\bar{x}_{diff} = \bar{x}_{A-B} = 3.2$ and $s_{diff} = 3.533$. Construct and interpret a 99% confidence interval for the true mean difference (Version A – Version B) in final exam scores for AP® Statistics students in this district.

for the plan
you can refer
to the dot plot
already made.

STATE 99% CI for μ_{DIFF}

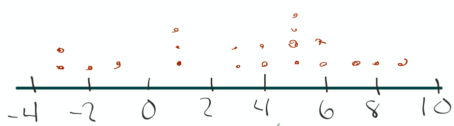
= true mean difference (Version A - Vers. B) in the Final Exam scores in the DISTRICT.

PLAN One-Sample t interval for μ_{diff}

Random Rand. sample of 20 AP students

10% Assume 20 < 10% of all students in large district.

Normal/Large Sample



Difference (A-B) in exam score

Dot plot doesn't show strong skewness or any outliers.

DO

With 99% confidence and df = 20-1 = 19

$t^* = 2.861$

$$3.2 \pm 2.861 \cdot \frac{3.533}{\sqrt{20}}$$

$$= (0.94, 5.46)$$

CONCLUDE

We are 99% confident that the interval from 0.94 to 5.46 captures the true mean difference (Version A - Ver. B) in final exam scores for AP Stats in this district.

Signif. tests for μ_{diff}

Effects of Caffeine Withdrawal

$$\text{stand. test statistic} = \frac{\text{statistic} - \text{parameter}}{\text{std. dev. of stat}}$$

$$t = \frac{\bar{x}_{diff} - \mu_{diff}}{\frac{s_{diff}}{\sqrt{n_{diff}}}}$$

Significance Tests for Mean Difference (μ_{diff})

Researchers designed an experiment to study the effects of caffeine withdrawal. They recruited 11 volunteers who were diagnosed as being caffeine dependent to serve as subjects. Each subject was barred from coffee, colas, and other substances with caffeine for the duration of the experiment. During one 2-day period, subjects took capsules containing their normal caffeine intake. During another 2-day period, they took placebo capsules. The order in which subjects took caffeine and the placebo was randomized. At the end of each 2-day period, a test for depression was given to all 11 subjects. Researchers wanted to know whether being deprived of caffeine would lead to an increase in depression.

The table below contains data on the subjects' scores on the depression test. Higher scores show more symptoms of depression.

Subject	1	2	3	4	5	6	7	8	9	10	11
Depression (caffeine)	5	5	4	3	8	5	0	0	2	11	1
Depression (placebo)	16	23	5	7	14	24	6	3	15	12	0

Do the data provide convincing evidence at the $\alpha = 0.05$ significance level that caffeine withdrawal increases depression score, on average, for subjects like the ones in this experiment?

Consider using a histogram on GDC

State $H_0: \mu_{diff} = 0$ $H_a: \mu_{diff} > 0$

where μ_{diff} = true mean difference (Placebo − Caffeine) in depression test score for subjects like these.

We'll use $\alpha = 0.05$

Or $H_0: \mu_{diff} = 0$

$H_a: \mu_{diff} < 0$

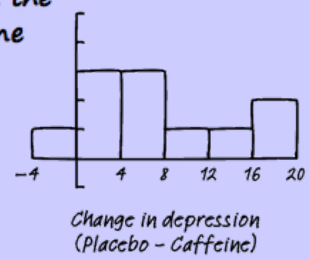
if using (Caffeine − Placebo)

PLAN

PLAN *Paired t test for μ_{diff}*

- *Random: Researchers randomly assigned the treatments— placebo then caffeine, caffeine then placebo—to the subjects.✓*

- *Normal/Large Sample:*
The sample size is small, but the histogram of differences doesn't show any outliers or strong skewness. ✓



DO

$\bar{x}_{diff} = 7.364$ $S_{diff} = 6.918$

$n_{diff} = 11$

$$t = \frac{7.364 - 0}{\frac{6.918}{\sqrt{11}}} = 3.53$$

P-Value $df = 11 - 1 = 10$

$$t_{cdf} = (\underset{\text{lower}}{3.53}, \underset{\text{Upper}}{1000}, \underset{\text{df}}{10}) = .0027$$

CONCLUDE

Sice the P-Value of 0.0027 $< \alpha = .05$
we reject Ho

We have convincing evidence that caffeine withdrawal increases depression test scores, on average, for subjects like the ones in the study.

Let's read bottom of p. 681

LCQ

Tyler - Michelle

Brita - Lina

Carson-Anna

August - Maia

Kendra - Jackson

Aidan - Madison - Sarah

Laura - Kai

Natalie - Zeke

10.3........75, 79, 85

+ LCQ

10.2