<u>Friday, February 3</u>: 9.1 Significance Tests

According to an article in the San Gabriel Valley Tribune (2-13-03), "Most people are kissing the 'right way'." That is, according to the study, the majority of couples tilt their heads to the right when kissing. In the study, a researcher observed 124 couples kissing in various public places and found that 83/124 ($\hat{p}$ = 0.669) of the couples tilted to the right when kissing. Is this *convincing* evidence that the true proportion p of *all* couples that kiss the right way is greater than 0.50?

Give two explanations for why the sample proportion was above 0.5.

How can we decide which of the two explanations is more plausible?

Read 529-530

What is the basic idea of a significance test?

Read 531-532

What is the difference between a null and an alternative hypothesis? What notation is used for each? What are some common mistakes when stating hypotheses?

What is the difference between and one-sided and a two-sided alternative hypothesis? How can you decide which to use?

Alternate Example: *A better golf club?*
Mike is an avid golfer who would like to improve his play. A friend suggests getting new clubs and lets Mike try out his 7-iron. Based on years of experience, Mike has established that the mean distance that balls travel when hit with his old 7-iron is μ = 175 yards with a standard deviation of σ = 15 yards. He is hoping that this new club will make his shots with a 7-iron more consistent (less variable), and so he goes to the driving range and hits 50 shots with the new 7-iron.
(a) Describe the parameter of interest in this setting.
(b) State appropriate hypotheses for performing a significance test.

HW #15: page 546 (1–9 odd)

<u>Monday, February 6</u>: 9.1 *P*-values and Conclusions

Read 533-534

What is a *P*-value? What does a *P*-value measure?

In the kissing example, the *P*-value = $P(\hat{p} \geq 0.669 \mid p = 0.5)$ = ____. Interpret this value.

Alternate Example: *A better golf club?*
When Mike was testing a new 7-iron, the hypotheses were $H_0: \sigma = 15$ versus $H_a: \sigma < 15$ where σ = the true standard deviation of the distances Mike hits golf balls using the new 7-iron. Based on 50 shots with the new 7-iron, the standard deviation was s_x = 10.9 yards. A significance test using the sample data produced a *P*-value of 0.002.
(a) Interpret the *P*-value in this context.
(b) Do the data provide convincing evidence against the null hypothesis? Explain.

Read 534-537

What are the two possible conclusions for a significance test?

What are some common errors that students make in their conclusions?

What is a significance level? When are the results of a study statistically significant?

Alternate example: *Tasty chips*
For his second semester project in AP Statistics, Zenon decided to investigate whether students at his school prefer name-brand potato chips to generic potato chips. He randomly selected 50 students and had each student try both types of chips, in random order. Overall, 34 of the 50 students preferred the name-brand chips. Zenon performed a significance test using the hypotheses $H_0: p = 0.5$ versus $H_a: p > 0.5$ where p = the true proportion of students at his school who prefer name-brand chips. The resulting P-value was 0.0055. What conclusion would you make at each of the following significance levels?
(a) $\alpha = 0.01$ (b) $\alpha = 0.001$

What should be considered when choosing a significance level?

HW #16: page 546 (2–10 even, 11–17 odd)

<u>Tuesday, February 7</u>: 9.1 Errors in Significance Testing

Read 538-542

In a jury trial, what two errors could a jury make?

In a significance test, what two errors can we make? Which error is worse?

Alternate Example: *Faster fast food?*
The manager of a fast-food restaurant want to reduce the proportion of drive-through customers who have to wait more than two minutes to receive their food once their order is placed. Based on store records, the proportion of customers who had to wait at least two minutes was p = 0.63. To reduce this proportion, the manager assigns an additional employee to assist with drive-through orders. During the next month the manager will collect a random sample of drive-through times and test the following hypotheses: $H_0: p = 0.63$ versus $H_a: p < 0.63$ where p = the true proportion of drive-through customers who have to wait more than two minutes after their order is placed to receive their food. Describe a Type I and a Type II error in this setting and explain the consequences of each.

What is the probability of a Type I error? What can we do to reduce the probability of a Type I error? Are there any drawbacks to this?

What the power of a test? How is power related to the probability of a Type II error? Will you be expected to calculate the power of a test on the AP exam?

What four factors affect the power of a test? Why does this matter?

Read 542-545

HW #17: page 547 (19–25 odd)

Wednesday/Thursday, February 8/9: 9.2 Significance Tests for a Population Proportion

What are the three conditions for conducting a significance test for a population proportion? How are these different than the conditions for constructing a confidence interval for a population proportion?

What is a test statistic? What does it measure? Is the formula on the formula sheet?

Read 552–556

What are the four steps for conducting a significance test? What is required in each step?

What test statistic is used when testing for a population proportion? Is this on the formula sheet?

What happens when the data don't support H_a?

Can you use your calculator for the Do step? Are there any drawbacks to this method?

Alternate Example: *Better to be last?*
On shows like *American Idol*, contestants often wonder if there is an advantage to performing last. To investigate this, a random sample of 600 *American Idol* fans is selected, and they are shown the audition tapes of 12 never-before-seen contestants. For each fan, the order of the 12 videos is randomly determined. Thus, if the order of performance doesn't matter, we would expect approximately 1/12 of the fans to prefer the last contestant they view. In this study, 59 of the 600 fans preferred the last contestant they viewed. Do these data provide convincing evidence that there is an advantage to going last?

HW #18: page 548 (26–30), page 562 (41–45 odd)

<u>Friday, February 10</u>: 9.2 Two-sided tests for a proportion

Read 556–557

Alternate Example: *Benford's law and fraud*
When the accounting firm AJL and Associates audits a company's financial records for fraud, they often use a test based on Benford's law. Benford's law states that the distribution of first digits in many real-life sources of data is not uniform. In fact, when there is no fraud, about 30.1% of the numbers in financial records begin with the digit 1. However, if the proportion of first digits that are 1 is significantly different from 0.301 in a random sample of records, AJL and Associates does a much more thorough investigation of the company. Suppose that a random sample of 300 expenses from a company's financial records results in only 68 expenses that begin with the digit 1. Should AJL and Associates do a more thorough investigation of this company?

Describe a Type I and Type II error in this context.

Read 558–560

Can you use confidence intervals to decide between two hypotheses? What is an advantage to using confidence intervals for this purpose? Why don't we always use confidence intervals?

Alternate Example: *Benford's law and fraud*
(a) Find and interpret a confidence interval for the true proportion of expenses that begin with the digit 1 for the company in the previous Alternate Example.
(b) Use your interval from (a) to decide whether this company should be investigated for fraud.

HW #19: page 562 (47–55 odd)

<u>Monday, February 13</u>: 9.3 Significance Tests for a Population Mean

Read 565–566

What are the three conditions for conducting a significance test for a population mean?

Alternate Example: *Less music?*
A classic rock radio station claims to play an average of 50 minutes of music every hour. However, it seems that every time you turn to this station, there is a commercial playing. To investigate their claim, you randomly select 12 different hours during the next week and record what the radio station plays in each of the 12 hours. Here are the number of minutes of music in each of these hours:

44 49 45 51 49 53 49 44 47 50 46 48

Check the conditions for carrying out a significance test of the company's claim that it plays an average of 50 minutes of music per hour.

Read 567–569

What test statistic do we use when testing a population mean? Is the formula on the formula sheet?

How do you calculate *P*-values using the *t* distributions?

Alternate Examples:
(a) Calculate the test statistic and *P*-value for the Less Music example

(b) Find the *P*-value for a test of $H_0: \mu = 10$ versus $H_a: \mu > 10$ that uses a sample of size 75 and has a test statistic of $t = 2.33$.

Read 570–573

Alternate Example: *Construction zones*

Every road has one at some point—construction zones that have much lower speed limits. To see if drivers obey these lower speed limits, a police officer used a radar gun to measure the speed (in miles per hour, or mph) of a random sample of 10 drivers in a 25 mph construction zone. Here are the results:

27	33	32	21	30	30	29	25	27	34

(a) Can we conclude that the average speed of drivers in this construction zone is greater than the posted 25 mph speed limit?

(b) Given your conclusion in part (a), which kind of mistake—a Type I or a Type II error—could you have made? Explain what this mistake means in this context.

HW #20 page 564 (57–60), page 587 (63, 67, 73)

Tuesday, February 14: 9.3 Two-sided tests for μ

Read 574–576

Alternate Example: *Don't break the ice*

Collection 1

Width	29.4943 mm 50 0.0877121 mm 0.0124044 mm

S1 = mean ( )
S2 = count ( )
S3 = stdDev ( )
S4 = stdError ( )

In the children's game Don't Break the Ice, small plastic ice cubes are squeezed into a square frame. Each child takes turns tapping out a cube of "ice" with a plastic hammer, hoping that the remaining cubes don't collapse. For the game to work correctly, the cubes must be big enough so that they hold each other in place in the plastic frame but not so big that they are too difficult to tap out. The machine that produces the plastic cubes is designed to make cubes that are 29.5 millimeters (mm) wide, but the actual width varies a little. To ensure that the machine is working well, a supervisor inspects a random sample of 50 cubes every hour and measures their width. The Fathom output summarizes the data from a sample taken during one hour.

(a) Interpret the standard deviation and the standard error provided by the computer output.
(b) Do these data give convincing evidence that the mean width of cubes produced this hour is not 29.5 mm? Use a significance test with α = 0.05 to find out.
(c) Calculate a 95% confidence interval for μ. Does your interval support your decision from part (b)?

HW #21: page 588 (69, 79, 81, 83, 85)

<u>Wednesday/Thursday, February 15/16</u>: 9.3 Significance Tests for Paired Data

Read 577–580

Alternate Example: *Is the express lane faster?*
For their second semester project in AP Statistics, Libby and Kathryn decided to investigate which line was faster in the supermarket: the express lane or the regular lane. To collect their data, they randomly selected 15 times during a week, went to the same store, and bought the same item. However, one of them used the express lane and the other used a regular lane. To decide which lane each of them would use, they flipped a coin. If it was heads, Libby used the express lane and Kathryn used the regular lane. If it was tails, Libby used the regular lane and Kathryn used the express lane. They entered their randomly assigned lanes at the same time, and each recorded the time in seconds it took them to complete the transaction. Carry out a test to see if there is convincing evidence that the express lane is faster.

Time in express lane (seconds)	**Time in regular lane (seconds)**
337	342
226	472
502	456
408	529
151	181
284	339
150	229
357	263
349	332
257	352
321	341
383	397
565	694
363	324
85	127

Read 581–585

HW #22: page 588 (75, 77, 89, 94–104)

Friday, February 17: Review Chapter 9 / FRAPPY

HW #23: page 594 Chapter 9 Review Exercises

Monday, February 20: Review Chapter 9 / Review for Midterm (Tuesday schedule due to Rodeo)

HW #24: page 597 Chapter 9 AP Statistics Practice Test

Tuesday/Wednesday, February 21/22: Chapter 9 Test (Wed/Thurs schedule due to Rodeo)

Monday, February 27: Begin Chapter 10

Tuesday/Wednesday, February 28/29: Continue Chapter 10 (late-in schedule for AIMS)

Thursday, March 1: Review for Midterm (Friday schedule) (sub)

Friday, March 2: Midterm (sub)

Problem Set for Midterm: From the following questions I will choose 3 for you to answer on the midterm, along with a mysterious 4th question. I will not be collecting these problems, but you are welcome to ask me about them at the end of class (when there is time), during tutorial, or afterschool. The rubrics for these questions can be found at the following website: apcentral.collegeboard.com/stats (click on AP Statistics Exam information). Good Luck!

1. 1999 #1 (Commercial Aircraft)
2. 2003 #2 (Class action lawsuits)
3. 2003B #2 (Age vs. Income)
4. 2004 #3 (Brontosaurs)
5. 2004B #3 (Bauxite cars)
6. 2006B #1 (Monthly sales)
7. 2006B #3 (Golf balls)
8. 2006B #4 (Dexterity)
9. 2006B #5 (Tractors)
10. 2010 #3 (Humane Society)
11. 2010B #1 (Polluted Rivers)
12. 2010B #2 (Cafeteria food)