Communication Nuances that Nudge Scores Higher

College Board Webinar, March 7, 2019

Dave Ferris, Presenter
dave_ferris@nobl.k12.in.us

Welcome!

Why focus on communication?

Why focus on *wrong* answers?

Make it Stick

Mindset

AP Central:

2018: Free-Response Questions			
Questions	**Scoring**	**Samples and Commentary**	**Score Distributions**
Free-Response Questions	Scoring Guidelines Chief Reader Report Scoring Statistics	Sample Responses Q1 Sample Responses Q2 Sample Responses Q3 Sample Responses Q4 Sample Responses Q5 Sample Responses Q6	Scoring Distributions

Finding good AP Problems:

- Subscribe to AP Community e-mail group; Read posts after the 2019 AP Exam
- Read Chief Reader Report
- Apply to be a Reader. You will be on the first to find problems and nuances!
- Review student samples when they are released, especially those that earn 3's, 2's.
- Read the "Scoring" portion of the AP Exam

Part (b) is scored as follows:

Essentially correct (E) if the probability is computed correctly, with work shown that includes appropriate numerical values for both the numerator and denominator.

Partially correct (P) if the response includes a numerator and denominator in calculating the conditional probability, with one appropriate term (numerator or denominator) and the other inappropriate.

Incorrect (I) if the response does not meet the criteria for E or P.

Note: Appropriate values include incorrectly calculated values from part (a).

2018 #2: Environmental Science Survey

2. An environmental science teacher at a high school with a large population of students wanted to estimate the proportion of students at the school who regularly recycle plastic bottles. The teacher selected a random sample of students at the school to survey. Each selected student went into the teacher's office, one at a time, and was asked to respond yes or no to the following question.

Do you regularly recycle plastic bottles?

Based on the responses, a 95 percent confidence interval for the proportion of all students at the school who would respond yes to the question was calculated as (0.584, 0.816).

(a) How many students were in the sample selected by the environmental science teacher?

(b) Given the method used by the environmental science teacher to collect the responses, explain how bias might have been introduced and describe how the bias might affect the point estimate of the proportion of all students at the school who would respond yes to the question.

This is a summary of an activity I did in class…just *one* example of how you might use student samples in class.

(Problem: 2018 #2b, Environmental Science Recycling Survey)

1. Give 5 minutes for students to write their answer to 2b.
2. "Stand up. Take turns reading your answer out loud to your group. DO NOT make any judgments, comments, suggestions or weird faces—just listen… Sit down when group is finished."
3. "Stand up if you heard an answer that was better than yours in some way." (most/all stood up)
4. "Now look at these two student samples." (pass out sheets with samples 2C1 and 2F1). "Read and score them with the E, P, I rating."
5. "Stand if you thought the first response was better. Stand if you thought the second response was better. Stand if you thought they were equally good/bad."
   (Most said #2 was better. And I agreed; it WAS better…BUT…they both were partially correct. BUT DON'T TELL THEM THIS YET.)
6. "Take 1-2 minutes in your group: why was the first response a 'partially correct' response?" (After group discussion, lead a whole group discussion: "The student did not say HOW it was biased. Unless students lie and say they do recycle when they really don't recycle, then the response is not biased. Only when they say something that is not true does bias enter into the survey.")
7. (Do the same for #2. This student gave TWO examples of why bias might creep into the survey, but they still did not say HOW bias could occur. They needed to say something like, "and they tell the teacher they recycle when they really do not recycle…")
8. Have students score their own (or their neighbor's) response, based on what they know about the correct, complete answer.

   There are three parts to the rubric for an Essentially Correct response:
   1. WHY the responses differ from the truth…
   2. HOW they differ…
   3. What direction/effect this will have on survey results…

2018 #2: Environmental Science Survey

Student Sample 2B from released student samples.

(b) Given the method used by the environmental science teacher to collect the responses, explain how bias might have been introduced and describe how the bias might affect the point estimate of the proportion of all students at the school who would respond yes to the question.

The students might feel pressure to answer yes since they were being surveyed by an environmental science teacher who would likely promote recycling. Thus, the point-estimate might overestimate the true proportion of all students who recycle plastic bottles.

Student Sample 2C1 from 2018 AP Reading:

(b) Given the method used by the environmental science teacher to collect the responses, explain how bias might have been introduced and describe how the bias might affect the point estimate of the proportion of all students at the school who would respond yes to the question.

There is bias in this sample because students do not want the teacher to know that they do not recycle. Therefore, the proportion of the sample of student who regularly recycle plastic bottles is higher than the true proportion.

Student Sample 2F1 from 2018 AP Reading:

(b) Given the ~~method used by the~~ environmental science teacher to collect the responses, explain how bias might have been introduced and describe how the bias might affect the point estimate of the proportion of all students at the school who would respond yes to the question.

There could be bias when the environmental Science teacher who's in favor of recycling askes the students if they recycle. The student may be embarassed to say no or pressured to say yes in fear of the teacher judging or disapproving of them. This is an example of response bias. This could cause an over prediction of the population proportion of students at that school who recycle.

From CRR: Common misconceptions/knowledge gaps

Failing to contrast how students would respond to the question (what they say) with the truth about whether they recycle (what they do).

NOTE: When I presented these samples to an experienced English teacher whom I respect, she concluded the last component (contrasting what they say with what they do) was implied in this student's answer, so she would have given these students a score of Essentially Correct.

<u>Note the nuances</u> in the Environmental Science problem (2018 #2):

Last student example:

recycling plastic bottles, so students might have been more likely to respond yes to the question, even if that was not their true answer. This bias might affect th

Score: _______

2017 #1: Gray Wolves

1. Researchers studying a pack of gray wolves in North America collected data on the length x, in meters, from nose to tip of tail, and the weight y, in kilograms, of the wolves. A scatterplot of weight versus length revealed a relationship between the two variables described as positive, linear, and strong.

   (a) For the situation described above, explain what is meant by each of the following words.

   (i) Positive:

   (ii) Linear:

   (iii) Strong:

Sample 1B from released student samples:

1. Researchers studying a pack of gray wolves in North America collected data on the length x, in meters, from nose to tip of tail, and the weight y, in kilograms, of the wolves. A scatterplot of weight versus length revealed a relationship between the two variables described as positive, linear, and strong.

(a) For the situation described above, explain what is meant by each of the following words.

(i) Positive:

Positive means there is a positive correlation between length, in meters, of the wolves and weight, in kilograms, of the wolves. As the length of the wolf increases it is expected that the weight will also increase.

(ii) Linear:

The data follows a linear pattern. As length increases, weight increases. The data points follow approximately the same pattern. Because it is linear we can also assume the residual plot follows no clear pattern.

(iii) Strong:

Strong implies that the data points for each wolf are closely related to each other. The points in the scatterplot follow fairly closely to the least squares regression line, also implying that the residuals for the points are small.

Sample: 1B
Score: 3

The statement about "positive correlation" in the first sentence of the response to part (a-i) essentially defines a positive relationship as a positive relationship. It does not explain the meaning of a linear relationship. The sentence is extraneous and does not affect the scoring. The second sentence provides a reasonable definition of a positive relationship, and it is presented in the context of the wolf study. The first sentence of the response to part (a-ii) merely introduces the concept that the response addresses. The second and third sentences describe a positive relationship. This was considered extraneous and not a parallel response. The fourth sentence describes a consequence of a straight line relationship but does not explain what is meant by a linear relationship. This response is not a reasonable definition of a linear relationship, but it does satisfy the context requirement for part (a). The first sentence of the response to part (a-iii) is too vague to provide a reasonable definition of the strength of a relationship. The first part of the second sentence is a reasonable definition of strength because it indicates that the points on the scatterplot are close to a straight line, in this case, the least-squares regression line. The rest of the second sentence enhances the explanation. Because reasonable definitions are given for two of the three components, and at least one is given in context, part (a) was scored as partially correct.

Sample 1C from released student samples

1. Researchers studying a pack of gray wolves in North America collected data on the length x, in meters, from nose to tip of tail, and the weight y, in kilograms, of the wolves. A scatterplot of weight versus length revealed a relationship between the two variables described as positive, linear, and strong.

(a) For the situation described above, explain what is meant by each of the following words.

(i) Positive:

When the data is plotted, there is an upward-right trend between the length and weight of the wolves

(ii) Linear:

when the data is plotted on the scatterplot, there is a strong relationship between the length and weight of the wolves with r close to 1 or -1 (in this case r is close to 1).

(iii) Strong:

When the data is plotted, the relationship between the length and weight of the wolves is strong which can be seen when drawing a line of best fit, the data should be near the regression line (no outliers).

<u>Note the nuances</u> from the Wolves problem (2018 #1):

Sample: 1C
Score: 2

The reference to "an upward-right trend" in the response to part (a-i) is not sufficiently precise to qualify as a reasonable definition of a positive relationship. The inclusion of the graph, however, provides the additional explanation needed for an acceptable response. This is a reasonable definition of a positive relationship, and it satisfies the context requirement for part (a). The response to part (a-ii) does not address the meaning of a linear relationship. Instead it uses a value of a correlation coefficient close to 1 or -1 to describe a strong relationship. Although it is not a reasonable definition of a linear relationship, the response does satisfy the context requirement for part (a). In the response to part (a-iii), indicating that the "data should be near the regression line" provides a reasonable definition of a strong relationship. The response also satisfies the context requirement for part (a). Because reasonable definition are given for two of the three components, and at least one is given in context, part (a) was scored as partially correct.

Notes:

- The description of a positive relationship should clearly indicate that relatively low values of one variable tend to appear with relatively low values of the other variable, and relatively high values of the first variable tend to appear with relatively high values of the other variable.
  - Examples of acceptable responses:
    - As length increases, so does weight.
    - Longer wolves weigh more.
    - The points on the graph go up as you move from left to right.
  - Examples of unacceptable responses:
    - As length goes up, weight changes.
    - Both length and weight get bigger.
    - The correlation is greater than 0.
- The description of a linear relationship can take one of two approaches: the data pattern (data points exhibit the pattern of a line in the graph) or the constant rate of change (as the explanatory variable changes, the response variable exhibits a constant rate of change).
  - Examples of acceptable responses:
    - The points generally follow a straight line.
    - The relationship between x and y is straight.
    - Length and weight have a constant slope.
  - Examples of unacceptable responses:
    - The points all line up.
    - You can draw a straight line through the points.
    - There is a positive correlation.
    - Every increase in x yields a 35.02 increase in y.
- The description of strong should indicate how close points are to a line.
  - Examples of acceptable responses:
    - Observed values are close to predicted values.
    - Deviations from the least-squares regression line are small.
    - The correlation coefficient is close to 1.
  - Examples of unacceptable responses:
    - All the points are close together.
    - The scatterplots are clustered together.
    - There is a high positive correlation.
- Context can be shown by referring to length and weight or by using meters and kilograms.
- Sketches and graphs can be used to help clarify definitions, but a sketch alone cannot satisfy a definition component.

2014 #2: Sales Representatives

2. Nine sales representatives, 6 men and 3 women, at a small company wanted to attend a national convention. There were only enough travel funds to send 3 people. The manager selected 3 people to attend and stated that the people were selected at random. The 3 people selected were women. There were concerns that no men were selected to attend the convention.

   (a) Calculate the probability that randomly selecting 3 people from a group of 6 men and 3 women will result in selecting 3 women.

   $$\frac{3}{9}\times\frac{2}{8}\times\frac{1}{7}=\frac{1}{84}=0.012$$

   (b) Based on your answer to part (a), is there reason to doubt the manager's claim that the 3 people were selected at random? Explain.

Student Sample "T" from the 2014 AP Reading

(b) Based on your answer to part (a), is there reason to doubt the manager's claim that the 3 people were selected at random? Explain.

Yes, there is reason to doubt the managers claim that the 3 people were selected at random because the probability for him to randomly select them is 1.2%. This means that it is very unlikely that he selected them at random.

Sample Identifier: T

Part (b)—There is no description of "small probability," or "unlikely outcome" attached to the appropriate probability, 1.2%. The only time "unlikely" is stated in the response, it isn't stating that selecting three women would be an unlikely *outcome*; it's stating that it is "very unlikely *that he selected them at random,"* which does not have the same meaning. (Essentially the last sentence stated that "it is very unlikely that the manager told the truth.") Part (b) is scored as Incorrect.

<u>Note the nuances</u> from Sales Representatives (2014 #2)

Part (b) is scored as follows:

Essentially correct (E) if the response states that the probability from part (a) is small (or insufficiently small), makes an appropriate decision consistent with the probability being small (or insufficiently small), and does so in the context of this situation.

Partially correct (P) if the response does any of the following:

- Otherwise satisfies the criteria for an (E) but does so without any context;
  OR
- States a significance level and makes a decision in context that is appropriate to the given probability in part (a) and the stated significance level, but does not explicitly compare the probability and the significance level;
  OR
- Otherwise satisfies the criteria for an (E) but does not explicitly make a decision about whether there is reason to doubt the manager's claim. (For example: "The probability of selecting the three women from among the nine employees is very small so it is unlikely to occur by chance.")

Incorrect (I) if the response does not meet the criteria for an (E) or a (P).

Notes:

- Each of the following situations is one in which a response that otherwise would score (E) should be scored (P), and a response that otherwise would score (P) should score (I):
  - The response includes a statement that the small probability *proves* that the manager did not make the selection at random (or any equivalent wording).
  - The response includes a statement that <u>clearly</u> interprets the probability from part (a) to be the probability that the manager selected the three people at random.

- Each of the following situations is one in which the response is scored (I):
  - The decision is inconsistent with the justification (*e.g.*, "the probability is very small, so there is no reason to doubt the manager's claim").
  - The response states or implies that because the selection of 3 women was *not impossible,* there is *no reason to doubt* the manager's claim.

2016 #3: Alzheimer's and Smoking

3. Alzheimer's disease results in a loss of cognitive ability beyond what is expected with typical aging. A local newspaper published an article with the following headline.

Study Finds Strong Association Between Smoking and Alzheimer's

The article reported that a study tracked the medical histories of 21,123 men and women for 23 years. The article stated that, for those who smoked at least two packs of cigarettes a day, the risk of developing Alzheimer's disease was 2.57 times the risk for those who did not smoke.

(a) Identify the explanatory and response variables in the study.

Explanatory variable:

Response variable:

Part (a) is scored as follows:

Essentially correct (E) if both variables are described correctly
Note: Describing the variables as "smoking" and "Alzheimer's" is not sufficient.

Partially correct (P)
if one variable is described correctly and one is not described correctly;
OR
if neither variable is described correctly but smoking is mentioned as the explanatory variable and Alzheimer's is mentioned as the response variable;
OR
if the explanatory and response variables are correctly described but are interchanged;

Incorrect (I) if the response does not meet the criteria for E or P.

Student sample 3F from the 2016 AP Reading

3. Alzheimer's disease results in a loss of cognitive ability beyond what is expected with typical aging. A local newspaper published an article with the following headline.

Study Finds Strong Association Between Smoking and Alzheimer's

The article reported that a study tracked the medical histories of 21,123 men and women for 23 years. The article stated that, for those who smoked at least two packs of cigarettes a day, the risk of developing Alzheimer's disease was 2.57 times the risk for those who did not smoke.

(a) Identify the explanatory and response variables in the study.

Explanatory variable:
# of packs of cigarettes smoked per day

Response variable:
the risk of developing alzheimers

Sample Identifier: 3F
Score: 2

- Part (a) is scored as partially correct (P). The explanatory variable is correctly defined as a variable that represents a degree a smoking by defining it as the number of cigarettes smoked per day. The response variable is incorrect, as it refers to the "risk" of developing Alzheimer's disease. Risk is a measurement that is taken after a study has been conducted and is not a characteristic of an observational unit.

"not a characteristic of an observational unit" Interesting that one of the most basic concepts we talk about, variables, can be difficult to explain completely and correctly.

<u>Note the nuances</u> from Alzheimer's and Smoking (2016 #3)

2012 #4:

4. A survey organization conducted telephone interviews in December 2008 in which 1,009 randomly selected adults in the United States responded to the following question.

At the present time, do you think television commercials are an effective way to promote a new product?

Of the 1,009 adults surveyed, 676 responded "yes." In December 2007, 622 of 1,020 randomly selected adults in the United States had responded "yes" to the same question. Do the data provide convincing evidence that the proportion of adults in the United States who would respond "yes" to the question changed from December 2007 to December 2008 ?

Student Sample "N" from the 2012 AP Reading

Define: P_1 = proportion of adults who said 'yes' in 2007.

P_2 = proportion of adults who said 'yes' in 2008.

Test:

1. $H_0: P_1 = P_2$ The true proportion of adults who said 'yes' to the stated question did not change between 2007 & 2008.

$H_a: P_1 \neq P_2$ The true proportion of adults who said 'yes' to the stated question in 2008 was different from 2007.

Note the nuance in Adult Surveys (2012 #4):

Sample 4B from released student samples:

2008	2007
$n_1 = 1009$	$n_2 = 1020$
$x_1 = 676$	$x_2 = 622$
$p_1 = 0.670$	$p_2 = 0.61$

$H_0: p_1 = p_2$

$H_A: p_1 \neq p_2$

Sample 4
Score: 3

In step 1, although the form of the hypotheses is correct, p1 and p2 are defined as the sample proportions with the equations p1 = 0.670 and p2 = 0.61. Step 1 was scored as partially correct.

From the Chief Reader's Report:

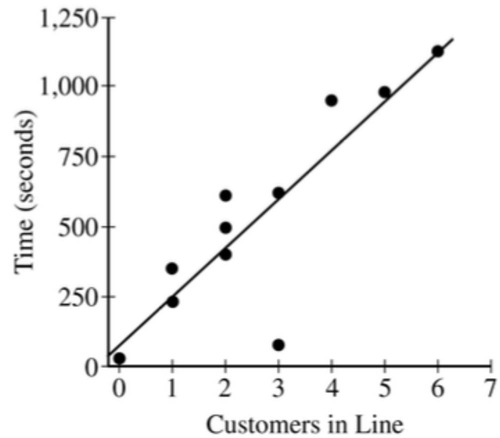
What were common student errors or omissions?

Some students included written descriptions of hypotheses that pertained to samples rather than populations, even though few students used symbols for sample statistics in their hypothesis statements. One common way to do this was to mistakenly say that one of the parameters is the proportion of adults who answered "yes" in 2008; **the use of past *tense*** ("answered" rather than "would have answered") made the description about the sample rather than the population.

Practice Problems:

2018 #1c: Grocery Store Wait Times

1. The manager of a grocery store selected a random sample of 11 customers to investigate the relationship between the number of customers in a checkout line and the time to finish checkout. As soon as the selected customer entered the end of a checkout line, data were collected on the number of customers in line who were in front of the selected customer and the time, in seconds, until the selected customer was finished with the checkout. The data are shown in the following scatterplot along with the corresponding least-squares regression line and computer output.



Predictor	Coef	SE Coef	T	P
Constant	72.95	110.36	0.66	0.525
Customers in line	174.40	35.06	4.97	0.001
S = 200.01		R-Sq = 73.33%	R-Sq (adj) = 70.37%	

c) One of the data points was determined to be an outlier. Circle the point on the scatterplot and explain why the point is considered an outlier.

Student Sample 1C from the 2018 Reading:

(c) One of the data points was determined to be an outlier. Circle the point on the scatterplot and explain why the point is considered an outlier.

The point is an outlier because its y value is so far off from its $\hat{y}$ value.

Student Sample 1H from the 2018 Reading:

(c) One of the data points was determined to be an outlier. Circle the point on the scatterplot and explain why the point is considered an outlier.

Data point (3, 62.5) is considered an outlier, because it is the farthest point from the line of best fit, it is a negative residual. On average, according to the line of best fit, when there are 3 customers in line it takes about 625 seconds to complete the checkout, but the data point above took 62.5 seconds, there can be confounding variables such as the amount of items each customer had.

Student Sample 1G from the 2018 Reading:

(c) One of the data points was determined to be an outlier. Circle the point on the scatterplot and explain why the point is considered an outlier.

This is considered an outlier because it deviates from the linear pattern. The regression line doesn't pass through or run near this data point so it is an outlier.

1C: "so far off" ≠ most far off; no comparison to other points; P

1H: "farthest point from the line" = E ("farthest" implies other points)

1G: "deviates from the linear pattern" implies comparison to other points; E

Student Sample 1E from the 2018 Reading:

(c) One of the data points was determined to be an outlier. Circle the point on the scatterplot and explain why the point is considered an outlier.

That point is an outlrer because it is the point that has the highest residual from the least-squares regression line.

Student Sample 1D from the 2018 Reading:

(c) One of the data points was determined to be an outlier. Circle the point on the scatterplot and explain why the point is considered an outlier.

This point is considered to be an outlier because it has a very large, negative residual. It's also an influential point meaning it would drastically change the slope of the regression line.

Student Sample 1B from the 2018 Reading:

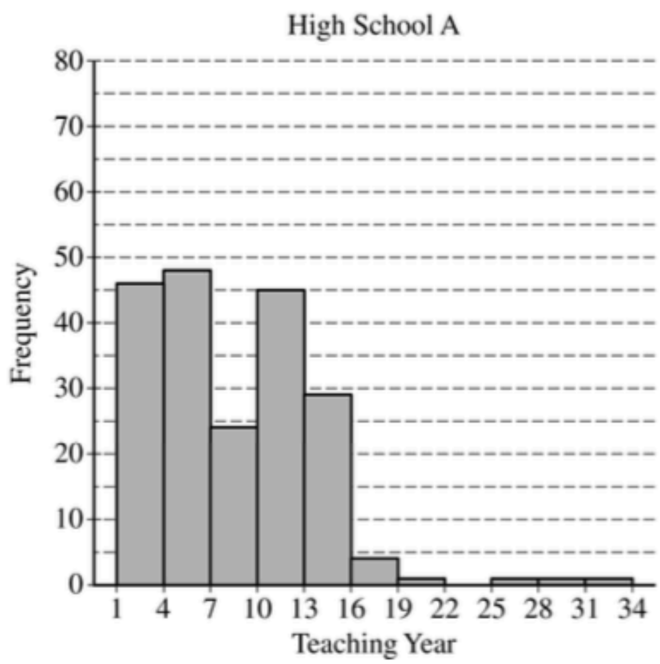
(c) One of the data points was determined to be an outlier. Circle the point on the scatterplot and explain why the point is considered an outlier.

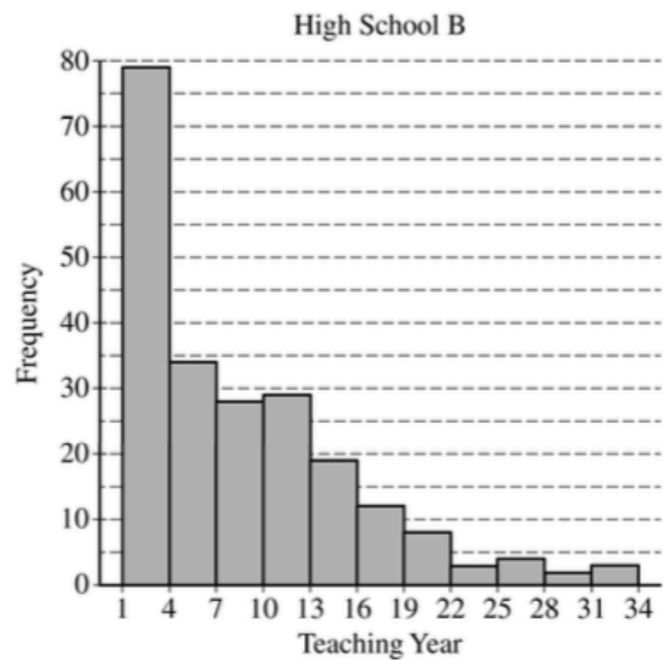
This point is much farther away from the least-squares regression line than the others points. It has a very negative residual. With the amount of customers in line ahead of the selected customer, it takes a relatively short amount of time to finish checking out compared to the other instances.

1E: "highest residual" = E
1D: no comparison to the other dots = P
1B: "much further away…than the other points." = E

2018 #5a: Teaching Years

5. The following histograms summarize the teaching year for the teachers at two high schools, A and B.





Teaching year is recorded as an integer, with first-year teachers recorded as 1, second-year teachers recorded as 2, and so on. Both sets of data have a mean teaching year of 8.2, with data recorded from 200 teachers at High School A and 221 teachers at High School B. On the histograms, each interval represents possible integer values from the left endpoint up to but not including the right endpoint.

(a) The median teaching year for one high school is 6, and the median teaching year for the other high school is 7. Identify which high school has each median and justify your answer.

2018 #5a

Student sample 5G from the 2018 Reading:

(a) The median teaching year for one high school is 6, and the median teaching year for the other high school is 7. Identify which high school has each median and justify your answer.

High School A = median of 7

High School B = median of 6

High School B is skewed right and has a much higher frequency of teacher's who have a smaller teaching year. This would make the median of high school B be less than that of high school A. High School B's Shape is unimodel and skewed right while A is not.

Student sample 5H from the 2018 Reading:

(a) The median teaching year for one high school is 6, and the median teaching year for the other high school is 7. Identify which high school has each median and justify your answer.

The median teacing year for High school A is 7 years and The median teaching year for high School B is 6 years. When calculated, The median year for A falls between 7 and 10 and The medium year for B falls between 4 and 7

5G: Skewness argument given, but no statement about separation of mean and median: P

5H: Uses location argument, but did not say HOW location was determined: P

Dave's Top Ten Teaching Strategies

1. Use FRAPPY's (see instructions on extended handout; also in TPS textbook).
   See: apstatsmonkey.com/StatsMonkey/FRAPPYs.html

2. Use FRAPPY Jr's: shorter versions of the longer FRAPPYs (what we did). Perhaps focus on only ONE part (see example in extended handout). Try different ways to integrate student samples and rubrics into your classroom.

3. Give AP Exam problems while material is fresh so students have confidence.

4. Spiral important concepts later. Example: Halloween candy and comparing distributions of fun size bag counts (MM's vs. Skittles, etc.)

5. Don't call on students that have their hands up. Randomly select students during discourse. Before choosing, have students discuss with a neighbor or group. This provides comfort and confidence. Emphasize you are really calling on their group/pair, since they discussed the answer ahead of time.

6. Use your ***<u>own</u>*** students' samples—you will have plenty to choose from! Then archive these for future years. See 7.2 Warmup and Confidence Interval Quiz on the following pages.

7. Ask students to *predict* rubrics for a problem, especially later in the year. Use the E, P, I scheme early and often.

8. Engage your students in productive discourse when discussing their written answers and common nuances between good and great answers. There are some good books that address how to do this well, like:
   5 Practices for Orchestrating Productive Mathematics Discussions [NCTM] by Margaret Smith and Mary Kay Stein

9. Use Academic Vocabulary, Close Reading (purposeful annotation) strategies
   altonschools.org/media/pdf/Marzano_Vocab.pdf
   (Google "close reading," and you will find lots of resources)

10. Habits, Perspective, and Atmosphere
    - Embed AP Exam questions and parts into quizzes and tests.
    - Talk @ rubrics early and often
    - Most scores of 2 out of 4 are above average! Means are usually between 1-2

      Engagement (Robert Marzano):

      Students must answer in the affirmative:

      How do I feel? Is it interesting? (This gets their attention…)

      Is it important? Can I do it? (This begins authentic engagement…)

***See noblestatman.com under "1-Day Workshops" for full (extended) handout...**

AP Statistics Warm-up over 7.2

Name________________________________

Just before a presidential election, a national opinion poll increases the size of its weekly random sample from the usual 1500 people to 4000 people.

(a) Does the larger random sample reduce the bias of the poll result? Explain.

(b) Does it reduce the variability of the result? Explain.

AP Statistics HW Student Samples

Bias vs. Variation NAME________________________________

Mark each student's answer as <u>E</u>ssentially correct, <u>P</u>artially correct or <u>I</u>ncorrect. (Yes, these are authentic student answers!)

18. Just before a presidential election, a national opinion poll increases the size of its weekly random sample from the usual 1500 people to 4000 people.

(a) Does the larger random sample reduce the bias of the poll result? Explain.

Yes, more samples = less variance.*

Yes more samples will reduce bias.*

No, it would just decrease the variation.

No it would only reduce the variation in the sample distribution.

It doesn't reduce bias it would only reduce variation.

No, a larger sample would not reduce bias.

No, a larger random sample doesn't reduce bias because bias means your systematically off so no matter how many you have it will still be off.

No it doesn't reduce the bias because a larger sample doesn't change bias so no matter how many there is it will still be off.

No—since bias is systematic, adding more individuals to a sample will not reduce a systematic flaw in the survey.

No, because bias is systematic, so adding more people wouldn't change the results.

No, because bias is a systematic favoring of an outcome. Adding more people to a systematic error doesn't lessen it.

The bias will not be reduced because a larger amount of samples won't necessarily cause a systematic change.

(b) Does it reduce the variability of the result? Explain.

Yes, more people you survey more correct your answer is.*

Yes, the more people, the more accurate the result.*

Yes, because there's more possible outcomes.

Yes, because the larger sample increases more possible outcomes.

Yes, because the larger pool covers more possible outcomes.

Yes, a larger sample would reduce variability of the result because there are more people and a more accurate result.

Yes, a larger sample size reduces variability.

Yes, because larger samples give a smaller spread, which means a reduced variability.

Yes—larger samples decrease a sample's spread and, therefore, its variability.

The variability will increase with larger sample.

A larger sample, however, will reduce the variability because less variability will be in between each sample.

*These answers were from the same group. Students were instructed to collaborate and ensure group members answers were at least equivalently stated.

ANSWERS: TO HW DEBRIEF 7.1

18. Just before a presidential election, a national opinion poll increases the size of its weekly random sample from the usual 1500 people to 4000 people.

(a) Does the larger random sample reduce the bias of the poll result? Explain.

Yes, more samples = less variance.* **DIDN'T ANSWER THE QUESTION ABOUT <u>*BIAS*</u>**

Yes more samples will reduce bias.* **DIDN'T EXPLAIN**

No, it would just decrease the variation. **OF WHAT?**

No it would only reduce the variation in the sample distribution. **SHOULD BE "<u>*SAMPLING*</u>"**

It doesn't reduce bias it would only reduce variation. **NO EXPLANATION**

No, a larger sample would not reduce bias. **NO EXPLANATION**

No, a larger random sample doesn't reduce bias because bias means your systematically off so no matter how many you have it will still be off. **PRETTY GOOD! (EXCEPT FOR YOUR)**

No it doesn't reduce the bias because a larger sample doesn't change bias so no matter how many there is it will still be off. **REDUNDANT; TOO MANY USES OF "IT"…UNCLEAR**

No—since bias is systematic, adding more individuals to a sample will not reduce a systematic flaw in the survey. **PRETTY GOOD!**

No, because bias is systematic, so adding more people wouldn't change the results. **TOO VAGUE (SYSTEMATIC WHAT?? … RESULTS OF WHAT?)**

No, because bias is a systematic favoring of an outcome. Adding more people to a systematic error doesn't lessen it. **EXCELLENT!**

The bias will not be reduced because a larger amount of samples won't necessarily cause a systematic change. **UNDERSTANDING OF BIAS IS IMPLIED BY LAST 4 WORDS**

(b) Does it reduce the variability of the result? Explain.

Yes, more people you survey more correct your answer is.* **"MORE CORRECT" IS VAGUE; PERHAPS "CLOSER TO THE TRUE VALUE" WOULD BE BETTER. AN ESTIMATE IS EITHER CORRECT OR INCORRECT. IT CANNOT BE "MORE CORRECT."**

Yes, the more people, the more accurate the result.* **"RESULT?" TOO VAGUE--NEED TO REFERENCE THE *SAMPLING DISTRIBUTION* FOR FULL CREDIT**

Yes, because there's more possible outcomes. **VERY UNCLEAR…**

Yes, because the larger sample increases more possible outcomes. **EQUALLY UNCLEAR…**

Yes, because the larger pool covers more possible outcomes. **WHAT'S UP WITH THIS PHRASE?**

Yes, a larger sample would reduce variability of the result because there are more people and a more accurate result. **STILL NO EXPLANATION TIED TO THE *SAMPLING DISTRIBUTION***

Yes, a larger sample size reduces variability. **VARIABILITY IN WHAT EXACTLY? (ANS: THE *SAMPLING DISTRIBUTION*)**

Yes, because larger samples give a smaller spread, which means a reduced variability. **DITTO…NEEDED TO ADD "IN THE *SAMLPING DISTRIBUTION*" AT THE END**

Yes—larger samples decrease a sample's spread and, therefore, its variability. **NOT THE "SAMPLE'S" SPREAD…THE SPREAD OF THE *SAMPLING DISTRIBUTION***

The variability will increase with larger sample. **COMPLETELY BACKWARDS…**

A larger sample, however, will reduce the variability because less variability will be in between each sample. **AHA! FINALLY A CORRECT EXPLANATION, ALTHOUGH NOT WHAT I WAS EXPECTING (I WAS EXPECTING A REFERENCE TO THE *SAMPLING DISTRIBUTION*)**

Confidence Interval Wording Quiz

Name________________________________

Assume a 95% confidence interval has been calculated for the proportion of orange Reece's Pieces from a bag of 121 candies. The interval is (.32, .49). Forty-nine pieces were orange.

Answer each sample with C(orrect) or I(ncorrect).

_______1. In the long run, 95% of all sample proportions will yield an interval that contains the true population parameter.

_______2. 95% of samples this size will produce confidence intervals that capture the true proportion of orange Reece's Pieces.

_______3. 95% of the time, the true proportion will lie between .32 and .49.

_______4. I have 95% confidence in my methods.

_______5. We are 95% sure that bags of Reece's Pieces will contain between 32% and 49% orange pieces.

_______6. The interval (.32, .49) will be correct 95% of the time and wrong 5% of the time.

_______7. For a given sample size, lower confidence means a larger margin of error.

_______8. "if this method were used to generate an interval estimate over and over again with different samples, in the long run, 95% of the resulting intervals would capture the true value of the characteristic being estimated."

_______9. For a given confidence level, a sample twice as large will make a margin of error twice as big.

_______10. According to my sample bag, 95% of all Reece's Pieces bags of this size will have between 32% and 49% orange candies.

_______11. For a certain confidence level, you can get a smaller margin of error by selecting a bigger sample.

_______12. There is a 95% chance that the sample proportion lies in this interval.

_______13. There is a 95% chance that the true proportion lies between .32 and .49.

_______14. 95% of the population will be contained in this interval.

_______15. For a fixed margin of error, larger samples provide greater confidence.

_______16. 95% of the sample proportions lie in this interval.

_______17. If we computed repeated random sample proportions, about 95% of them would lie in this interval.

_______18. According to my sample bag, between 32% and 49% of Reeces' Pieces are orange.

_______19. When I behave in the way in which I am behaving (using statistically correct methods) I expect to be right about 95% of the time and wrong about 5% of the time.

_______20. It is probably true that 40.5% of Reece's Pieces are orange.

KEY to Confidence Interval Wording Quiz

1. CORRECT: This is a good one! "In the long run, 95% of all sample means will yield an interval that contains the true population parameter."

2. CORRECT: Same reasoning as #1, #4 and #19 (see BVD, p. 374)

3. WRONG: There is nothing magical about the numbers .32 and .49. If many additional CI's were generated from the same sample size, the true proportion will fall within 95% of these intervals.

4. CORRECT: The confidence comes from doing correct methods. We will be right 95% of the time we do these methods correctly.

5. WRONG: Again, the focus of a CI should not be on future sample proportions,

6. WRONG: The interval will be correct 95% of the time and wrong 5% of the time. (it isn't the interval that is wrong, it is *my* error for choosing to make a decision in this way (Paul Velleman).

7. WRONG: A larger margin of error would give you more confidence. Think about hitting a target. If the target is bigger (larger MOE), then you will have more confidence that you will hit it. Also, check out the formula. Your algebra background should be able to prove this.

8. CORRECT: "if this method were used to generate an interval estimate over and over again with different samples, in the long run, 95% of the resulting intervals would capture the true value of the characteristic being estimated." (POD p. 443)

9. WRONG: You will need a sample four times as big—see the formula.

10. WRONG: The interval isn't about sample proportions, but about the population proportion. (BVD, p. 373)

11. CORRECT: Do the algebra on the CI formula.

12. WRONG: I am 100% sure that my sample proportion lies in this interval!!

13. WRONG: The true population proportion is either IN there or it is NOT in there (100% chance or 0% chance). We just don’t know which.

14. WRONG: 95% of the population will be contained in this interval.

15. CORRECT: Think back to the target again. If the target stays the same size (same MOE), but you have more chances to throw at the target, then your confidence that you will hit is once increases. You can also do the algebra with the CI formula.

16. WRONG: Additional SAMPLE proportions do not have to land within this particular CI. We’re pretty sure the TRUE proportion lies in this CI. but rather on the true proportion. (BVD, p. 376, #6c.)

17. WRONG: Focus on the true proportion. Our confidence is not about future sample proportions, but on where the true proportion lies.

18. WRONG: This statement asserts that the population proportion cannot be outside that interval. We cannot be absolutely sure about that (just pretty sure). (BVD, p. 373)

19. CORRECT: When I behave in the way in which I am behaving (using statistically correct methods) I expect to be right about 95% of the time and wrong about 5% of the time.

20. WRONG: This was only OUR sample proportion. Another sample proportion would likely produce a different p-hat. There is nothing magical about our ONE p-hat.

F.R.AP.P.Y's

(Free Response AP Problems—Yay!)
(Problems can be found on StatsMonkey web site)

"FRAPPYs are not simply a test-preparation tool or a you-do-the-problem-and-I'll-grade-it-and-give-it-back-to-you exercise. The FRAPPY is an assessment FOR learning whose purpose is to provide students feedback and a means for self-reflection on their conceptual understanding as well as help them develop their communication skills. The students are the critical component...they not only do the problem, but they also become an AP Reader and evaluate their performance as well as that of others."

--Jason Molesky

1) Hand out FRAPPY! and give **12-15 minutes** to complete.

2) Then students turn their response over and briefly **discuss the "Intent** of the Question" from their perspective.

   Ask, "What do you think this question was getting at? What statistical concept or ability are they asking you to display?"

3) **Show 2-3 student responses and** have pairs of kids **classify them** as Minimal, Developing, Substantial, or Complete.

   Discuss why they classified them that way. What did/didn't the sample responses do? Note, they have NOT seen the rubric at this point. In a sense, they are developing it on their own.

4) Hand out and discuss the **actual scoring rubric**.

5) Have students pair up and **grade each other's responses**.

6) **Have students reflect** on what they would do differently to improve their response on similar questions.

   File away for AP review later...

Example of FRAPPY Jr.:

1. Find a good AP Exam problem for students to complete, but focus on ONE PART only. They may need to complete an earlier part in order to a later part, but have them focus on one part only. Give them an appropriate amount of time, perhaps 5-7 minutes.
2. In groups of 3-4, have students put all their papers face down in the middle of their group. Have one student shuffle them and pass them out to the group. They may get theirs back or they may not.
3. Tell students to [stand up and] take turns reading the paper they received out loud to their group. Have the group informally score each one with E, P or I. Write the group consensus score on the paper. Repeat until all papers are read and scored.
4. Have students look at several student samples of this problem part and score them as E, P, or I.
5. Lead a class discussion over the student samples, and point out the rubric used to score them. Try to simplify the rubrics so students do not get "lost in the weeds."
6. Using these newly revealed rubrics, have them score someone else's paper from their original work on the problem. Give them a chance to hand back and discuss the scoring decisions.

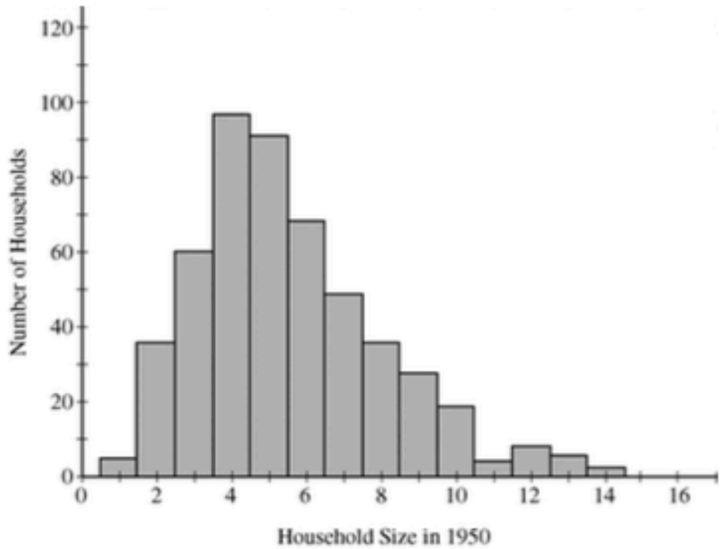
Notes about calling on students:

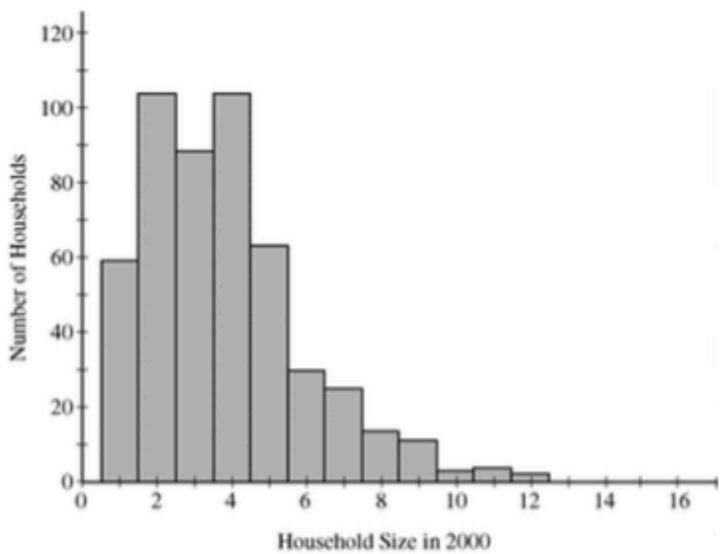
Do not solely call on students with their hands up. Randomly select students using a calculator or other random method! Before choosing them, include an appropriate opportunity to discuss with a neighbor or their group. This should provide appropriate comfort and confidence. Also, you can emphasize that although you are calling on a single student, you are really calling on their group/pair, since they discussed the answer ahead of time. The selected student is just the spokesperson for that particular group/pair.

12 Common Communication Errors

…on the AP Stats Exam…see if you can find them… NAME_______________________________

2012 Stud.A 3. Independent random samples of 500 households were taken from a large metropolitan area in the United States for the years 1950 and 2000. Histograms of household size (number of people in a household) for the years are shown below.





(a) Compare the distributions of household size in the metropolitan area for the years 1950 and 2000.

shape - Both distributions are skewed right; distribution of Year 2000 has a stronger skew

center - The center has moved slightly (left) lower from 1950's to 2000's distribution.

spread - the spread of 1950's distribution is slightly greater than that of 2000's.

2012 #3: Student B

The distribution of household size in 1950 is slightly skewed to the right, but does not show any extreme outliers. It also includes households with sizes between 1 and 14.

The distribution of household size in 2000 is extremely skewed to the right and households with 10-12 people may be considered outliers. This distribution includes households with sizes between 1 and 12, which is different than the 1950's distribution.

2014

3. Schools in a certain state receive funding based on the number of students who attend the school. To determine the number of students who attend a school, one school day is selected at random and the number of students in attendance that day is counted and used for funding purposes. The daily number of absences at High School A in the state is approximately normally distributed with mean of 120 students and standard deviation of 10.5 students.

   (a) If more than 140 students are absent on the day the attendance count is taken for funding purposes, the school will lose some of its state funding in the subsequent year. Approximately what is the probability that High School A will lose some state funding?

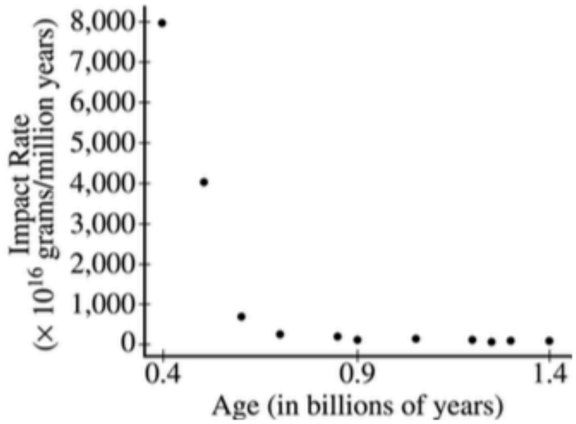
normalcdf (141, 10000, 120, 10.5)

= .0228

The probability High School A will lose some state funding is 2.28%.

2004B (focus on part (b))

1. The Earth's Moon has many impact craters that were created when the inner solar system was subjected to heavy bombardment of small celestial bodies. Scientists studied 11 impact craters on the Moon to determine whether there was any relationship between the age of the craters (based on radioactive dating of lunar rocks) and the impact rate (as deduced from the density of the craters). The data are displayed in the scatterplot below.



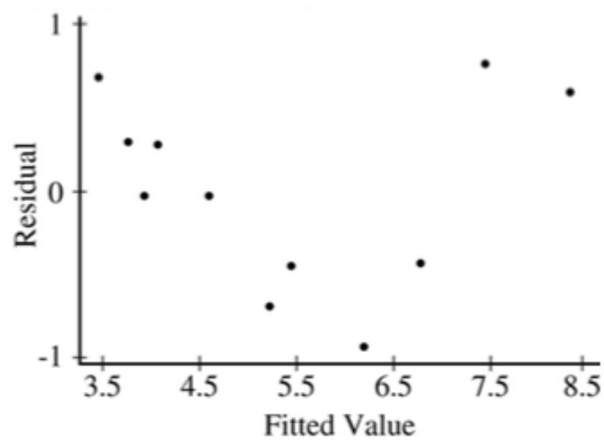
(a) Describe the nature of the relationship between impact rate and age.

Prior to fitting a linear regression model, the researchers transformed both impact rate and age by using logarithms. The following computer output and residual plot were produced.

Regression Equation: ln(rate) = 4.82 – 3.92 ln(age)

Predictor	Coef	SE Coef	T	P
Constant	4.8247	0.1931	24.98	0.000
ln(age)	–3.9232	0.4514	–8.69	0.000

S = 0.5977 R-Sq = 89.4% R-Sq (adj) = 88.2%



(b) Interpret the value of r^2.

89.4 % of the variability in impact rate is accounted for by the variability in age.

2007

2. As dogs age, diminished joint and hip health may lead to joint pain and thus reduce a dog's activity level. Such a reduction in activity can lead to other health concerns such as weight gain and lethargy due to lack of exercise. A study is to be conducted to see which of two dietary supplements, glucosamine or chondroitin, is more effective in promoting joint and hip health and reducing the onset of canine osteoarthritis. Researchers will randomly select a total of 300 dogs from ten different large veterinary practices around the country. All of the dogs are more than 6 years old, and their owners have given consent to participate in the study. Changes in joint and hip health will be evaluated after 6 months of treatment.

   (a) What would be an advantage to adding a control group in the design of this study?

An advantage to adding a control group to this design would be that it gives the experiment something to compare its results to, to see how much of a difference the treatments make.

2005

5. A survey will be conducted to examine the educational level of adult heads of households in the United States. Each respondent in the survey will be placed into one of the following two categories:

   - Does not have a high school diploma
   - Has a high school diploma

   The survey will be conducted using a telephone interview. Random-digit dialing will be used to select the sample.

   (a) For this survey, state one potential source of bias <u>and</u> describe how it might affect the estimate of the proportion of adult heads of households in the United States who do not have a high school diploma.

One possible source of bias would be that a greater amount of households that are head by a person who does not have a high school diploma would not have phones. This would be undercoverage bias because they would not reach these people. A greater prop. of heads of houses will be in this sample than the true pop. prop

2007 (focus on part (c))

2. As dogs age, diminished joint and hip health may lead to joint pain and thus reduce a dog's activity level. Such a reduction in activity can lead to other health concerns such as weight gain and lethargy due to lack of exercise. A study is to be conducted to see which of two dietary supplements, glucosamine or chondroitin, is more effective in promoting joint and hip health and reducing the onset of canine osteoarthritis. Researchers will randomly select a total of 300 dogs from ten different large veterinary practices around the country. All of the dogs are more than 6 years old, and their owners have given consent to participate in the study. Changes in joint and hip health will be evaluated after 6 months of treatment.

   (a) What would be an advantage to adding a control group in the design of this study?

   (b) Assuming a control group is added to the other two groups in the study, explain how you would assign the 300 dogs to these three groups for a completely randomized design.

   (c) Rather than using a completely randomized design, one group of researchers proposes blocking on clinics, and another group of researchers proposes blocking on breed of dog. How would you decide which <u>one</u> of these two variables to use as a blocking variable?

Rather than using a completely randomized design, I would encorporate blocking on the specific breed of dog, because the different treatments could possibly have a different effect on the different types of dogs, therefore I would use the blocking on breed of dog so it would eliminate any variables that could change the actual data. I would rather block on breed of dog than clinic, because breeds of dogs seem to be much more different than different clinics, therefore I would block on breeds of dogs.

2012

4. A survey organization conducted telephone interviews in December 2008 in which 1,009 randomly selected adults in the United States responded to the following question.

At the present time, do you think television commercials are an effective way to promote a new product?

Of the 1,009 adults surveyed, 676 responded "yes." In December 2007, 622 of 1,020 randomly selected adults in the United States had responded "yes" to the same question. Do the data provide convincing evidence that the proportion of adults in the United States who would respond "yes" to the question changed from December 2007 to December 2008 ?

<u>Define</u>: P_1 = proportion of adults who said 'yes' in 2007.

P_2 = proportion of adults who said 'yes' in 2008.

<u>Test</u>:

1. $H_0: P_1 = P_2$ The true proportion of adults who said 'yes' to the stated question did not change between 2007 & 2008.

$H_a: P_1 \neq P_2$ The true proportion of adults who said 'yes' to the stated question in 2008 was different from 2007.

2013 (focus on the interpretation of the interval)

1. An environmental group conducted a study to determine whether crows in a certain region were ingesting food containing unhealthy levels of lead. A biologist classified lead levels greater than 6.0 parts per million (ppm) as unhealthy. The lead levels of a random sample of 23 crows in the region were measured and recorded. The data are shown in the stemplot below.

Lead Levels

2	8
3	0
3	5 8 8
4	1 1 2
4	6 8 8
5	0 1 2 2 3 4
5	9 9
6	3 4
6	6 8

Key: 2|8 = 2.8 ppm

(a) What proportion of crows in the sample had lead levels that are classified by the biologist as unhealthy?

(b) The mean lead level of the 23 crows in the sample was 4.90 ppm and the standard deviation was 1.12 ppm. Construct and interpret a 95 percent confidence interval for the mean lead level of crows in the region.

conditions: Simple random sample. stated in problem

23(10) = 2300, safe to assume > 2300 crows in world

histgram of data (as shown by stemplot) is

approx normal w/ no outliers.

1 sample t-interval at 95% confidence:

(4.4157, 5.3843)

in repeated samples

we are 95% confident that, the mean lead level of crows in the region falls between the values of 4.4157 and 5.3843 ppm

2014 #1

(c) After verifying that the conditions for inference were satisfied, the administrator performed a chi-square test of the following hypotheses.

H_0 : There is no association between residential status and level of participation in extracurricular activities among the students at the university.

H_a : There is an association between residential status and level of participation in extracurricular activities among the students at the university.

The test resulted in a p-value of 0.23. Based on the p-value, what conclusion should the administrator make?

Based on the p-value of 0.23, the administrator shoud make the conclusion that there is a correlation between residential status and the level of participation in extracurricular activities among the students at the university.

2014 (focus on part (b))

2. Nine sales representatives, 6 men and 3 women, at a small company wanted to attend a national convention. There were only enough travel funds to send 3 people. The manager selected 3 people to attend and stated that the people were selected at random. The 3 people selected were women. There were concerns that no men were selected to attend the convention.

(a) Calculate the probability that randomly selecting 3 people from a group of 6 men and 3 women will result in selecting 3 women.

$P(3 \text{ women selected}) : \left(\frac{3}{9}\right)\left(\frac{2}{8}\right)\left(\frac{1}{7}\right) \approx 0.0119$, or $\approx 1.2\%$ chance

(b) Based on your answer to part (a), is there reason to doubt the manager's claim that the 3 people were selected at random? Explain.

Yes, because there is only a 1.2% chance that the 3 women were chosen at random, and we are generally suspicious of anything under a 5% chance of occuring randomly.

2012 #4

2 Proportion Z-Test is appropriate

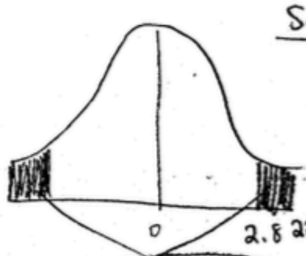
Randomization: 1,009 randomly selected adults in the US in 2008
1,020 randomly selected adults in the US in 2007

2008	2007
$n_1 = 1009$	$n_2 = 1020$
$x_1 = 676$	$x_2 = 622$
$P_1 = 0.670$	$P_2 = 0.61$

10% condition: 1,009 randomly selected adults in the United States in 2008 and 1,020 randomly selected adults in the United States in 2007 are likely less than 10% of their respective populations.

$H_0: P_1 = P_2$

$H_A: P_1 \neq P_2$



Success/Failure: $(1009)(0.670) \geq 10$
$(1009)(0.33) \geq 10$
$(1020)(0.61) \geq 10$
$(1020)(0.39) \geq 10$

Independent Groups: reponse of 1009 randomly selected adults in 2008 is likely independent from 1020 randomly selected adults in 2007.

$$z = 2.8226 = \frac{0.34 - 0}{\sqrt{\frac{(0.670)(0.33)}{1009} + \frac{(0.61)(0.39)}{1020}}}$$

$P = 0.00476$

Since a p-value of 0.00476 is less than any reasonable significance level, we reject the null hypothesis. There is evidence to suggest that the proportion of adults in the United States who would respond "yes" to the question changed from December 2007 to December 2008

ANSWERS: The point of emphasis in each example is <u>underlined</u> below.
Each graded part on the AP Statistics Exam is scored as: **E**(ssentially) correct, **P**(artially) correct or **I**(ncorrect)

2012 #3: Stud A: This student COMPARED shape, center and spread, but there is <u>no mention of context</u> (household sizes). Score = P.

2012 #3: Stud B: Part (a) is worth three points.
1) Correctly compares centers. Score = I (no direct comparison)
2) Correctly compares variability. Score = I (no direct comparison)
3) Correctly compares shapes AND includes context (somewhere)
Shapes ARE compared, and the student mentioned household sizes. Score = E
Takeaway: <u>ALWAYS write statements of direct comparison</u> when comparing distributions.

2014 #3: There is danger in only showing "<u>calculator speak</u>." To earn an "E" for this problem, students need to 1) identify a normal model AND the two parameters, 2) use a correct boundary, and 3) calculate the correct normal probability. Since the parameters are not clearly identified, this student earned a P.

2004B #1b: <u>Context is incorrect</u>. The variables are the LN(age) and LN(impact rate). Score = P.
(This brought the student's overall problem score down from a 4 to a 3.)

2007 #2: <u>No context</u>. Score = P

2005 #5: No explanation of how this bias "might affect the estimate of the proportion of adult heads…"
(The student <u>did not answer the second part of the question</u>, even with the hint of "<u>***and***</u>") Score=P

2007 #2c: There are several issues here. First, the writing is very disjointed and confusing to read. Second, there is a lack of <u>precision of language</u> with the phrase "change the actual data" and FIVE usages of the word "different." But the third is the fatal one: "eliminate any variables." This is a false statement—blocking does not eliminate variables. Score = P.
Takeaways: 1) use precise language, and 2) answer the question, then stop writing.

2012 #4: "…adults who said 'yes' to the stated question…" is referring to the ***sample,*** not the ***population.*** The hypotheses should always describe the <u>population parameters</u>. Score=P

2013 #1 [FYI: In Part (a), students need to <u>show "work,"</u> even if it's only a fraction: 4/23=17.4%).]
In Part (b), the added phrase "in repeated samples" makes the statement incorrect. In fact, this statement is neither a correct <u>interpretation of the interval</u> NOR a correct <u>explanation of confidence LEVEL</u>. Score on this part = I (they earned E's for conditions/name and interval).

2014 #1c: No <u>linkage from p-value to decision</u> ("since the p-value is greater than alpha=0.05…") AND, wrong conclusion based on a high p-value. Score = I

2014 #2b: This student <u>incorrectly interprets the probability</u> that was calculated. The CORRECT interpretation should be "Given the process is random, there is a 1.2% chance of selecting three women." [Even though this was not a hypothesis test, what the student stated is essentially the probability that the null hypothesis ("The process is random.") is true.]
The student DID make the correct decision AND supported it by comparing it to a reasonable standard (5%), so their score is a P.

2012 #4: Looking at their table, p_1 and p_2 are the statistics (should be parameters); t and p-value are correct but formula is for NON-pooled test. Both sections earned a P. Overall score: PPEE = 3.