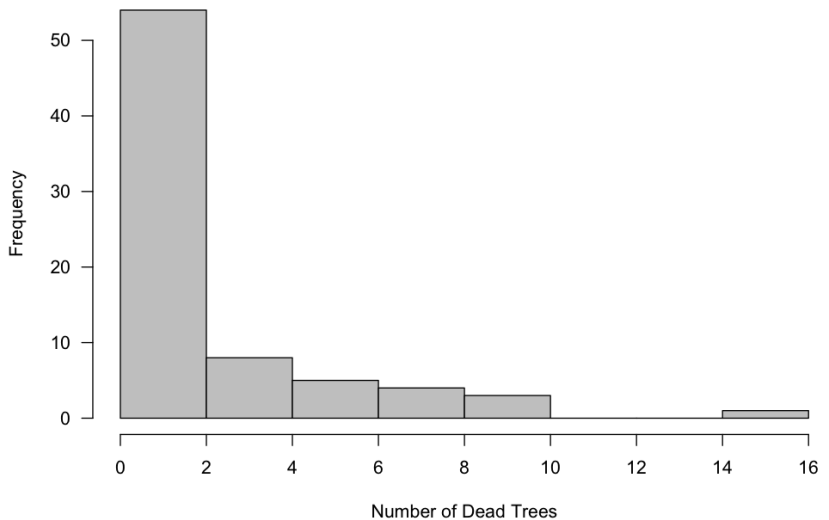


AP Statistics  
2020 Mock Exam

1. In a time of extended drought, citizens in a small city in Southern California were required to restrict how often they ran water to their lawns. As a result, some of the trees in the city died to a lack of water. A concerned citizen decides to take a sample in order to measure the number of dying trees. He uses a city map that divides the residential homes into several thousand blocks. He randomly selects 75 blocks. For each of these 75 blocks, he counts the number of dead trees on each block. Here is a histogram of his data, along with summary statistics. Each bin on the histogram has a width of two. The first bin contains 0 and 1, the next contains 2 and 3, and so on.

n	75
mean	2.36
s	2.883
se	0.333
median	1



- a) Describe the distribution of the number of dead trees.
- b) Interpret the sample standard deviation,  $s$ , in context.

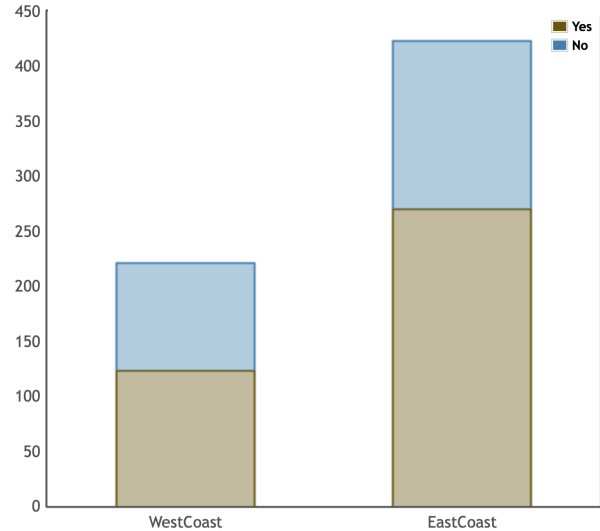
AP Statistics  
2020 Mock Exam

- c) Name the sampling method used by this citizen to estimate the number of dead trees per each home. Describe why this method of randomly selecting blocks was a more practical choice compared to using a simple random sample of homes.
- d) Estimate the probability that a randomly selected block in the city will have less than six dead trees.
- e) We will assume that each block is independent of the next. Given that 19 of the city blocks had no dead trees, what is the estimated probability that if 4 city blocks were chosen these blocks would have at least one dead tree?
- f) For further analysis, the concerned citizen is considering creating a 95%  $t$  confidence interval to estimate the average number of dead trees. Are the conditions satisfied for inference? Explain why or why not.

AP Statistics  
2020 Mock Exam

2. Researchers were curious if different regions of the country differed in their use of a new, popular, social media app. Two independent random samples were taken. One sample from a city on the west coast of America and another from the east coast. College students aged 18-25 were asked if they have used the app. Their responses are organized in a table and graph below.

	West Coast	East Coast	Total
Yes	123	270	393
No	98	153	251
Total	221	423	644



- a) Does there appear to be an association between which coast a student is from and whether or not they have used this social media app? (Note: Do **not** answer this question using an inference procedure.)
- b) What is the probability that a randomly selected student is from the west coast, given that this student has used the social media app?
- c) A curious statistics student wants to test if the proportion of students who have used the app is greater on the East Coast than on the West Coast. Write the appropriate null and alternative hypotheses for this hypothesis test.

AP Statistics  
2020 Mock Exam

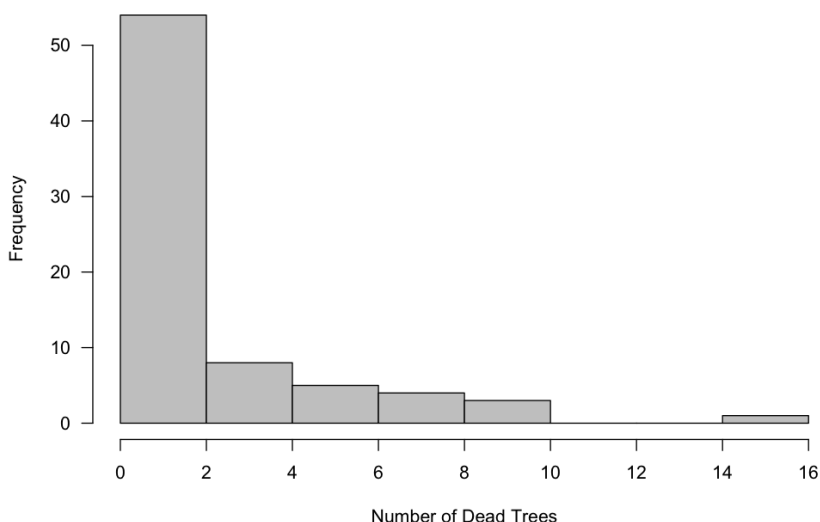
- d) The P-value for the hypothesis test is 0.022. Based on this  $p$ -value, what conclusion should the researcher make?
- e) In another study, college students on the East Coast were compared with 85 students in the Midwest and  $p$ -value of 0.13 was found. However, there is concern that a Type II has been made. Describe how a different sample size would decrease the likelihood of a Type II error.

AP Statistics  
2020 Mock Exam

**Solutions**

1. In a time of extended drought, citizens in a small city in Southern California were required to restrict how often they ran water to their lawns. As a result, some of the trees in the city died to a lack of water. A concerned citizen decides to take a sample in order to measure the number of dying trees. He uses a city map that divides the residential homes into several thousand blocks. He randomly selects 75 blocks. For each of these 75 blocks, he counts the number of dead trees on each block. Here is a histogram of his data, along with summary statistics. Each bin on the histogram has a width of two. The first bin contains 0 and 1, the next contains 2 and 3, and so on.

n	75
mean	2.36
s	2.883
se	0.333
median	1



- a) Describe the distribution of the number of dead trees.

*The distribution of the number of dead trees is skewed to the right and has a center around 1. Most of the data is from zero to four dead trees and there is an outlier around 14 or 15.*

*Scoring: 5 components: center, shape, spread, outlier, context.  
Partially correct for 3 or 4.*

- b) Interpret the sample standard deviation,  $s$ , in context.

*The typical distance of the number of dead trees from the mean is 2.883.*

*Scoring: Context is required. Distance from the mean should be part of the explanation.  
2007 #1 rubric can be consulted.*

AP Statistics  
2020 Mock Exam

**Solutions**

- c) Name the sampling method used by this citizen to estimate the number of dead trees per each home. Describe why this method of randomly selecting blocks was a more practical choice compared to using a simple random sample of homes.

*\*\*This is a cluster sample where the blocks are the clusters.*

*\*\*This is more practical because once the researcher had driven to a selected block, he could survey many homes at once. If he took a SRS he would have to drive to very, very many different places.*

*Scoring: Three components:*

*\*Correct Name (cluster)*

*\*Advantage of clusters*

*\*Disadvantage of SRS*

*Partial can be earned for 2/3 components OR for stratified with a strong communication regarding advantage of strata compared to SRS.*

- d) Estimate the probability that a randomly selected block in the city will have less than six dead trees.

$$(52 + 8 + 5)/75 = 86.7\%$$

*Scoring: Work shown + reasonable answer required*

*Any reasonable estimation of frequencies should be accepted.*

*Any use of any probability distribution (normal, binomial, etc..) will receive no points.*

- e) We will assume that each block is independent of the next. Given that 19 of the city blocks had no dead trees, what is the estimated probability that if 4 city blocks were chosen these blocks would have at least one dead tree?

$$1 - (1-19/75)^4 = 68.9\%$$

*Scoring: work shown and correct answer for E.*

*Partial for the complement of the correct answer or for using the complement of  $(19/75)^4$*

- f) For further analysis, the concerned citizen is considering creating a 95% *t* confidence interval to estimate the average number of dead trees. Are the conditions satisfied for inference? Explain why or why not.

*Yes. Random sample,  $75 < 10\%$  of thousands,  $75 > 30$ , thus the sample mean is approximately normal.*

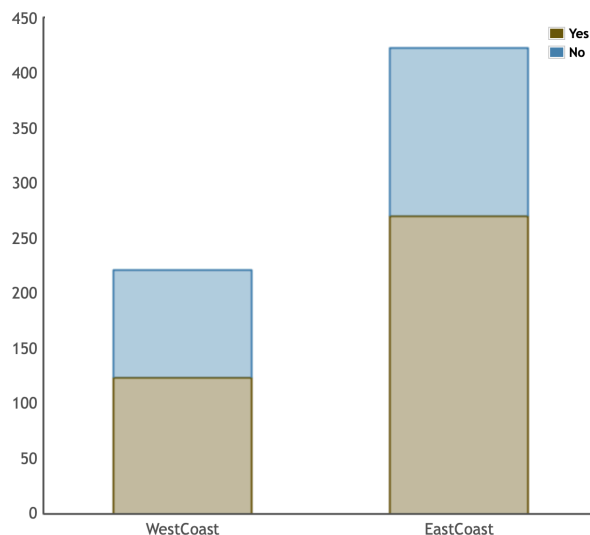
*Scoring: All three components for E. 2/3 for a P.*

AP Statistics  
2020 Mock Exam

**Solutions**

2. Researchers were curious if different regions of the country differed in their use of a new, popular, social media app. Two independent random samples were taken. One sample from a city on the west coast of America and another from the east coast. College students aged 18-25 were asked if they have used the app. Their responses are organized in a table and graph below.

	West Coast	East Coast	Total
Yes	123	270	393
No	98	153	251
Total	221	423	644



- a) Does there appear to be an association between which coast a student is from and whether or not they have used this social media app? (Note: Do **not** answer this question using an inference procedure.)

*Yes, there is an association. It appears that East Coast students are more likely to have used the app than West Coast students. This is apparent from the graph and from the percents on the table (63.8% yes vs 55.7% yes)*

*Scoring: An association with differences on either table, frequencies, %'s, or graph required. If student argues No association and explains that the differences are not that large, student may earn E.*

- b) What is the probability that a randomly selected student is from the west coast, given that this student has used the social media app?

$$123/393 = 31.3\%$$

*Scoring: Work shown and correct answer for E  
Partial can be earned for reversing the condition.*

- c) A curious statistics student wants to test if the proportion of students who have used the app is greater on the East Coast than on the West Coast. Write the appropriate null and alternative hypotheses for this hypothesis test.

*Ho:  $p_e = p_w$ ; Ha:  $p_e > p_w$ ;  $p_e$  = true % of east coast students who have used the app;  $p_w$  = true % of west coast students who have used the app.*

*Scoring: proper symbols, > (direction), symbols defined correctly.*

AP Statistics  
2020 Mock Exam

**Solutions**

- d) The P-value for the hypothesis test is 0.022. Based on this  $p$ -value, what conclusion should the researcher make?

*With a  $p$ -value of  $0.022 < \alpha$ , we reject  $H_0$ . We have found statistically significant evidence that the percentage of east coast students who have used this app is greater than the percentage of west coast students who have used this app.*

*Scoring: Correct decision, linked to alpha, correct conclusion.  
Partial for 2/3 components.*

- e) In another study, college students on the East Coast were compared with 85 students in the Midwest and  $p$ -value of 0.13 was found. However, there is concern that a Type II has been made. Describe how a different sample size would decrease the likelihood of a Type II error.

*If the sample size(s) were larger, this would reduce the probability of a type II error. Increasing the sample size(s) would decrease the standard deviation of the sampling distribution of the sample proportions. This would increase our ability to detect a difference between the two populations.*

*Scoring:*

*\*\*Larger  $n$*

*\*\*Standard deviation of the sampling distribution decreases*

*\*\*Ability to detect a difference*

*Partial for:*

*\*\*Weak communication*

*\*\*Insufficient vocabulary (“sampling distribution” missing)*



AP Statistics  
2020 Mock Exam

Teacher Notes:

This exam was configured with 11 parts for several reasons.

- A 6/5 split for the two questions is close to the stated 55/45% weighting.
- Old FRQ's have 4 parts and are to be finished in 13 minutes. CB will want students writing for the full 25/15 minute time allotments to prevent cheating. Thus the exam will be longer than shorter.

Reading the percentage breakdowns in the CED, I concluded that a reasonable split of the 11 parts is as follows.

- 3 parts Exploring Data—Units 1 & 2
- 1 part Data Collection—Unit 3
- 3 parts Probability—Units 4 & 5
- 4 parts Inference—Units 6 & 7

I don't have any inside information. These are guesses as to the nature of the exam. As my friend Corey says, the best way to help students pass the AP Statistics exam is to help students understand statistics. My biggest goal was to write questions where students need to demonstrate statistical understanding in order to answer the questions.

As much as possible, I avoided questions where the answer could be given using a memorized phrase. This proved to be a difficult goal!

My thanks to Jeff Eicher for editorial input and a nicer histogram. If you find things you don't like, Jeff probably told me to fix it and I ignored him.