Comprehensive **A**ssessment of **O**utcomes for a first course in **S**tatistics (CAOS)

CAOS 4

Developed by the Web ARTIST Project
https://app.gen.umn.edu/artist/

Funded by a grant from the National Science Foundation
NSF CCLI ASA- 0206571

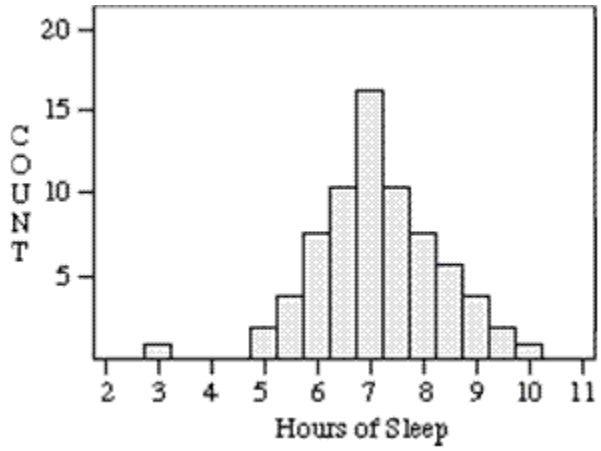
Principal Investigators:
Joan Garfield and Bob delMas, University of Minnesota
Beth Chance, Cal Poly – San Luis Obispo
Post-doctoral Research Assistant:
Ann Ooms, University of Minnesota

Version 31

September 8, 2005

ARTIST CAOS 4 POSTTEST

The following graph shows a distribution of hours slept last night by a group of college students.



1. Select the statement below that gives the most complete description of the graph in a way that demonstrates an understanding of how to statistically describe and interpret the distribution of a variable.

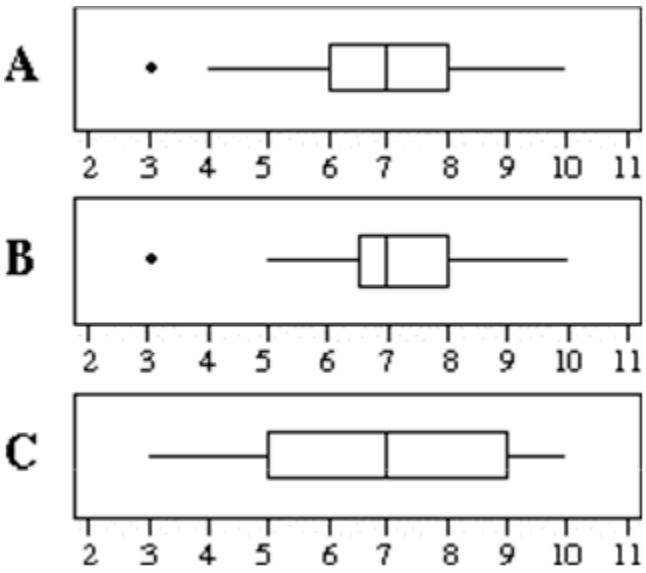
    a. The bars go from 3 to 10, increasing in height to 7, then decreasing to 10. The tallest bar is at 7. There is a gap between three and five.

    b. The distribution is normal, with a mean of about 7 and a standard deviation of about 1.

    c. Most students seem to be getting enough sleep at night, but some students slept more and some slept less. However, one student must have stayed up very late and got very few hours of sleep.

    d. The distribution of hours of sleep is somewhat symmetric and bell-shaped, with an outlier at 3. The typical amount of sleep is about 7 hours and overall range is 7 hours.

2. Which box plot seems to be graphing the same data as the histogram in question 1?

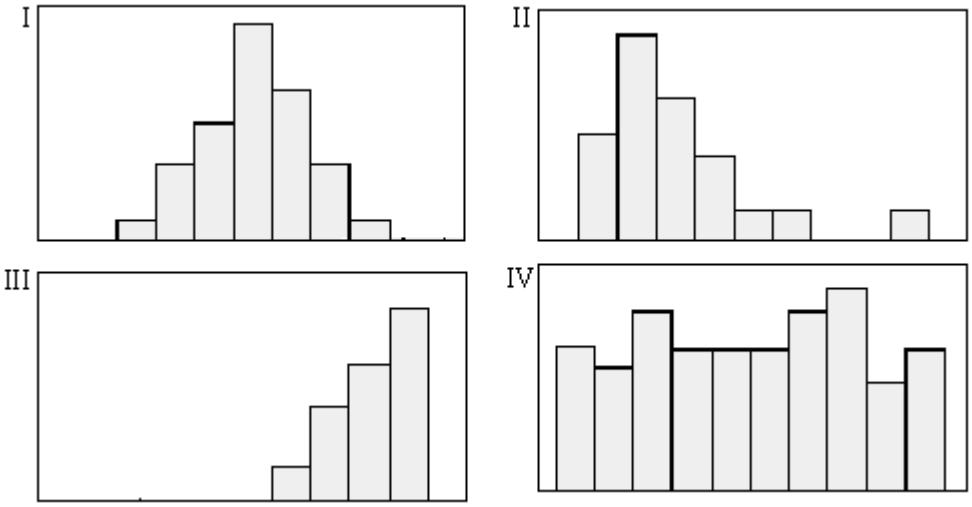


a. Boxplot A.

b. Boxplot B.

c. Boxplot C.

Items 3 to 5 refer to the following situation:
Four histograms are displayed below. For each item, match the description to the appropriate histogram.



3. A distribution for a set of quiz scores where the quiz was very easy is represented by:
    a. Histogram I.
    b. Histogram II.
    c. Histogram III.
    d. Histogram IV.

4. A distribution for a set of wrist circumferences (measured in centimeters) taken from the right wrist of a random sample of newborn female infants is represented by:
    a. Histogram I.
    b. Histogram II.
    c. Histogram III.
    d. Histogram IV.

5. A distribution for the last digit of phone numbers sampled from a phone book (i.e., for the phone number 968-9667, the last digit, 7, would be selected) is represented by:
    a. Histogram I.
    b. Histogram II.
    c. Histogram III.
    d. Histogram IV.

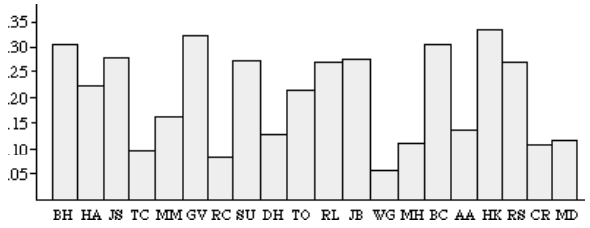
6. A baseball fan likes to keep track of statistics for the local high school baseball team. One of the statistics she recorded is the proportion of hits obtained by each player based on the number of times at bat as shown in the table below. Which of the following graphs gives the best display of the distribution of proportion of hits in that it allows the baseball fan to describe the shape, center and spread of the variable, proportion of hits?

Player	Proportion of hits
BH	0.305
HA	0.229
JS	0.281
TC	0.097
MM	0.167
GV	0.333
RC	0.085

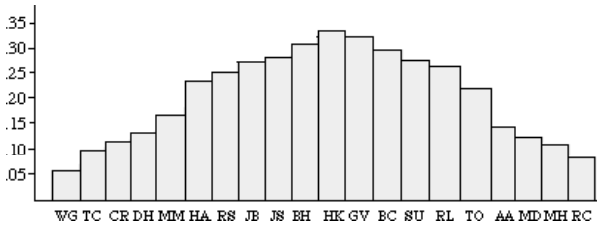
Player	Proportion of hits
SU	0.270
DH	0.136
TO	0.218
RL	0.267
JB	0.270
WG	0.054
MH	0.108

Player	Proportion of hits
BC	0.301
AA	0.143
HK	0.341
RS	0.261
CR	0.115
MD	0.125

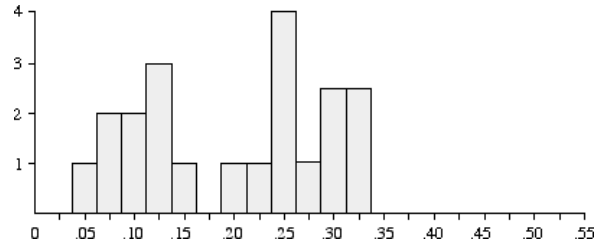
A



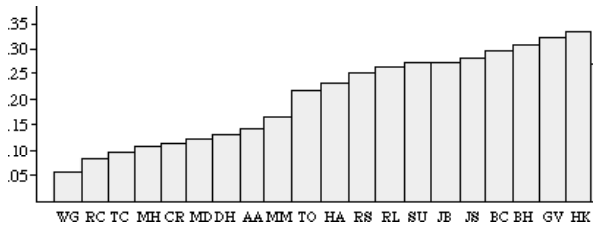
B



C



D



7. A recent research study randomly divided participants into groups who were given different levels of Vitamin E to take daily. One group received only a placebo pill. The research study followed the participants for eight years to see how many developed a particular type of cancer during that time period. Which of the following responses gives the best explanation as to the purpose of randomization in this study?

    a. To increase the accuracy of the research results.

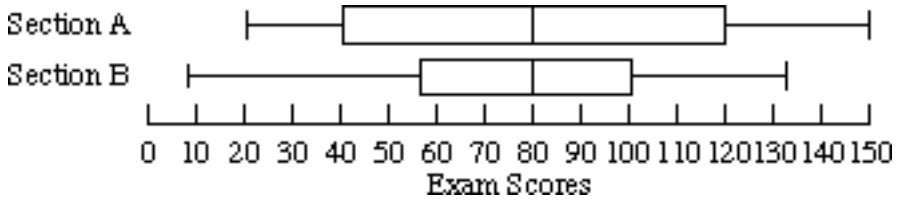
    b. To ensure that all potential cancer patients had an equal chance of being selected for the study.

    c. To reduce the amount of sampling error.

    d. To produce treatment groups with similar characteristics.

    e. To prevent skewness in the results.

Items 8 to 10 refer to the following situation:
The two boxplots below display final exam scores for all students in two different sections of the same course.



8. Which section would you expect to have a greater standard deviation in exam scores?

    a. Section A.

    b. Section B.

    c. Both sections are about equal.

    d. It is impossible to tell.

9. Which data set has a greater percentage of students with scores at or below 30?

    a. Section A.

    b. Section B.

    c. Both sections are about equal.

    d. It is impossible to tell.

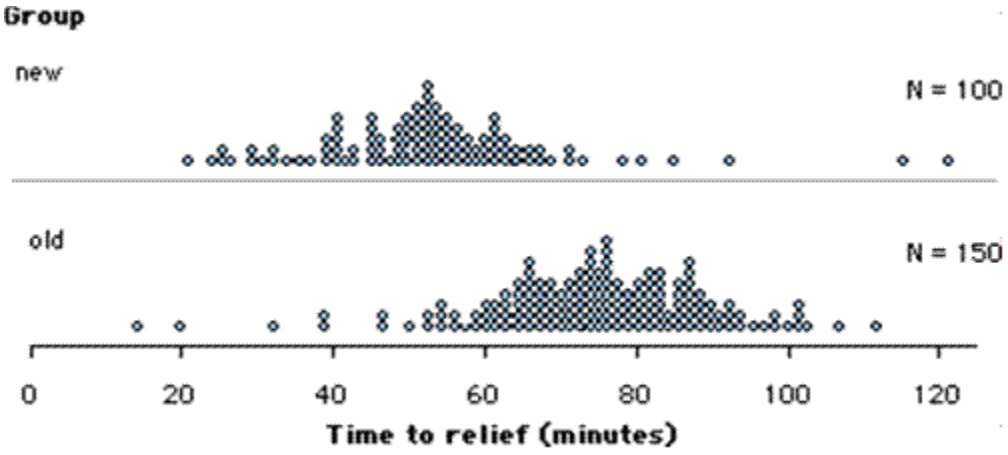
10. Which section has a greater percentage of students with scores at or above 80?

    a. Section A.

    b. Section B.

    c. Both sections are about equal.

Items 11 to 13 refer to the following situation:
A drug company developed a new formula for their headache medication. To test the effectiveness of this new formula, 250 people were randomly selected from a larger population of patients with headaches. 100 of these people were randomly assigned to receive the new formula medication when they had a headache, and the other 150 people received the old formula medication. The time it took, in minutes, for each patient to no longer have a headache was recorded. The results from both of these clinical trials are shown below. Items 11, 12, and 13 present statements made by three different statistics students. For each statement, indicate whether you think the student's conclusion is valid.



11. The old formula works better. Two people who took the old formula felt relief in less than 20 minutes, compared to none who took the new formula. Also, the worst result - near 120 minutes - was with the new formula.

    a. Valid.

    b. Not valid.

12. The average time for the new formula to relieve a headache is lower than the average time for the old formula. I would conclude that people taking the new formula will tend to feel relief about 20 minutes sooner than those taking the old formula.

    a. Valid.

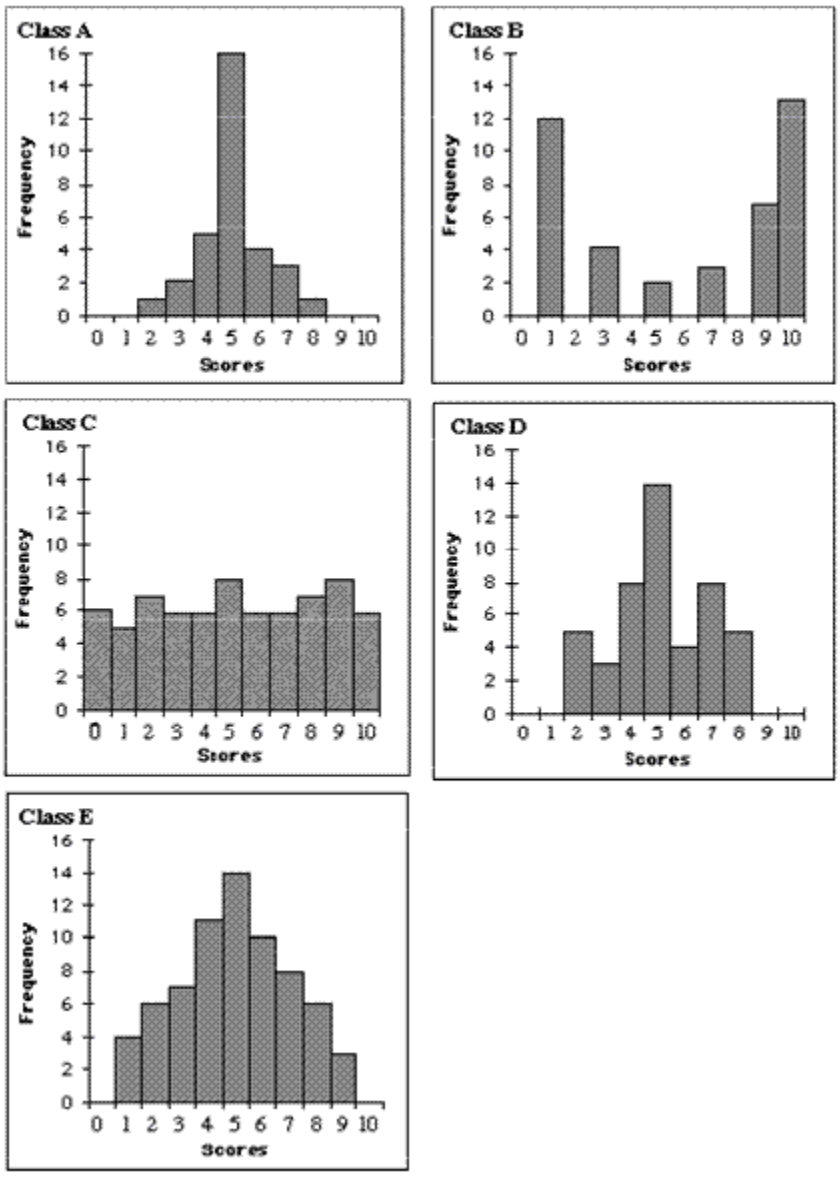
    b. Not valid.

13. I would not conclude anything from these data. The number of patients in the two groups is not the same so there is no fair way to compare the two formulas.

    a. Valid.

    b. Not valid.

Items 14 and 15 refer to the following situation:
Five histograms are presented below. Each histogram displays test scores on a scale of 0 to 10 for one of five different statistics classes.



14. Which of the classes would you expect to have the lowest standard deviation, and why?

    a. Class A, because it has the most values close to the mean.

    b. Class B, because it has the smallest number of distinct scores.

    c. Class C, because there is no change in scores.

    d. Class A and Class D, because they both have the smallest range.

    e. Class E, because it looks the most normal.

15. Which of the classes would you expect to have the highest standard deviation, and why?

    a. Class A, because it has the largest difference between the heights of the bars.

    b. Class B, because more of its scores are far from the mean.

    c. Class C, because it has the largest number of different scores.

    d. Class D, because the distribution is very bumpy and irregular.

    e. Class E, because it has a large range and looks normal.

16. A certain manufacturer claims that they produce 50% brown candies. Sam plans to buy a large family size bag of these candies and Kerry plans to buy a small fun size bag. Which bag is more likely to have more than 70% brown candies?

    a. Sam, because there are more candies, so his bag can have more brown candies.

    b. Sam, because there is more variability in the proportion of browns among larger samples.

    c. Kerry, because there is more variability in the proportion of browns among smaller samples.

    d. Kerry, because most small bags will have more than 50% brown candies.

    e. Both have the same chance because they are both random samples.

17. Imagine you have a barrel that contains thousands of candies with several different colors. We know that the manufacturer produces 35% yellow candies. Five students each take a random sample of 20 candies, one at a time, and record the percentage of yellow candies in their sample. Which sequence below is the most plausible for the percent of yellow candies obtained in these five samples?

    a. 30%, 35%, 15%, 40%, 50%.

    b. 35%, 35%, 35%, 35%, 35%.

    c. 5%, 60%, 10%, 50%, 95%.

    d. Any of the above.

18. Jean lives about 10 miles from the college where she plans to attend a 10-week summer class. There are two main routes she can take to the school, one through the city and one through the countryside. The city route is shorter in miles, but has more stoplights. The country route is longer in miles, but has only a few stop signs and stoplights. Jean sets up a randomized experiment where each day she tosses a coin to decide which route to take that day. She records the following data for 5 days of travel on each route.

    Country Route - 17, 15, 17, 16, 18

    City Route - 18, 13, 20, 10, 16

    It is important to Jean to arrive on time for her classes, but she does not want to arrive too early because that would increase her parking fees. Based on the data gathered, which route would you advise her to choose?

    a. The Country Route, because the times are consistently between 15 and 18 minutes.

    b. The City Route, because she can get there in 10 minutes on a good day and the average time is less than for the Country Route.

    c. Because the times on the two routes have so much overlap, neither route is better than the other. She might as well flip a coin.

19. A graduate student is designing a research study. She is hoping to show that the results of an experiment are statistically significant. What type of *p*-value would she want to obtain?

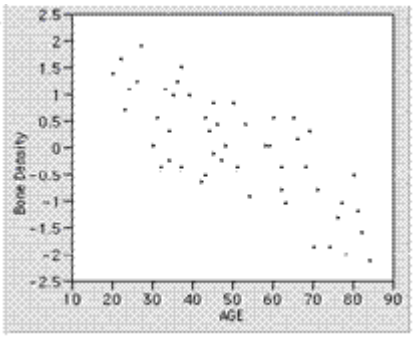
    a. A large *p*-value.

    b. A small *p*-value.

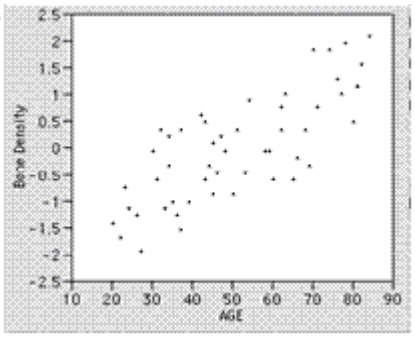
    c. The magnitude of a *p*-value has no impact on statistical significance.

20. Bone density is typically measured as a standardized score with a mean of 0 and a standard deviation of 1. Lower scores correspond to lower bone density. Which of the following graphs shows that as women grow older they tend to have lower bone density?

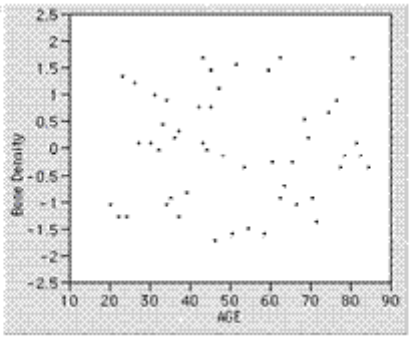
A



B



C

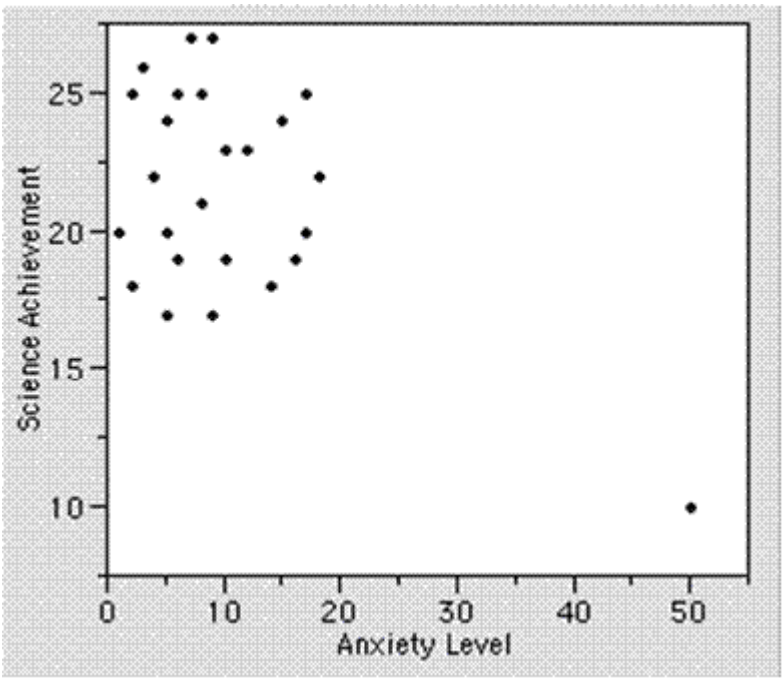


a. Graph A.

b. Graph B.

c. Graph C.

21. The following scatterplot shows the relationship between scores on an anxiety scale and an achievement test for science. Choose the best interpretation of the relationship between anxiety level and science achievement based on the scatterplot.



a. This graph shows a strong negative linear relationship between anxiety and achievement in science.

b. This graph shows a moderate linear relationship between anxiety and achievement in science.

c. This graph shows very little, if any, linear relationship between anxiety and achievement in science.

22. Researchers surveyed 1,000 randomly selected adults in the U.S. A statistically significant, strong positive correlation was found between income level and the number of containers of recycling they typically collect in a week. Please select the best interpretation of this result.

a. We can not conclude whether earning more money causes more recycling among U.S. adults because this type of design does not allow us to infer causation.

b. This sample is too small to draw any conclusions about the relationship between income level and amount of recycling for adults in the U.S.

c. This result indicates that earning more money influences people to recycle more than people who earn less money.

Items 23 and 24 refer to the following situation:
A researcher in environmental science is conducting a study to investigate the impact of a particular herbicide on fish. He has 60 healthy fish and randomly assigns each fish to either a treatment or a control group. The fish in the treatment group showed higher levels of the indicator enzyme.

23. Suppose a test of significance was correctly conducted and showed no statistically significant difference in average enzyme level between the fish that were exposed to the herbicide and those that were not. What conclusion can the graduate student draw from these results?

    a. The researcher must not be interpreting the results correctly; there should be a significant difference.

    b. The sample size may be too small to detect a statistically significant difference.

    c. It must be true that the herbicide does not cause higher levels of the enzyme.

24. Suppose a test of significance was correctly conducted and showed a statistically significant difference in average enzyme level between the fish that were exposed to the herbicide and those that were not. What conclusion can the graduate student draw from these results?

    a. There is evidence of association, but no causal effect of herbicide on enzyme levels.

    b. The sample size is too small to draw a valid conclusion.

    c. He has proven that the herbicide causes higher levels of the enzyme.

    d. There is evidence that the herbicide causes higher levels of the enzyme for these fish.

Items 25 to 27 refer to the following situation:
A research article reports the results of a new drug test. The drug is to be used to decrease vision loss in people with Macular Degeneration. The article gives a *p*-value of .04 in the analysis section. Items 25, 26, and 27 present three different interpretations of this *p*-value. Indicate if each interpretation is valid or invalid.

25. The probability of getting results as extreme as or more extreme than the ones in this study if the drug is actually not effective.
    a. Valid.
    b. Invalid.

26. The probability that the drug is not effective.
    a. Valid.
    b. Invalid.

27. The probability that the drug is effective.
    a. Valid.
    b. Invalid.

Items 28 to 31 refer to the following situation:
A high school statistics class wants to estimate the average number of chocolate chips in a generic brand of chocolate chip cookies. They collect a random sample of cookies, count the chips in each cookie, and calculate a 95% confidence interval for the average number of chips per cookie (18.6 to 21.3). Items 28, 29, 30 and 31 present four different interpretations of these results. Indicate if each interpretation is valid or invalid.

28. We are 95% certain that each cookie for this brand has approximately 18.6 to 21.3 chocolate chips.
    a. Valid.
    b. Invalid.

29. We expect 95% of the cookies to have between 18.6 and 21.3 chocolate chips.
    a. Valid.
    b. Invalid.

30. We would expect about 95% of all possible sample means from this population to be between 18.6 and 21.3 chocolate chips.
    a. Valid.
    b. Invalid.

31. We are 95% certain that the confidence interval of 18.6 to 21.3 includes the true average number of chocolate chips per cookie.
    a. Valid.
    b. Invalid.

32. It has been established that under normal environmental conditions, adult largemouth bass in Silver Lake have an average length of 12.3 inches with a standard deviation of 3 inches. People who have been fishing Silver Lake for some time claim that this year they are catching smaller than usual largemouth bass. A research group from the Department of Natural Resources took a random sample of 100 adult largemouth bass from Silver Lake and found the mean of this sample to be 11.2 inches. Which of the following is the most appropriate statistical conclusion?

    a. The researchers cannot conclude that the fish are smaller than what is normal because 11.2 inches is less than one standard deviation from the established mean (12.3 inches) for this species.

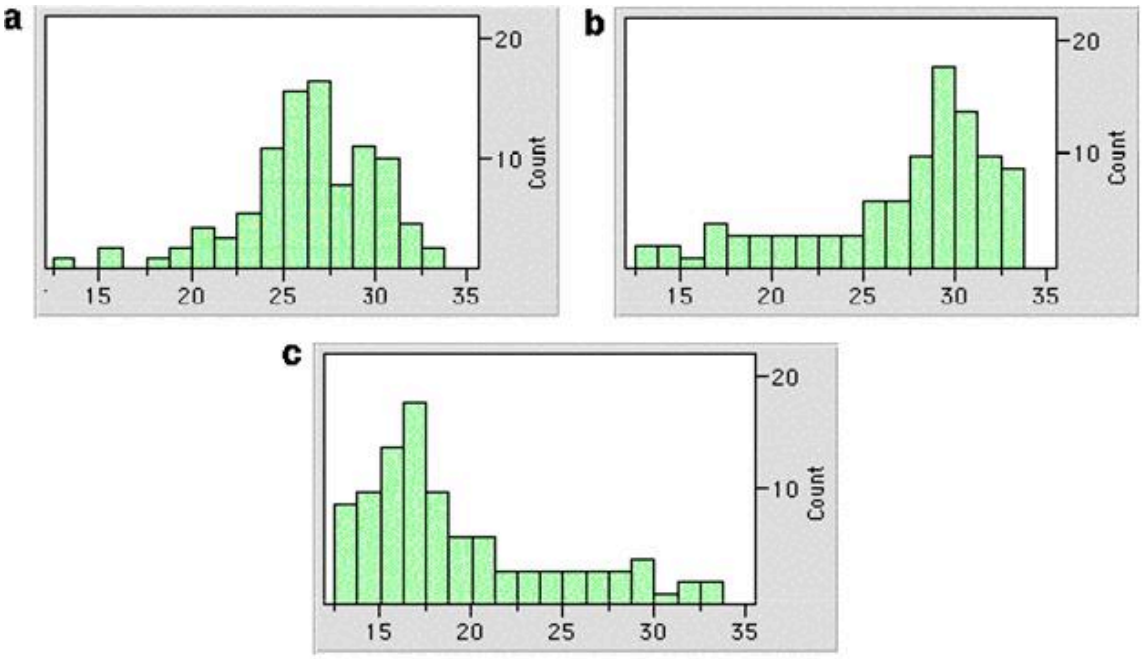
    b. The researchers can conclude that the fish are smaller than what is normal because the sample mean should be almost identical to the population mean with a large sample of 100 fish.

    c. The researchers can conclude that the fish are smaller than what is normal because the difference between 12.3 inches and 11.2 inches is much larger than the expected sampling error.

A study examined the length of a certain species of fish from one lake. The plan was to take a random sample of 100 fish and examine the results. Numerical summaries on lengths of the fish measured in this study are given.

Mean	26.8mm
Median	29.4mm
Standard Deviation	5.0 mm
Minimum	12.mm
Maximum	33.4mm

33. Which of the following histograms is most likely to be the one for these data?

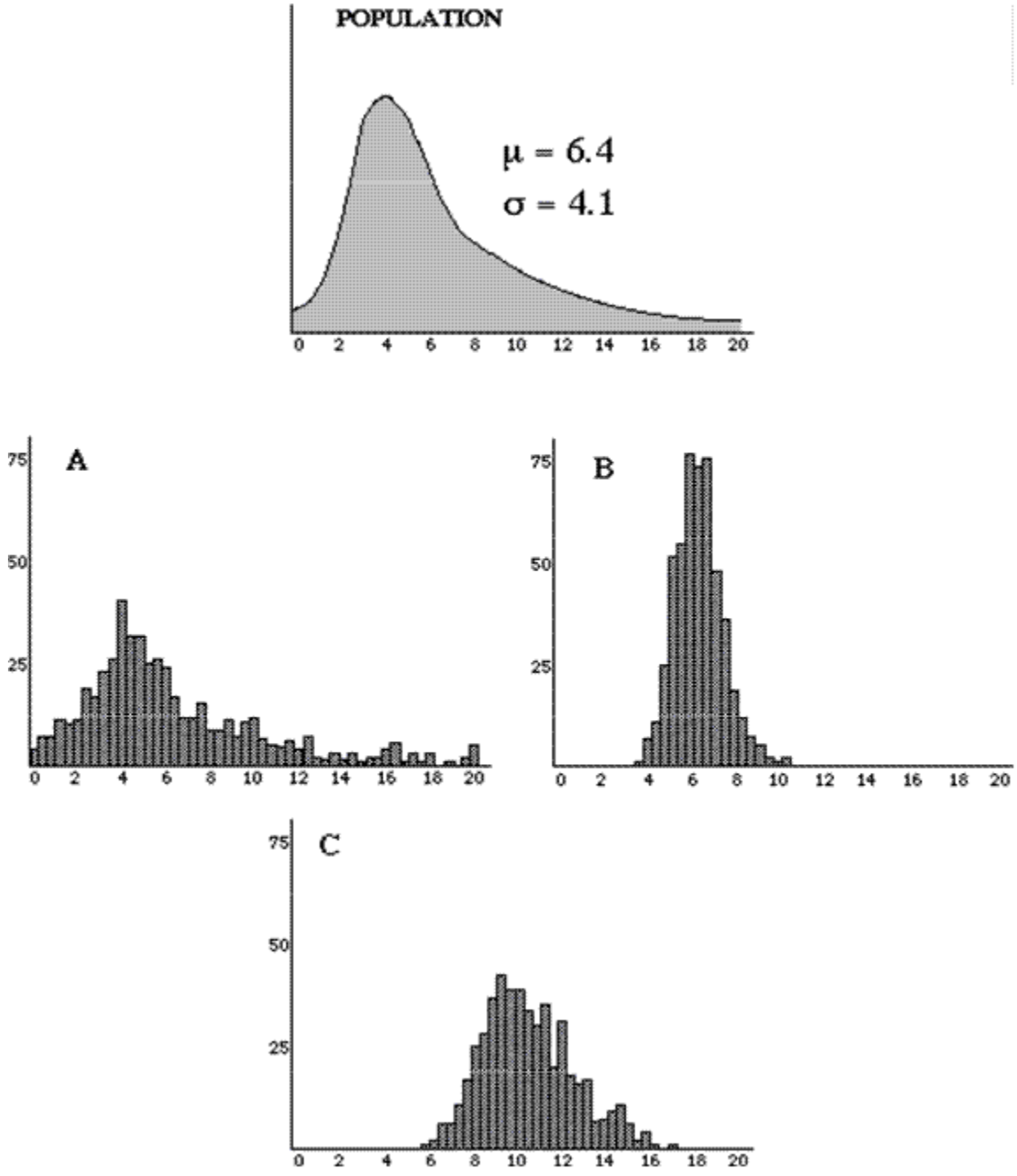


a. Histogram a.

b. Histogram b.

c. Histogram c.

Items 34 and 35 refer to the following situation:
Four graphs are presented below. The graph at the top is a distribution for a population of test scores. The mean score is 6.4 and the standard deviation is 4.1.



34. Which graph (A, B, or C) do you think represents a single random sample of 500 values from this population?

    a. Graph A

    b. Graph B

    c. Graph C

35. Which graph (A, B, or C) do you think represents a distribution of 500 sample means from random samples each of size 9?

    a. Graph A

    b. Graph B

    c. Graph C

36. This table is based on records of accidents compiled by a State Highway Safety and Motor Vehicles Office. The Office wants to decide if people are less likely to have a fatal accident if they are wearing a seatbelt. Which of the following comparisons is most appropriate for supporting this conclusion?

Safety Equipment in Use	Injury		ROW TOTAL
	Nonfatal	Fatal	
Seat Belt	412,368	510	412,878
No Seat Belt	162,527	1,601	164,128
COLUMN TOTAL	574,895	2,111	577,006

    a. Compare the ratios 510/412,878 and 1,601/164,128

    b. Compare the ratios 510/577,006 and 1,601/577,006

    c. Compare the numbers 510 and 1,601

37. A student participates in a Coke versus Pepsi taste test. She correctly identifies which soda is which four times out of six tries. She claims that this proves that she can reliably tell the difference between the two soft drinks. You have studied statistics and you want to determine the probability of anyone getting at least four right out of six tries just by chance alone. Which of the following would provide an accurate estimate of that probability?

    a. Have the student repeat this experiment many times and calculate the percentage time she correctly distinguishes between the brands.

    b. Simulate this on the computer with a 50% chance of guessing the correct soft drink on each try, and calculate the percent of times there are four or more correct guesses out of six trials.

    c. Repeat this experiment with a very large sample of people and calculate the percentage of people who make four correct guesses out of six tries.

    d. All of the methods listed above would provide an accurate estimate of the probability.

38. A college official conducted a survey to estimate the proportion of students currently living in dormitories about their preference for single rooms, double rooms, or multiple (more than two people) rooms in the dormitories on campus. Which of the following does NOT affect the college official's ability to generalize the survey results to all dormitory students?

    a. Five thousand students live in dormitories on campus. A random sample of only 500 were sent the survey.

    b. The survey was sent to only first-year students.

    c. Of the 500 students who were sent the survey, only 160 responded.

    d. All of the above present a problem for generalizing the results.

39. The number of people living on American farms has declined steadily during the last century. Data gathered on the U.S. farm population (millions of people) from 1910 to 2000 were used to generate the following regression equation: Predicted Farm Population = 1167 - .59 (YEAR). Which method is best to use to predict the number of people living on farms in 2050?

    a. Substitute the value of 2050 for YEAR in the regression equation, and compute the predicted farm population.

    b. Plot the regression line on a scatterplot, locate 2050 on the horizontal axis, and read off the corresponding value of population on the vertical axis.

    c. Neither method is appropriate for making a prediction for the year 2050 based on these data.

    d. Both methods are appropriate for making a prediction for the year 2050 based on these data.

40. The following situation models the logic of a hypothesis test. An electrician uses an instrument to test whether or not an electrical circuit is defective. The instrument sometimes fails to detect that a circuit is good and working. The null hypothesis is that the circuit is good (not defective). The alternative hypothesis is that the circuit is not good (defective). If the electrician rejects the null hypothesis, which of the following statements is true?

    a. The circuit is definitely not good and needs to be repaired.

    b. The electrician decides that the circuit is defective, but it could be good.

    c. The circuit is definitely good and does not need to be repaired.

    d. The circuit is most likely good, but it could be defective.