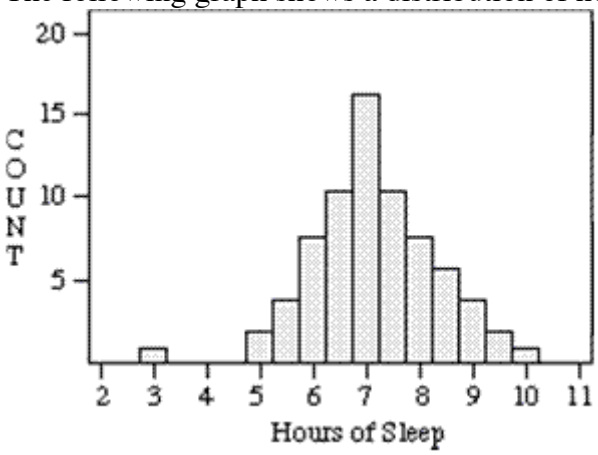
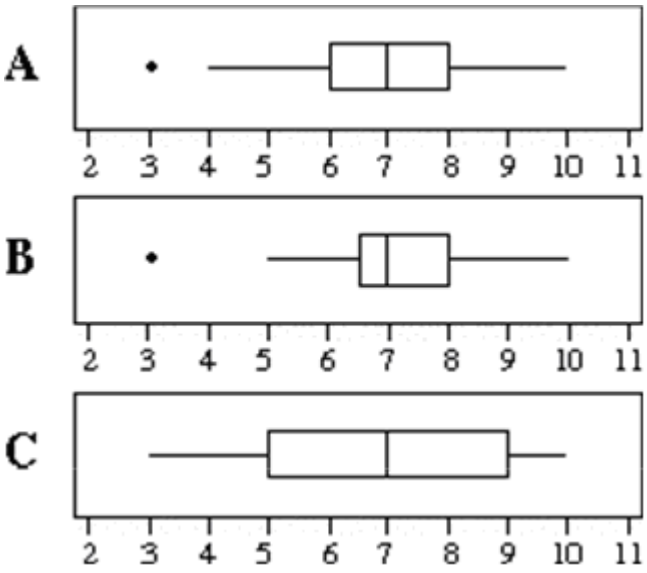
The following graph shows a distribution of hours slept last night by a group of college students.



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



Boxplot A.

Boxplot B.

Boxplot C.

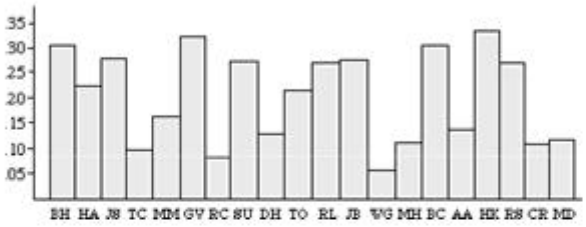
1b. A baseball fan likes to keep track of statistics for the local high school baseball team. One of the statistics she recorded is the proportion of hits obtained by each player based on the number of times at bat as shown in the table below. Which of the following graphs gives the best display of the 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

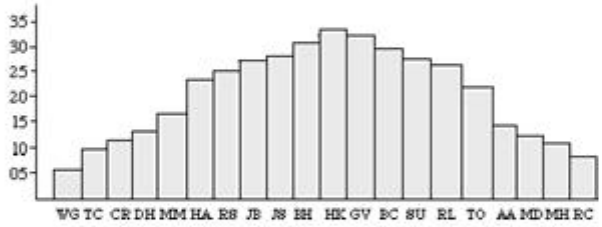
Player	Proportion of hits
SU	0.270
DH	0.136
TO	0.218
RL	0.267
JB	0.270
WG	0.054
MH	0.108

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

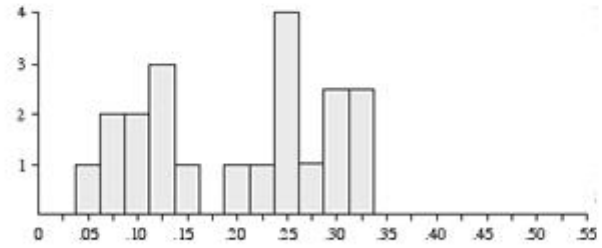
A



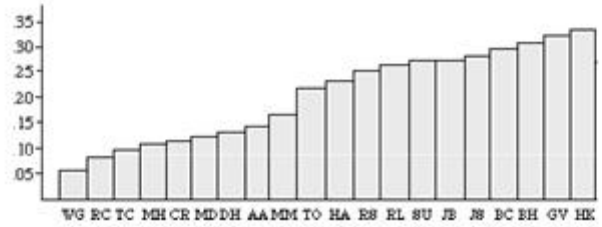
B



C



D



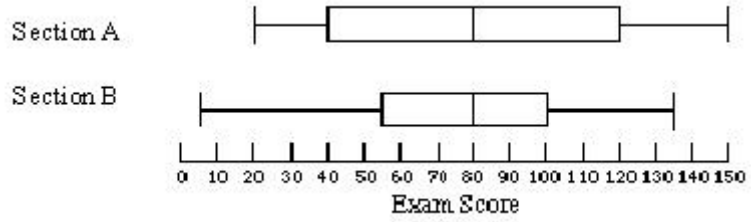
Graph A.

Graph B.

Graph C.

Graph D.

The two boxplots below display final exam scores for all students in two different sections of the same course.



2a. Which section would you expect to have a greater standard deviation in exam scores?

Section A.

Section B.

Both sections are about equal.

It is impossible to tell.

2b. Which data set has a greater percentage of students with scores at or below 30?

Section A.

Section B.

Both sections are about equal.

It is impossible to tell.

2c. Which section has a greater percentage of students with scores at or above 80?

Section A.

Section B.

Both sections are about equal.

3a. 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?

To increase the accuracy of the research results.

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

To reduce the amount of sampling error.

To produce treatment groups with similar characteristics.

To prevent skewness in the results.

3b. 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?

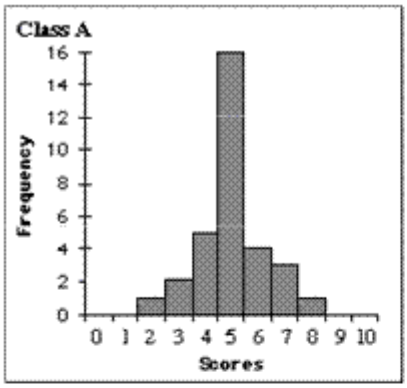
30%, 35%, 15%, 40%, 50%.

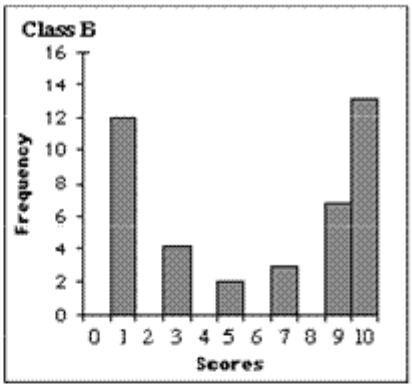
35%, 35%, 35%, 35%, 35%.

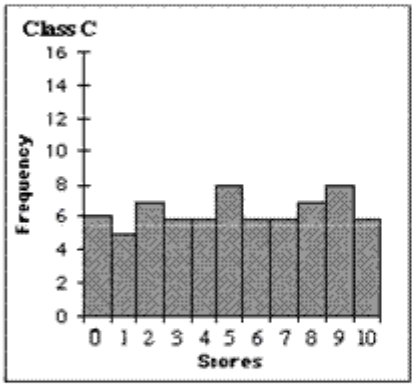
5%, 60%, 10%, 50%, 95%.

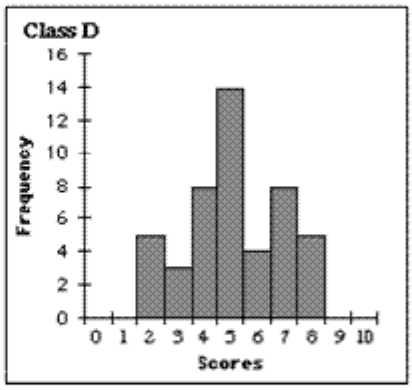
Any of the above.

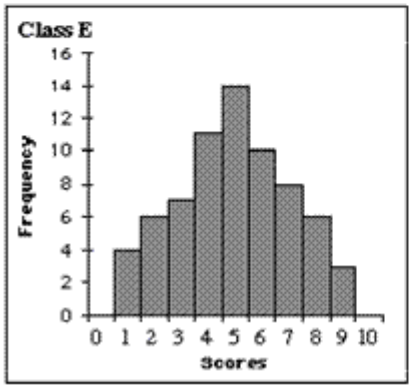
Five histograms are presented below. Each histogram displays test scores on a scale of 0 to 10 for one of five different statistics classes.











3c. Which of the classes would you expect to have the highest standard deviation, and why?

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

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

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

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

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

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, and 30 present four different interpretations of these results. Indicate if each interpretation is valid or invalid.

4a. We expect 95% of the cookies to have between 18.6 and 21.3 chocolate chips.

Valid.

Invalid.

4b. We would expect about 95% of all possible sample means from this population to be between 18.6 and 21.3 chocolate chips.

Valid.

Invalid.

4c. 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?

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.

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.

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.

5a. 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?

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

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.

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.

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

5b. 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?

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

The survey was sent to only first-year students.

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

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

5c. 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?

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

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.

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

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