AP Statistics Exam Review
Topic I: Describing Data

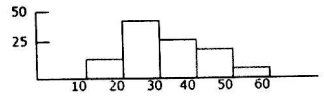
1. The following is a stem-and-leaf plot of a data set of size 80. What are the median and the mode of this set?

0	8
1	357
2	0468
3	467899
4	024577899
5	1334455677889
6	2345566778899
7	0112234456788
8	1123344567
9	22357
10	46

A. 66, 67
B. 64, 68
C. 65, 67
D. 65, 66
E. 64, 68

2. The following are math scores for a class of 30 students: 50, 60, 76, 77, 79, 76, 78, 80, 82, 84, 86, 90, 92, 99, 88, 86, 84, 70, 76, 78, 74, 72, 68, 66, 70, 70, 72, 74, 76, 78. The distribution of scores is:
A. symmetric.
B. skewed to the left.
C. uniform.
D. skewed to the right.
E. bell-shaped.

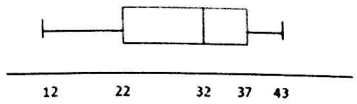
3. Which of the following statements about the data in this histogram is (are) true?



I. There are no outliers in the data.
II. There are no gaps in the data.
III. The mean is less than the median.
IV. The data set is symmetric.

A. I only
B. I and II only
C. I, II, and IV only
D. II and IV only
E. I, II, III, and IV

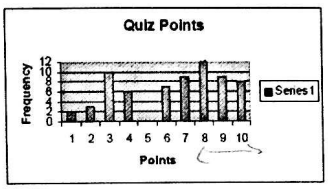
4. Consider this box plot of a set of scores. What is the median score? What is the shape of the distribution? On which side of the median is the mean?



A. 32, skewed left, right
B. 32, skewed right, left
C. 30, symmetric, right
D. 30, skewed right, right
E. 32, skewed left, left

5. Consider the following math quiz scores for a class of 30 students: 30, 40, 48, 72, 70, 76, 78, 80, 82, 84, 86, 90, 92, 100, 74, 88, 86, 84, 70, 76, 78, 72, 60, 55, 64, 68, 70, 74, 76, 78. What are the mean and median? Which measure would be a better indicator of how well you did on the quiz?
A. 73, 76, the mean
B. 74, 77, the median
C. 74, 74, either one
D. 73, 76, the median
E. 73, 88, the mean

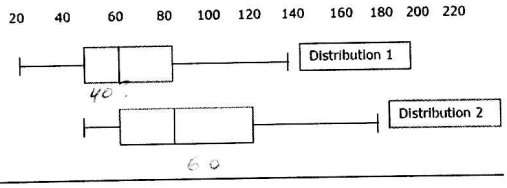
6. **What percent of scores are between eight and ten on the following graph?**



A. 66%
B. 61%
C. 50%
D. 44%
E. 41%

7. The average score on a test is 150 with a standard deviation of 15. Each score is then increased by 25. What are the new mean and standard deviation?
A. 175, 40
B. 280, 24
C. 175, 15
D. 200, 17
E. 3750, 375

8. Consider these boxplots for the two distributions. What is the interquartile range for each of the distributions?



A. Range for Distribution 1 = 15, Range for Distribution 2 = 25
B. Range for Distribution 1 = 45, Range for Distribution 2 = 40
C. Range for Distribution 1 = 120, Range for Distribution 2 = 130
D. Range for Distribution 1 = 50, Range for Distribution 2 = 70
E. Range for Distribution 1 = 40, Range for Distribution 2 = 60

9. Which statements represent the most accurate comparison of the two distributions of Question 8?
A. Both distributions are relatively symmetric with the median located near the center of the box. The whiskers are very similar for both distributions and show a large spread in the data.
B. Both distributions appear to be skewed to the right. There is more variation in the data for Distribution 2, because the interquartile range for Distribution 2 is larger than the interquartile range for Distribution 1.
C. Both distributions appear to be skewed to the right. There is more variation in the data for Distribution 1, because the interquartile range for Distribution 1 is larger than the interquartile range for Distribution 2.
D. Both distributions appear to be skewed to the right. The variation in both distributions is approximately the same.
E. None of the above statements represents an accurate description of the distributions.

10. A five-number summary for a set of data is (20, 25, 40, 51, 82). Which of the following could be an outlier?
A. 20
B. 25
C. 20 and 82
D. 82
E. There are no outliers.

FREE RESPONSE

Use the given data set of test grades from a college statistics class for this question.
85 72 64 65 98 78 75 76 82 80 61 92 72 58 65 74 92 85 74 76 77 77 62 68 68 54 62 76 73 85 88 91 99 82 80 74 76 77 70 60

A. **Construct two different graphs of these data**
B. **Calculate the five-number summary and the mean and standard deviation of the data.**
C. **Describe the distribution of the data, citing both the plots and the summary statistics found in questions 1 and 2.**

AP Statistics Exam Review
Topic II: Normal Distribution

1. Consider a normal distribution that has a mean of 78 and a standard deviation of 6. What is the percentile rank, to the nearest integer, of a value of 82 in this distribution?
A. 82%
B. 78%
C. 75%
D. 84%
E. 80%

2. Scores on a hypothetical IQ test are normally distributed with a mean of 110 and a standard deviation of 13. How high must your IQ be, to the nearest integer, to fall within the top 7% of the scores?
A. 125
B. 118
C. 140
D. 129
E. 131

3. Consider normally distributed data with a mean of 42 and a standard deviation of 2. If there are 450 values between 40 and 44, how many values are there in the distribution?
A. 594
B. 650
C. 465
D. 675
E. 660

4. The weights of adult men in a hypothetical population are approximately normally distributed with a mean of 168 lbs and a standard deviation of 18 lbs. Alvin is at the 92nd percentile in weight for adult men. What is his approximate weight, in pounds?
A. 168
B. 185
C. 235
D. 193
E. 180

5. Which of the following statements about the normal curve are true?
I. The mean and the median are the same and are located at the peak of the curve.
II. The area under the curve is exactly one, and it represents a relative cumulative frequency.
III. The curve is always on or above the horizontal axis.
IV. The shape of the curve is determined by the value of the standard deviation.

A. I and II only
B. I, II, and III only
C. II, III, and IV only
D. II and IV only
E. I, II, III, and IV

5. What is the standard deviation of a normal distribution with a mean of 27, given that an x value of 18 has a z score of -1.93?

A. 5.0
B. 5.7
C. 17.4
D. 4.7
E. 7.5

$-1.93 = \frac{18-27}{\sigma}$

$\sigma = 4.66$

7. What proportion of the terms in a normal distribution has z scores between -3.16 and -0.59?

A. 0.277
B. 0.262
C. 0.15
D. 0.45
E. 0.0277

.2768 area calc

8. A random number generator was used to generate the following sample of 30 scores: 62, 69, 77, 55, 70, 69, 66, 67, 59, 65, 59, 79, 72, 63, 71, 73, 74, 62, 71, 66, 79, 64, 84, 63, 60, 77, 82, 69, 72, 75. What is the approximate z score and percentile rank for the data point 77?

A. 1.50, 93%
B. 0.75, 77%
C. 1.00, 84%
D. 1.08, 86%
E. 1.25, 89%

$z = 1.08$ $\bar{x} = 69.13$ $\sigma = 7.27$

So $\frac{77-69.13}{7.27}$

9. Suppose that the air pollution index for your city averages 73.5 with a standard deviation of 16. Assuming this is a normally distributed data set, within what interval will the index fall 95% of the time?

A. (25.5, 121.5)
B. (57.5, 89.5)
C. (41.5, 105.5) closest
D. (49.5, 97.5)
E. None of the above

$\bar{x} = 73.5$ $\sigma = 16$

$73.5 \pm 1.96\left(\frac{16}{\sqrt{77}}\right)$

(69.15, 104.50)

$1.96 = \frac{x-73.5}{16}$ $-1.96 = \frac{x-73.5}{16}$ -31.36 31.36 $+73.5$

10. A hypothetical soda company produces soda in 12 oz cans. However, the actual volume of soda in the can varies normally with a mean of 11.9 oz and a standard deviation of 0.3 oz. What is the probability that the mean amount of soda in a can marked 12 oz is actually less than 12 oz?

A. 0.57
B. 0.61
C. 0.59
D. 0.63
E. 0.62

$P(x < 12) = \frac{12-11.9}{.3}$

$z = .333$ .629

FREE RESPONSE

A set of 2,000 measurements had a symmetric, mound-shaped distribution. The mean is 5.3 and the standard deviation is 0.7. Determine an interval that contains approximately 1,360 date values.

AP Statistics Exam Review
Topic III: Bivariate Data

1. What is the regression equation for the following set of data?

Game: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Points Scored: 73, 72, 83, 82, 79, 84, 85, 91, 86, 94

A. $y = 0.81x + 71.53$
B. $y = 71.53x + 2.07$
C. $y = 2.07x + 71.53$
D. $y = 0.81x + 2.07$
E. None of the above

$y = 71.53 + 2.06x$

2. Consider the regression equation obtained in Question 1. What is the best interpretation of the slope of the regression equation?

A. For each additional game played, the score is predicted to increase by just over two points.
B. For every two games played, the score is predicted to increase by one point.
C. For each additional game played, the score is predicted to increase by just under one point.
D. For every four games played, the score is predicted to increase by five points.
E. None of the above statements represents an accurate interpretation of the slope of the regression equation.

3. Which of the following statements about residuals are always true?

I. A random scattering of residuals means that the relationship is linear.
II. The sum of the residuals is always zero.
III. Residuals are plotted against the original x values.

A. I and II only
B. I and III only
C. II and III only
D. I, II, and III
E. None of the above

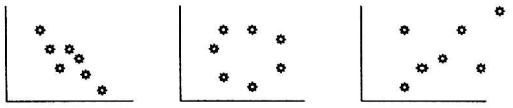
4. Which of the following statements is (are) true about the correlation coefficient r?

I. The square of the correlation coefficient, r^2, represents the proportion of t variation in the y variable that is explained by the x variable.
II. A correlation coefficient of -0.9 indicates a weak negative association between x and y.
III. A correlation coefficient of 0.5 means that 50% of the variation in the y values is explained by the x values.
IV. When $r = 1$, there is a perfect cause and effect relationship between the variables.

A. I only
B. I and II only
C. I, III, and IV only
D. III and IV only
E. I, II, III, and IV

$r = .8981$ $r^2 = .8065$

5. Which of the following is true about the correlations for the three scatterplots?



A. One is zero, and the other two are positive.
B. Two are zero, and the other is close to one.
C. Two are negative and the other is zero.
D. One is zero, one is negative, and one is positive.
E. None are less than or equal to 0.

6. Which of the following statements about influential scores is (are) true?

I. Influential points always have large residuals.
II. Removal of an influential point sharply affects the regression line.
III. A point with an x value that is an outlier can be a more influential point than a point with a y value that is an outlier.
IV. Removal of influential points affects the correlation coefficient.

A. I only
B. I, II, and III only
C. III and IV only
D. II, III, and IV only
E. I, II, III, and IV

7. A hypothetical survey was taken after an election to find out how voters felt about the manner in which it was handled. Seven hundred registered voters were asked their political affiliation and whether they felt that there were significant irregularities in the way the election was conducted.

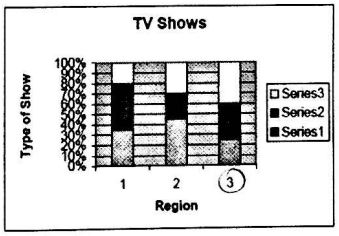
	Agree	Disagree	No Opinion
Democrat	200	65	25
Republican	75	175	50
Independent	75	25	10

290
300
110

Which group was least likely to have no opinion about the voting procedures, based on the table above?

A. Independents .091
B. Republicans .16
C. Democrats .086
D. Democrats and Independents equally
E. Democrats, Republicans, and Independents equally

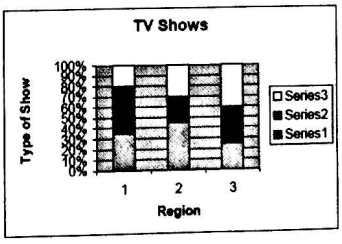
8. A hypothetical survey was taken regarding the types of TV shows that people like the best. Series 1 represents sit-coms, Series 2 represents talk shows, and Series 3 represents action dramas. Individuals from three regions of the country (1 represents the North, 2 represents the West, and 3 represents the South) were polled. What percentage of those surveyed in the South prefer talk shows?



60
23

A. 20%
B. 25%
C. 30%
D. 35%
E. 40%

9. A hypothetical survey was taken regarding the types of TV shows that people like the best. Series 1 represents sit-coms, Series 2 represents talk shows, and Series 3 represents action dramas. Individuals from three regions of the country (1 represents the North, 2 represents the West, and 3 represents the South) were polled. Based on the chart, does there seem to be a relationship between geographic location and TV show preference?



A. Yes, because all three bars are the same height.
B. Yes, because the corresponding segments of the three bars have different sizes.
C. Yes, because all three bars have three segments.
D. No, because the segment sizes in each bar are different.
E. No, because the sums of the corresponding segments of the three bars are different.

10. In the following table, what value for *n* results in a table showing perfect independence?

7	28	
18	*n*	

A. 36
B. 90
C. 144
D. 84
E. 72

FREE RESPONSE

A survey was conducted recently in ten large American cities to determine whether there is any relationship between the average weekday hotel rates and average car rental rates. The following data was collected.

Daily Hotel Rate (x) (in dollars)	Daily Car Rental Rate (y) (in dollars)
149	49
187	50
171	52
122	49
115	39
147	44
128	37
212	63
168	46
181	51

A. Construct a scatterplot for this data
B. Use the scatterplot to determine if there is a linear relationship between the two variables
C. If there is a linear relationship, numerically describe the strength of this relationship and construct a least squares regression model
D. Find the residual associated with the point (168, 46).
E. What percent of the variation in the car rental rates is explained by the regression of y on x?

AP Statistics Exam Review
Topic IV: Planning a Study

1. An observational study based on survey data concluded that fertilizer A leads to improved plant growth over fertilizer B. You want to replicate this study using an experimental approach and obtain 30 plants to work with. Which of the following is an acceptable design for your study?
A. Treat 20 plants with both fertilizers A and B and treat another 10 plants with a third type of fertilizer.
B. Treat 15 plants with fertilizer A and the other 15 plants with fertilizer B.
C. Assign the numbers 1 through 30 to the subject plants and use a random number table to determine which 15 plants will get fertilizer A and which 15 plants will get fertilizer B.
D. All of the above are acceptable designs.
E. None of the above is an acceptable design.

2. You're interested in determining how many students in your high school are planning on going to college right after high school. You pose the question to 50 students in your school. The population of interest is:
A. the chosen 50 students.
B. the senior class.
C. all high school students.
D. the students in your high school.
E. the teachers and students in your high school.

3. Which of the following statements is (are) true about a double-blind experiment?
I. In a double-blind experiment, neither the experimenters nor the subjects know which treatment a subject received.
II. A double-blind experiment eliminates bias.
III. In a double-blind experiment, one set of subjects is given all treatments.
A. I only
B. I and II
C. I and III
D. I, II, and III
E. None of these are true.

4. You are designing a study to examine the effect on cholesterol level of eating eggs versus egg derivatives for breakfast. Fifty women and fifty men with average cholesterol levels were selected to participate. After the treatment, you measure the cholesterol level of each subject. What would be the best design for this study?
A. Completely randomized (everyone chooses what they eat)
B. Comparative randomized (blocked by gender)
C. Randomized block (blocked by gender and type of treatment by randomly assigning half of the women and half of the men to the egg-eating group and the rest to the egg-derivative-eating group)
D. None of the above is an appropriate design.
E. All of the above are appropriate designs.

5. A toothpaste manufacturer conducted a set of experiments to examine the effects of a new anti-plaque compound. Subjects assigned to the control group were given toothpaste without the anti-plaque compound; subjects assigned to the treatment group were given the same type of toothpaste as the control group, but with the anti-plaque compound added. All subjects were given new toothbrushes. At the end of the experiment, those who were in the control group had slightly less plaque compared to those in the treatment group. This could be an example of which of the following?
A. Hawthorne effect
B. Placebo effect
C. Over-sampling
D. Interviewer bias
E. Voluntary bias

6. You are shopping for a new laptop computer and favor brand X, but decide to look at laptop reviews before buying. After reading that 62% of the people who reviewed the performance of the brand X laptop are very unhappy with it, you change your mind about buying brand X. This is an example of which of the following?
A. Voluntary response bias
B. Confounding
C. Hidden bias
D. Nonresponse bias
E. Interviewer response bias

7. Which of these survey questions could result in a biased response?
I. Do you agree or disagree with the following statement?
Because many student perpetrators in school shootings admit that their actions were inspired by the media, there is a direct relationship between violence in the media and youth violence.
II. Do you agree or disagree with the following statement?
Because there are only a few incidents of shootings in schools, there is no direct relationship between violence in the media and violence in our youth.
III. A questionnaire is sent to 10,000 teachers in a city asking them to express their preference for mayor, candidate A or candidate B. The questionnaire contains the following statement about candidate A: Candidate A supports merit increases for teachers based on how students perform on annual standardized tests, a position opposed by the teacher's union.
IV. Are you in favor of a three-day waiting period between the filing of an application for a marriage license and the receiving of the marriage license?
A. I and II.
B. I, II, and III.
C. I, II, and IV.
D. III and IV.
E. All of the above are biased.

8. Which of the following sampling designs is (are) least likely to result in bias?
I. Voluntary response sampling
II. Convenience sampling
III. Quota sampling
IV. Multistage cluster sampling
A. I, III, and IV equally
B. II, III, and IV equally
C. III and IV equally
D. III only
E. IV only

9. A hypothetical company is hired to conduct a nationwide telephone survey on whether people consider the movie-rating system effective. The company selects 100 area codes at random and then uses random digit dialers within each area code to contact participants. This is an example of:
A. simple random sampling.
B. convenience sampling.
C. voluntary response sampling.
D. multistage cluster sampling.
E. None of the above.

10. A census of all teachers in a certain school indicates that 42% of them require their students to use Internet-based resources. Which of the following is not true?
A. The population of interest is the teachers in this school.
B. Sampling design is not an issue in this situation.
C. This proportion is of limited value in inferring information about all teachers.
D. 42% is a statistic resulting from this census.
E. All of these statements are true.