AP Statistics
Mr Murphy
Sample Fall Final

Name: Key
Date:
Period:

Multiple Choice (1 pt. each)

1. The head of the admissions office at a small college wants to understand why minority students who visit her school do not eventually enroll. The college holds a preview weekend for students who have been admitted. Two months later, after the students have decided while college to attend, a survey is sent out to all minority students who attended the weekend visit but who did not choose to attend this college. About a third of them returned the survey, with 48% of those indicating that they received a larger scholarship offer elsewhere. Which is true?

I. The population of interest is all potential college students.
II. This survey design suffered from non-response bias.
III. Because it comes from a sample, 48% is a parameter, not a statistic.

(a) I only
(b) II only
(c) I and II only
(d) II and III only
(e) I, II, and III

2. Which of the following is the best description of a systematic random sample?

(a) A sample chosen in such a way that every possible sample of a given size has an equal chance to be the sample.

(b) After a population is separated into distinct groups, one or more of these groups are randomly selected in their entirety to be the sample.

(c) A value is randomly selected from an ordered list and then every nth value in the list after that first value is selected for the sample.

(d) Select a sample in such a way that the proportion of some variables thought to impact the response is approximately the same in the sample as in the population.

(e) A sample in which respondents volunteer their response.

3. Given P(A) = 0.4, P(B) = 0.3, P(B|A) = 0.2. What are P(A and B) and P(A or B)?

(a) P(A and B) = 0.12, P(A or B) = 0.58
(b) P(A and B) = 0.08, P(A or B) = 0.62
(c) P(A and B) = 0.12, P(A or B) = 0.62
(d) P(A and B) = 0.08, P(A or B) = 0.58
(e) P(A and B) = 0.08, P(A or B) = 0.70

Conditional Prob: $P(B|A) = \frac{P(B \text{ and } A)}{P(A)}$

$0.2 = \frac{P(B \text{ and } A)}{0.4}$

$P(B \text{ and } A) = 0.08$

$P(A \text{ and } B) = 0.08$

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

$= 0.4 + 0.3 - 0.08 = 0.62$

Independence here means $P(\text{Left} \mid F) = P(\text{Left}) = \frac{20}{120} = \frac{1}{6}$

4. A survey of some AP Stats students recorded gender and whether or not the student was left or right-handed. Results were summarized in the table below. If it turned out that handedness were independent of gender, how many of the AP Stats students were lefty girls?

	Lefty	Righty	Total
Male	11		66
Female	???9		54
Total	20	100	120

$\frac{1}{6}(54) = 9$

(a) 4
(b) 7
(c) 9
(d) 10
(e) The number cannot be determined.

5. A study is to be conducted on a new weatherproofing product for outdoor decks. Four houses with outdoor decks in one suburban neighborhood are selected for the study. Each deck is to be divided into two halves, one half receiving the new product and the other half receiving the product the company currently has on the market. Each of the four decks is divided into North/South sections. Either the new or the old product is randomly assigned to the North side of each of the decks and the other product is assigned to the South side. The major reason for doing this is that

(a) the study is much too small to avoid using randomization.

(b) there are only two treatments being studied.

(c) this controls for known differences in the effect of the sun on the North and South sides of decks.

(d) randomization is a necessary element of any experiment.

(e) this controls for the unknown differential effects of the weather on the North and South sides of decks in this neighborhood.

6. Which of the following best describes a cluster sample of size 20 from a population of size 320?

(a) All 320 names are written on slips of paper and the slips are put into a box. Twenty slips are selected at random from the box.

(b) The 320 names are put into an alphabetical list. One of the first 16 names on the list is selected at random as part of the sample. Every 16th name on the list is then selected for the sample.

(c) The sample will consist of the first 20 people who volunteer to be part of the sample.

(d) Each of the 320 people is assigned a number. Twenty numbers are randomly selected by a computer and the people corresponding to these 20 numbers are the sample.

(e) The 320 names are put into an alphabetical list and the list numbered from 1 to 320. A number between 1 and 320 (inclusive) is selected at random. The person corresponding to that number and the next 19 people on the list are selected for the sample.

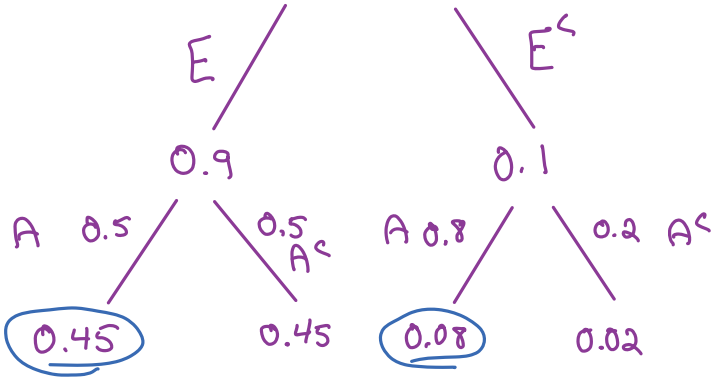
Questions 7 and 8 refer to the following information:

At a local community college, 90% of students take English. 80% of those who don't take English take art courses, while only 50% of those who do take English take art.

7. What is the probability that a student takes art?

(a) 0.80
(b) 0.53
(c) 0.50
(d) 1.3
(e) 0.45

$P(A) = 0.45 + 0.08$
$= 0.53$



8. What is the probability that a student who takes art doesn't take English?

(a) 0.08
(b) 0.10
(c) 0.8
(d) 0.85
(e) 0.15

$P(E^c|A) = \frac{P(E^c \cap A)}{P(A)} = \frac{(.1)(.8)}{(.53)}$

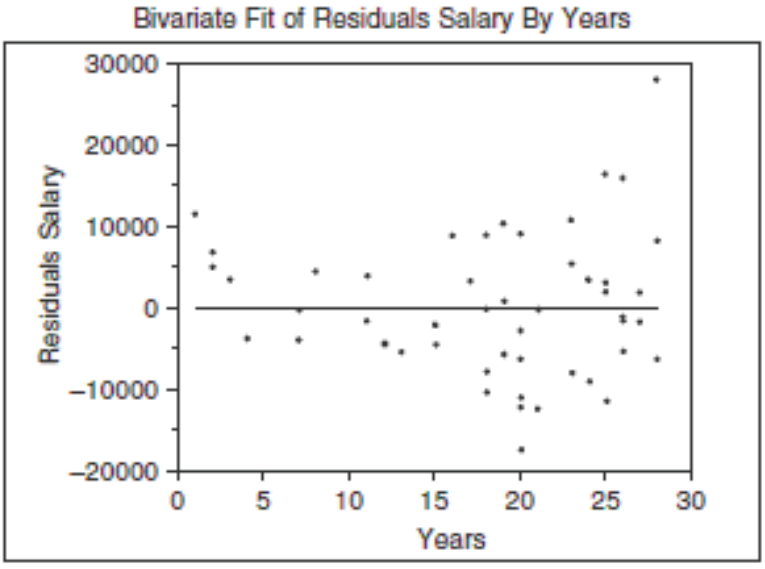
$P(E^c|A) = \frac{P(E^c \cap A)}{P(A)} = \frac{(0.1)(0.8)}{0.53} = 0.15$

9. The following numbers are given in ascending order: 3, 4, x, x, 9, w, 13, 28, y, z. Which of the following gives a five-number summary of the data?

(a) $\left\{3, x, \frac{w+9}{2}, 28, z\right\}$
(b) $\left\{3, \frac{x+4}{2}, \frac{w+9}{2}, 28, z\right\}$
(c) $\{3,4,9,13,28\}$
(d) $\{3,x,w,28,z\}$
(e) There isn't enough information to identify all five numbers in the five-number summary.

$\{\min Q_1, \text{Med}, Q_3, \max\}$

10. The salaries and years of experience for 50 social workers was collected and a regression analysis was conducted to investigate the nature of the relationship between the two variables. R-sq. = 0.79. The results are as follows:



Linear Fit

Salary = 11369.416 + 2141.3092*Years

Parameter Estimates

TERM	ESTIMATE	STD ERROR	t RATIO	PROB > |t|
Intercept	11369.416	3160.249	3.60	0.0008
Years	2141.3092	160.8358	13.31	<.0001

Which of the following statements is least correct?

(a) Starting salaries are about $11,370.
(b) The residual plot indicates that a line is a good model for the data for all years.
(c) There appears to be an outlier in the data at about 28 years of experience.
(d) The variability of salaries increases as years of experience increases.
(e) For each additional year of experience, salary is predicted to increase by about $2141.

because of the outlier mainly but this could still be true

11. A researcher was interested in determining the relationship between pulse rate (in beats/minute) and the time (in minutes) it took to swim a fixed distance. Based on 25 trials in the pool, the correlation coefficient between time and pulse rate was found to be –0.654 (that is, large times—going slowly—were associated with slower pulses). Prior to publication, the researcher decided to change the time measurements to seconds (each of the 25 times was multiplied by 60). What would this conversion do to the correlation between the two variables?

(a) Since the units on only one of the variables was changed, the correlation between the two variables would decrease.

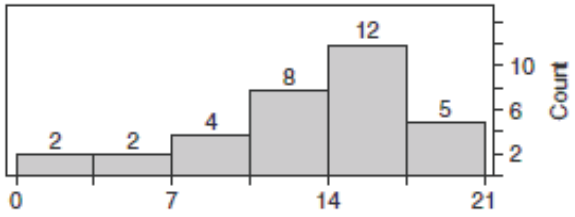
(b) The correlation would change proportional to the change in the units for time.

(c) The correlation between the two variables would change, but there is no way, based on the information given, to know by how much.

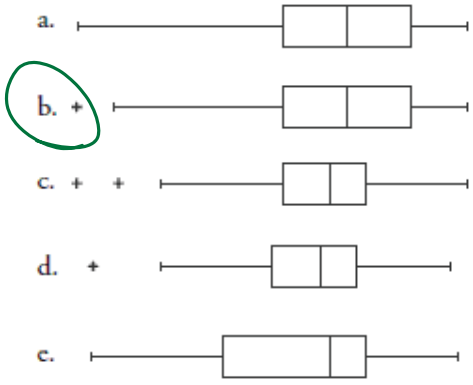
(d) Changing the units of measurement has no effect on the correlation coefficient. Hence, the correlation would be the same.

(e) Since changing from minutes to seconds would result in larger times, the correlation would actually increase.

12. The following histogram displays the scores of 33 students on a 20-point Introduction to Statistics quiz. The lowest score, 0, is an outlier. The next lowest score, 2, is not an outlier.



Which of the following boxplots best represents the data shown in the histogram?



13. Which of the following statements is (are) correct?

I. The area under a probability density curve for a continuous random variable is 1.

II. A random variable is a numerical outcome of a random event.

III. The sum of the probabilities for a discrete random variable is 1.

(a) II only

(b) I and II

(c) I and III

(d) II and III

(e) I, II, and III

14. Alligators captured in Florida are found to have a mean length of 2 meters and a standard deviation of 0.35 meters. The lengths of alligators are believed to be approximately normally distributed. What is the approximate length of an alligator at the 67th percentile of alligator lengths?

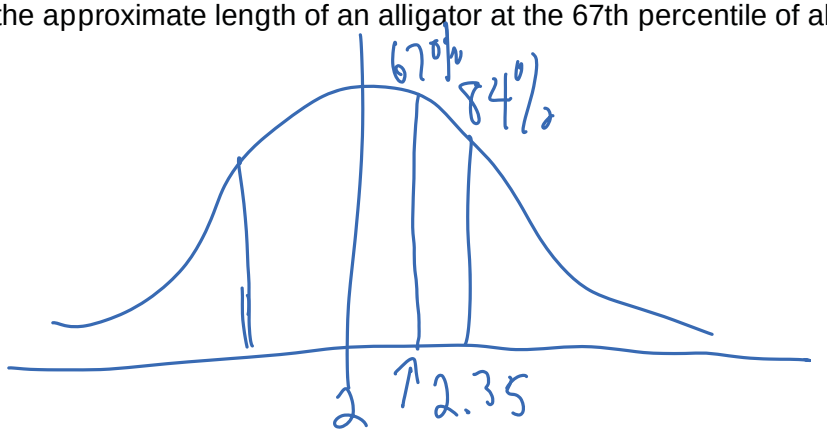
(a) 2.01 meters.

(b) 2.44 meters.

(c) 2.21 meters.

(d) 2.15 meters.

(e) 2.09 meters.



15. A psychiatrist is studying the effects of regular exercise on stress reduction. She identifies 40 people who exercise regularly and 40 who do not. Each of the 80 people is given a questionnaire designed to determine stress levels. None of the 80 people who participated in the study knew that they were part of a study. Which of the following statements is true?

(a) This is an observational study.

(b) This is a randomized comparative experiment.

(c) This is a double-blind study.

(d) This is a matched-pairs design.

(e) This is an experiment in which exercise level is a blocking variable.

16. A study published in the Journal of the National Cancer Institute (Feb. 16, 2000) looked at the association between cigar smoking and death from cancer. The data reported were as follows:

Death from Cancer			
	Yes	No	Total
Never Smoked	782	120,747	121,529
Former Smoker	91	7,757	7,848
Current Smoker	141	7,725	7,866
Total	1,014	136,229	137,243

Which of the following statements is true?

(a) A former smoker is more likely to have died from cancer than a person who has never smoked.

(b) Former smokers and current smokers are equally likely to have died from cancer.

(c) The events "Current Smoker Dies from Cancer" and "Died from Cancer" are independent.

(d) It is more likely that a person who is a current smoker dies from cancer than a person has never smoked and dies from cancer.

(e) Among those whose death was not from cancer, the proportion of current smokers is higher than the proportion of former smokers.

17. You play a game that involves rolling a die. You either win or lose $1 depending on what number comes up on the die. If the number is even, you lose $1, and if it is odd, you win $1. However, the die is weighted and has the following probability distribution for the various faces:

Face	1	2	3	4	5	6
Win (x)	+1	−1	+1	−1	+1	−1
$p(x)$	0.15	0.20	0.20	0.25	0.1	0.1

Given that you win rather than lose, what is the probability that you rolled a "5"? Prob of a 5 given a win

(a) 0.50

(b) 0.10

(c) 0.45

(d) 0.22

(e) 0.55

$$P(5 \mid \text{win}) = \frac{P(5 \cap \text{win})}{P(\text{win})} = \frac{0.1}{0.2 + 0.25 + 0.1}$$

all the different ways you can win

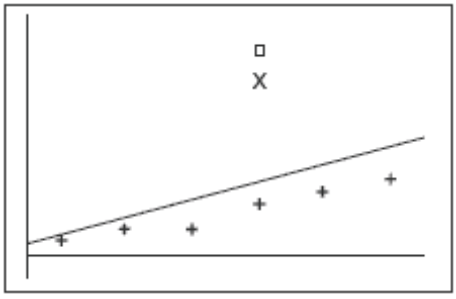
$$= 0.22$$

18.

For the graph given at right,

which of the following statements

is (are) true?



I. The point marked with the "X" is better described as an outlier than as an influential point.

II. Removing the point "X" would cause the correlation to increase.

III. Removing the point "X" would have a significant effect on the slope of the regression line.

(a) I and II only

(b) I only

(c) II only

(d) II and III only

(e) I, II, and III

19. 40% of the staff in a local school district have a master's degree. One of the schools in the district has only 4 teachers out of 15 with a master's degree. You are asked to design a simulation to determine the probability of getting this few teachers with master's degrees in a group this size. Which of the following assignments of the digits 0 through 9 would be appropriate for modeling this situation?

(a) Assign "0,1,2" as having a master's degree and "4,5,6,7,8,9" as not having a degree.

(b) Assign "1,2,3,4,5" as having a master's degree and "0,6,7,8,9" as not having a degree.

(c) Assign "0,1" as having a master's degree and "2,3,4,5,6,7,8,9" as not having a degree.

(d) Assign "0,1,2,3" as having a master's degree and "4,5,6,7,8,9" as not having a degree.

(e) Assign "7,8,9" as having a master's degree and "0,1,2,3,4,5,6," as not having a degree.

0,1,2,3 40%

4-9 60%

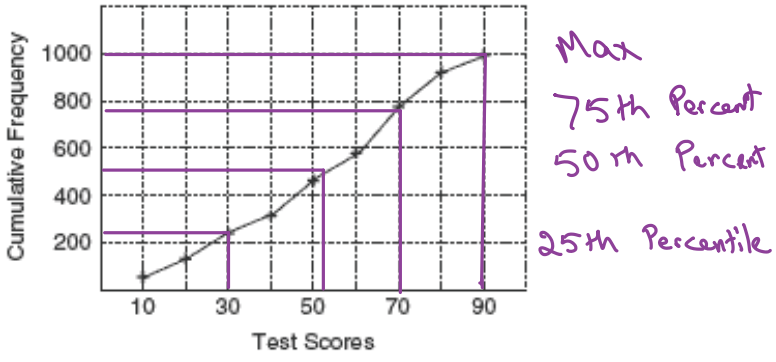
20. Given the cumulative frequency table shown below, what are the median of the distribution?

VALUE	CUMULATIVE FREQUENCY
2	0.15
3	0.25
5	0.45
7	0.95
10	1.00

← 45% up the chart means that 50% is higher

(a) 2
(b) 3
(c) 5
(d) 7
(e) 10

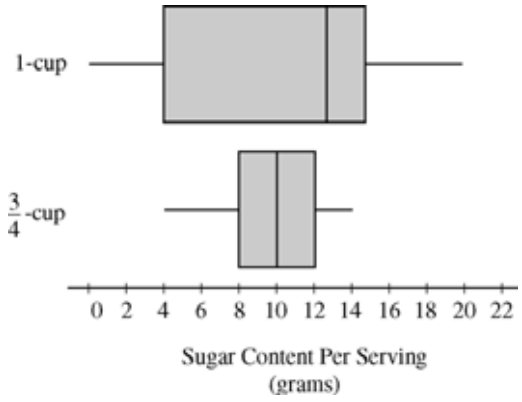
21. A spelling test was given to 1000 elementary students in a large urban school district. The graph below is a cumulative frequency graph of the results. Which of the following is closest to the five-number summary (minimum, first quartile, median, third quartile, maximum) for the distribution of spelling scores?



(a) {10, 30, 50, 70, 90}
(b) {10, 20, 50, 80, 90}
(c) {0, 30, 50, 70, 100}
(d) {20, 40, 60, 80, 100}
(e) There is not enough information contained in the graph to determine the five-number summary.

Free Response (4 pts.)

1. (2008 Q1) To determine the amount of sugar in a typical serving of breakfast cereal, a student randomly selected 60 boxes of different types of cereal from the shelves of a large grocery store. The student noticed that the side panels of some of the cereal boxes showed sugar content based on one-cup servings, while others showed sugar content based on three-quarter-cup servings. Many of the cereal boxes with side panels that showed three-quarter-cup servings were ones that appealed to young children, and the student wondered whether there might be some difference in the sugar content of the cereals that showed different-size servings on their side panels. To investigate the question, the data were separated into two groups. One group consisted of 29 cereals that showed one-cup serving sizes; the other group consisted of 31 cereals that showed three-quarter-cup serving sizes. The boxplots shown below display sugar content (in grams) per serving of the cereals for each of the two serving sizes.



(a) Write a few sentences to compare the distributions of sugar content per serving for the two serving sizes of cereals.

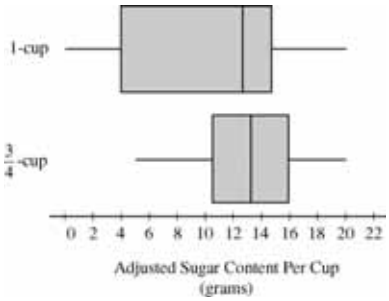
The one cup serving size cereals have a higher median value than the 3/4 cup cereals.

The 3/4 cup servings have a distribution that is mostly symmetric while

the one cup servings is clearly skewed left

The one cup servings also have more variability (IQR, Range).

After analyzing the boxplots on the preceding page, the student decided that instead of a comparison of sugar content per recommended serving, it might be more appropriate to compare sugar content for equal-size servings. To compare the amount of sugar in serving sizes of one cup each, the amount of sugar in each of the cereals showing three-quarter-cup servings on their side panels was multiplied by $4/3$. The bottom boxplot shown below displays sugar content (in grams) per cup for those cereals that showed a serving size of three-quarter-cup on their side panels.



(b) What new information about sugar content do the boxplots above provide?

Increasing the 3/4 cup size serving by a factor of 4/3 gives it a median that is now more similar to and even slightly larger than the 1 cup serving size

The variability of the 3/4 cup size servings, while larger now by a factor of 4/3, is still much less than the 1 cup serving

The distributions now have similar maximum values

(c) Based on the boxplots shown above on this page, how would you expect the mean amounts of sugar per cup to compare for the different recommended serving sizes? Explain.

Since the 3/4 cup servings distribution is mostly symmetric, we should expect the mean to be close to the median

Since the 1 cup servings distribution is skewed left, we should expect the mean to be smaller than the median so we expect that the 1 cup serving will have a lower mean than the 3/4 cup serving

2. (2006B Q5) When a tractor pulls a plow through an agricultural field, the energy needed to pull that plow is called the draft. The draft is affected by environmental conditions such as soil type, terrain, and moisture.

A study was conducted to determine whether a newly developed hitch would be able to reduce draft compared to the standard hitch. (A hitch is used to connect the plow to the tractor.) Two large plots of land were used in this study. It was randomly determined which plot was to be plowed using the standard hitch. As the tractor plowed that plot, a measurement device on the tractor automatically recorded the drat at 25 randomly selected points in the plot.

After the plot was plowed, the hitch was changed from the standard one to the new one, a process that take a substantial amount of time. Then the second plot was plowed using the new hitch. Twenty-five measurements of draft were also recorded at randomly selected points in this plot.

(a) What was the response variable? The amount of draft

Identify the treatments.

There were two treatments: The standard hitch and the new hitch

What were the experimental units?

The two plots of land

(b) Given that the goal of the study is to determine whether a newly developed hitch reduces draft compared to the standard hitch, was randomization properly used in this study? Justify your answer.

Yes because the two hitches were randomly assigned to the two plots

(c) Given that the goal of the study is to determine whether a newly developed hitch reduces draft compared to the standard hitch, was replication properly used in this study? Justify your answer.

There is no real replication used in this study because each type of hitch was only applied to one plot of land. It would have been better to have replicated and reassigned treatments such that both plots of land would have received both treatments for comparison

(d) Plot of land is a confounding variable in this experiment. Explain why.

Since both plots did not receive the same treatment, the plots became confounding variables as a result. This could have been addressed by using both hitches on both plots but since it was not we can't be sure whether the difference in draft is affected by the plot

3. (2011 Q5) Windmills generate electricity by transferring energy from wind to a turbine. A study was conducted to examine the relationship between wind velocity in miles per hour (mph) and electricity production in amperes for one particular windmill. For the windmill, measurements were taken on twenty-five randomly selected days, and the computer output for the regression analysis for predicting electricity production based on wind velocity is given below. The regression model assumptions were checked and determined to be reasonable over the interval of wind speeds represented in the data, which were from 10 miles per hour to 40 miles per hour.

Predictor	Coef	SE Coef	T	P
Constant	0.137	0.126	1.09	0.289
Wind velocity	0.240	0.019	12.63	0.000

S = 0.237 R-Sq = 0.873 R-Sq (adj) = 0.868

(a) Use the computer output above to determine the equation of the least squares regression line. Identify all variables used in the equation.

$\widehat{(\text{electricity production})} = 0.137 + 0.240(\text{wind velocity})$

$\hat{y}$ x

(b) How much more electricity would the windmill be expected to produce on a day when the wind velocity is 25 mph than on a day when the wind velocity is 15 mph? Show how you arrived at your answer.

A slope of 0.240 means that for each additional one mph of wind speed electricity production is predicted to increase by 0.24 amperers. Therefore, we expect an increase of $10(0.24) = 2.4$ amperers between a 15 mph day and a 25 mph day

(c) What proportion of the variation in electricity production is explained by its linear relationship with wind velocity?

This proportion is represented by r^2 so the proportion is 0.873

(d) Interpret the estimate of the slope parameter in context.

A slope of 0.240 means that for each additional one mph of wind speed electricity production is predicted to increase by 0.24 amperers.