Name _____

AP Statistics
Part 1 Review Test 1

1. A sample of 99 distances has a mean of 24 feet and a median of 24.5 feet. Unfortunately, it has just been discovered that an observation which was erroneously recorded as "30" actually had a value of "35." If we make this correction to the data, then
   a) the mean remains the same, but the median is increased.
   b) the mean and median remain the same.
   c) the median remains the same, but the mean is increased.
   d) the mean and median are both increased.
   e) we do not know how the mean and median are affected without further calculations, but the variance is increased.

2. A set of data has a median that is much larger than the mean. Which of the following statements is most consistent with this information?
   a) A stemplot of the data is assymetrical.
   b) A stemplot of the data is skewed left.
   c) A stemplot of the data is skewed right.
   d) The data set must be so large that it would be better to draw a histogram than a stemplot.
   e) A stemplot of the data is symmetric.

3. There are some children playing on the playground. Their mean age is four years old. If another four-year-old child joins the play group, the
   a) mean age will stay the same but the variance will increase.
   b) mean age will increase, but the variance will stay the same.
   c) mean age and variance will stay the same.
   d) mean age and variance will increase.
   e) mean age will stay the same but the variance will decrease.

4. The weights of the male and female students in a class are summarized in the following boxplots:
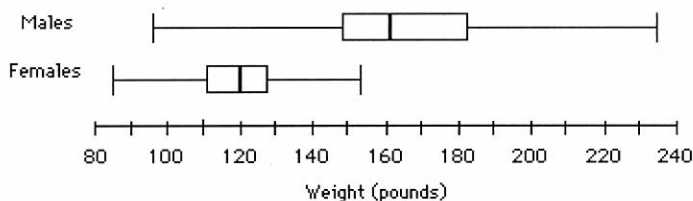
Which of the following is NOT correct?
   a) About 50% of the male students have weights between 150 and 185 pounds.
   b) About 25% of female students have weights more than 130 pounds.
   c) The median weight of male students is about 162 pounds.
   d) The mean weight of female students is about 120 pounds because of symmetry.
   e) The male students have less variability than the female students.

*The answer is E* (handwritten annotation)

5. A town contains three elementary schools. School A has a mean class size of 25 pupils for its three fifth-grade classrooms. School B has a mean class size of 30 pupils in its two fifth-grade classrooms. School C has 20 pupils in its only fifth-grade classroom. What is the average class size for fifth-grade classrooms in this town?
   a) 12.5
   b) 25
   c) 25.8
   d) 26.7
   e) Cannot be determined

total students: $25 \cdot 3 + 30 \cdot 2 + 20 = 155$

$\dfrac{155}{6} = 25.83$

6. The mean of a set of data is 35, its median is 33, its standard deviation is 6, and its IQR is 12. A new set is created by first multiplying every term by 5 and then subtracting 10. Which of the following is true?

  a) The standard deviation is 20.
  b) The sum of the standard deviation and inter quartile range is 18.
  c) The mean is 175.
  d) The inner quartile range is 50.
  e) The difference between the mean and the median is 10.

(handwritten:) SD = 30, IQR = 60, mean = 165, median = 155
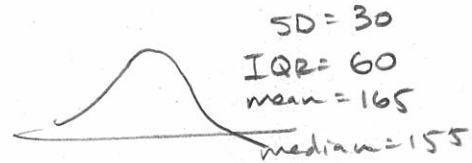
7. Consider the following data set: {14, 17, 37, 45, 51, 52, 53, 56, 56, 57, 60, 63, 65, 67, 89}. Which of the following are outliers?

  a) 89 only
  b) 14 and 89
  c) 14 only
  d) 14, 17, and 89
  e) 14 and 17

*(handwritten:)*
$Q_1 = 45$
$Q_3 = 63$
$IQR = 18$
$45 - 18(1.5) = 18$
$63 + 18(1.5) = 90$

8. Sally's percentile score on a math exam was 35 while Sara's percentile score on the same test was 70. We know that

  a) Sara correctly answered more items than Sally did
  b) Sara's math achievement is twice as good as Sally's
  c) Sally scored better than 35 of her classmates
  d) Sally correctly answered half as many items as Sara did
  e) they both scored better than average on the math exam

9.

mean = $ 2.90
median = $ 2.50
interquartile range = $ 9
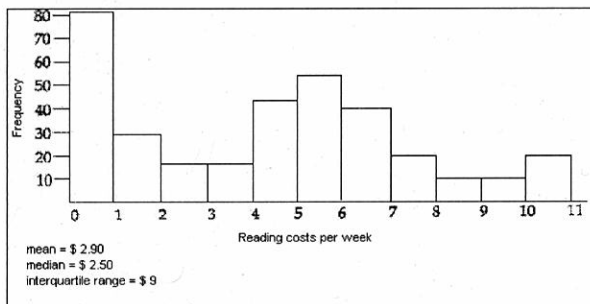
Based on the data and graph, what do you think of this statement: The mean is a poor indicator of average and the median should be used instead to represent the "average" amount of money spent on reading material.

  a) Agree; the median is better to use here because it's not sensitive to outliers.
  b) Agree; the median is better to use with nonsymmetric distributions like this one.
  c) Disagree; the median lacks accuracy because it is just the middle value.
  d) Disagree; the mean is better because it takes all the amounts of money into consideration and none are left out.
  e) Disagree; the mean is a more accurate measure.

10. Which of the following histograms pictured below has the smallest standard deviation?

  a)   b)   c)   d)   e)
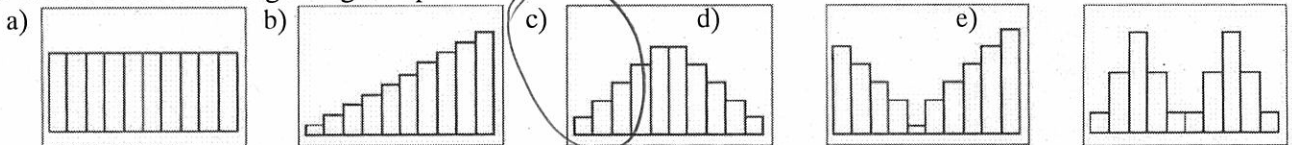
11.

Mary's best time for downhill skiing the challenging course has a z-score of 0.5 as compared to all skiers that are timed on the same course. Which statement best interprets her z-score?

(A) Mary's time is 0.5 seconds times faster than all skiers timed on the same course.

(B) Mary's time is 0.5 seconds faster than all skiers timed on the same course.

(C) Mary's time is 0.5 standard deviations below the mean for all skiers timed on the same course.

(D) Mary's time is 0.5 standard deviations above the mean time for all skiers timed on the same course.

(E) Mary skis worse than the majority of the skiers timed on the same course.

12.

	Job	No Job	Total
Juniors	13	5	18
Seniors	13	26	39
Total	26	31	57

A survey of 57 students was conducted to determine whether or not they held jobs outside of school. The two-way table above shows the numbers of students by employment status (job, no job) and class (juniors, seniors). Which of the following best describes the relationship between employment status and class?

(A) There appears to be no association, since the same number of juniors and seniors have jobs.
(B) There appears to be no association, since close to half of the students have jobs.
(C) There appears to be an association, since there are more seniors than juniors in the survey.
(D) There appears to be an association, since the proportion of juniors having jobs is much larger than the proportion of seniors having jobs.
(E) A measure of association cannot be determined from these data.

(May not be ready for) this question yet)

13.

At a college the scores on the chemistry final exam are approximately normally distributed, with a mean of 75 and a standard deviation of 12. The scores on the calculus final are also approximately normally distributed, with a mean of 80 and a standard deviation of 8. A student scored 81 on the chemistry final and 84 on the calculus final. Relative to the students in each respective class, in which subject did this student do better?

(A) The student did better in chemistry.
(B) The student did better in calculus.
(C) The student did equally well in each course.
(D) There is no basis for comparison, since the subjects are different from each other and are in different departments.
(E) There is not enough information for comparison, because the number of students in each class is not known.
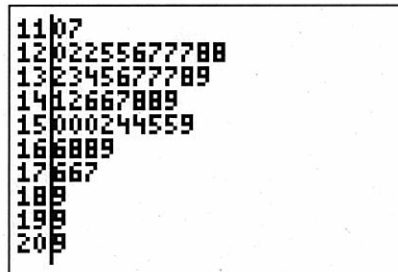
$$\frac{81-75}{12} = .5 \qquad \frac{84-80}{8} = .5$$

14. In order to rate TV shows, phone surveys are sometimes used. Such a survey might record several variables, some of which are listed below. Which of these variables are categorical?

      I.   The type of show being watched
      II.   The number of persons watching the show
      III.   The ages of persons watching the show
      IV.   The name of the show being watched
      V.   The number of times the show has been watched in the last month

(a) II, III, and V
(b) I only
(c) I and V
(d) I and IV
(e) None of the above describes the complete set of correct responses

```
11|07
12|02255677788
13|2345677789
14|12667889
15|000244559
16|6889
17|667
18|9
19|9
20|9
```

15. The stemplot displays the 1988 per capita income (in hundreds of dollars) of the 50 states. Which of the following best describes the data?

(a) Skewed distribution, mean greater than median
(b) Skewed distribution, median greater than mean
(c) Symmetric distribution, mean greater than median

(d) Symmetric distribution, median greater than mean
(e) Symmetric distribution with outliers on high end
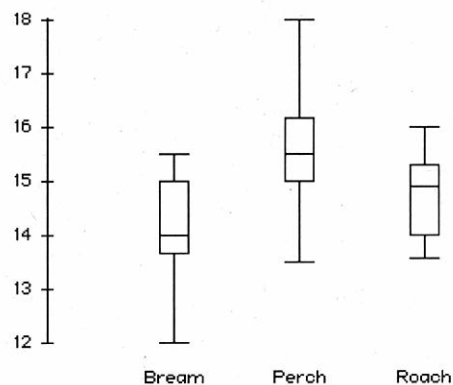
Bream   Perch   Roach

16. A study was conducted on the weights of three different species of fish (Bream, Perch & Roach) found in a lake in Finland. These three fish (bream, perch and roach) are commercial fish. Their weights are displayed in the boxplots. Which of the following statements comparing these boxplots is **NOT** correct?

(a) The median weights of the three species differ.
(b) The spread of roach is less than the spread of the other two species.
(c) The distributions of weights are approximately symmetric for all three species.
(d) There are no outliers in weight for the three species.
(e) The variability in the weights for the three species combined exceeds the variation in the medians of the three species.

17. The mean age of 14 of the members attending a mathematics department faculty meeting is 42. Mr. Myers, who is 57, arrives late. What is the average of all 15 members?

(a) 43
(b) 44
(c) 45
(d) 46
(e) cannot be determined

$$total = 14 \cdot 42 + 57 = 645$$

$$\frac{645}{15} = 43$$

18. The weights of cockroaches living in a typical college dormitory are approximately normally distributed with a mean of 80 grams and a standard deviation of 4 grams. The percentage of cockroaches weighing between 77 grams and 83 grams is about:

(a) 99.7%
(b) 95%
(c) 68%
(d) 55%
(e) 34%

$$\frac{77-80}{4} = -.75$$

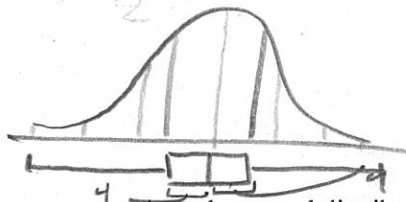$$\frac{83-80}{4} = .75$$

.5467

19. Scores on the American College Test (ACT) are normally distributed with a mean of 18 and a standard deviation of 6. The interquartile range of the scores is approximately:
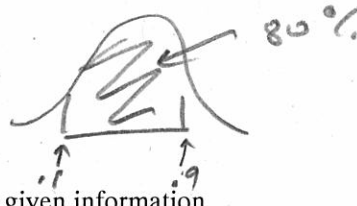
.25 → z score of -.67

(a) 8.1
(b) 12
(c) 6
(d) 10.3
(e) 7

20. The test grades at a large school have an approximately normal distribution with a mean of 50. What is the standard deviation of the data so that 80% of the students are within 12 points (above or below) the mean?

(a) 5.875
(b) 9.375
(c) 10.375
(d) 14.5
(e) cannot be determined from the given information
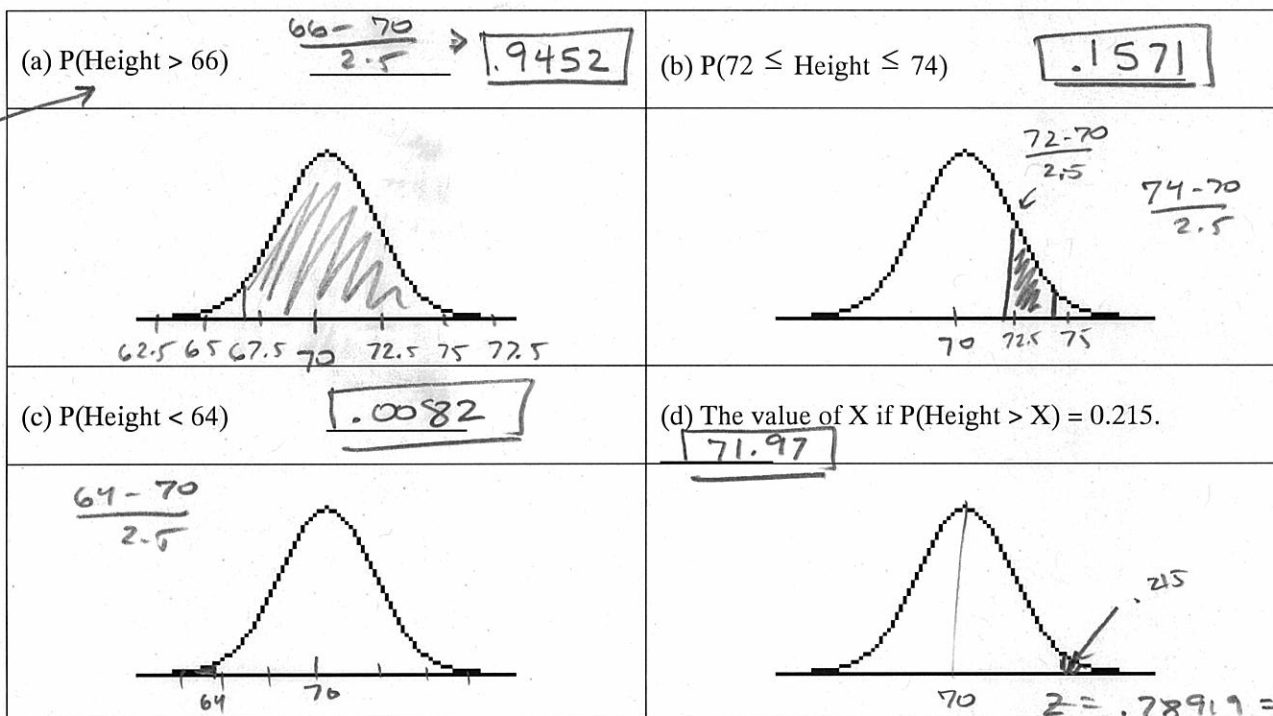

80%

$$1.28 = \frac{62-50}{\sigma}$$

$$\frac{1.28}{1} = \frac{12}{\sigma}$$

$$\sigma = 9.36$$

1. The heights of NCAA women basketball players are approximately normally distributed with $\mu = 70$ and $\sigma = 2.5$. For each of the following, illustrate with a picture and evaluate.

This notation may be confusing for you. P

(a) P(Height > 66) $\dfrac{66-70}{2.5} \rightarrow \boxed{.9452}$	(b) P(72 ≤ Height ≤ 74) $\boxed{.1571}$
62.5 65 67.5 70 72.5 75 77.5	$\dfrac{72-70}{2.5}$ $\dfrac{74-70}{2.5}$ 70 72.5 75
(c) P(Height < 64) $\boxed{.0082}$	(d) The value of X if P(Height > X) = 0.215. $\boxed{71.97}$
$\dfrac{64-70}{2.5}$ 64 70	.215 70 $z = .78919 = \dfrac{X-70}{2.5}$

2. The summary statistics for the number of inches of rainfall in Los Angeles for 177 years, beginning in 1877, are shown below.

N	Mean	Median	StDev	Min	Max	Q1	Q3
117	14.941	13.070	6.747	4.850	38.180	9.680	19.250

(a) Describe a procedure that uses these summary statistics to determine whether there are outliers.

oops

$IQR = Q3 - Q1 = 19.250 - 9.680 = 9.57$

$Q1 - 1.5(IQR)$
$9.680 - 1.5(9.57) = -4.675 < min (4.850)$
$Q3 + 1.5(IQR) = \boxed{33.605 < max (38.18)}$

There are outlier(s)

(b) Are there outliers in these data? Justify your answer based on the procedure that you described in (a).

Yes. Since the maximum is greater than Q3 + 1.5 · IQR there will be outliers.

(c) The news media reported that in a particular year, there were **only** 10 inches of rainfall. Use the information provided to comment on this reported statement.

$z = \dfrac{10 - 14.941}{6.747} = .23$

There is a 23% chance of there being "only" 10 inches of rainfall.

3. Two parents have each built a toy catapult for use in a game at an elementary school fair. To play the game, the students will attempt to launch Ping-Pong balls from the catapults so that the balls land within a 5-centimeter band. A target line will be drawn through the middle of the band, as shown in the figure below. All points on the target line are equidistant from the launching location.
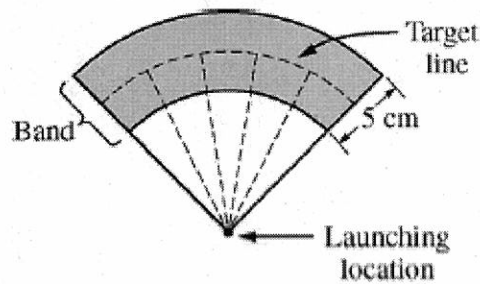
If a ball lands within the shaded band, the student will win a prize.

The parents have constructed the two catapults according to slightly different plans. They want to test these catapults before building additional ones. Under identical conditions, the parents launch 40 Ping-Pong balls from each catapult and measure the distance that the ball travels before landing. Distances to the nearest centimeter are graphed in the dotplot below.
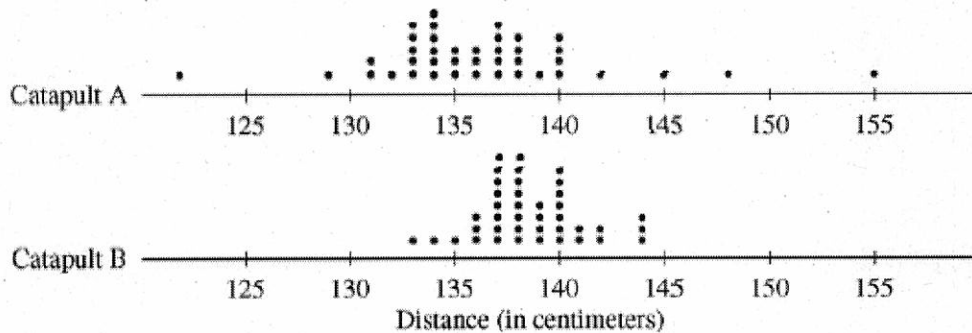
a) Comment on any similarities and any differences in the two distributions of distances traveled by balls launched from catapult A and catapult B.

b) If the parents want to maximize the probability of having the Ping-Pong balls land within the band, which one of the catapults, A or B, would be better to use than the other? Justify your choice.

c) Using the catapult that you chose in part (b), how many centimeters from the target line should this catapult be placed? Explain why you chose this distance.

AP® STATISTICS
2006 SCORING GUIDELINES

Question 1

Intent of Question

The primary goals of this question are: (1) to assess a student's ability to use simple graphical displays (dotplots in this case) to compare and contrast two distributions; and (2) to evaluate a student's ability to recognize what statistical information is most useful in making different practical decisions.

Solution

Part (a):

Both distributions of distances are roughly symmetric and somewhat mound-shaped. The center of the distances for catapult A (median A = 136 cm) is slightly lower than the center of the distances for catapult B (median B = 138 cm). There is more variability in the distances traveled by the Ping-Pong balls launched with catapult A. There are distances that are extreme enough to be called (potential) outliers in the catapult A distribution, but there are no outliers among the catapult B distances.

Part (b):

Catapult B would be best because the distances vary less about the center of the distribution for catapult B. If catapult B is properly placed, the balls launched will have a higher probability of landing in the narrow (only 5 cm wide) target band.

Part (c):

The catapult should be placed 138 cm from the target line. Since the distribution of distances for catapult B seems to be fairly symmetric and somewhat mound-shaped, the median (138 cm) is a good representation of the center of the distribution. Placing catapult B at this location would have resulted in a high proportion (30/40 = 0.75) of Ping-Pong balls from this sample of launches landing in the target band.

Scoring

Parts (a), (b), and (c) are scored as essentially correct (E), partially correct (P), or incorrect (I).

Part (a) is essentially correct (E) if the student correctly identifies similarities and differences in center, spread, and shape for the two distributions.

Part (a) is partially correct (P) if the student correctly identifies similarities and differences in two of the three characteristics (center, shape, and spread) for the two distributions.

Part (a) is incorrect (I) if the student correctly identifies no more than one similarity or difference of the three characteristics (center, shape, and spread) for the two distributions.

Notes:
- Correct comments regarding outliers should be viewed as a positive. However, comments about outliers do not count as one of the three required characteristics.

- Describing catapult A's distribution as "normal" or "skewed left" or "uniform" is not acceptable for the shape characteristic. Describing either distribution as "approximately normal" is acceptable.

- Giving separate lists of measures of center and/or spread for the two distributions with no linkage between them is not an acceptable discussion of similarities and differences for these characteristics.

Part (b) is essentially correct (E) if catapult B is chosen using a rationale based on the variability in the distances.

Part (b) is partially correct (P) if catapult B is chosen, but the explanation does not refer to the variability in the distances.

Part (b) is incorrect (I) if catapult B is chosen and no explanation is provided OR catapult A is chosen.

Part (c) is essentially correct (E) if:
the catapult is placed at the median (or mean) of the distances traveled by the Ping-Pong balls, and the explanation addresses why the median (or mean) was selected based on a property of the chosen statistic that relates to the context of the problem;
    OR
the catapult is placed at a distance of 137.5-139.5 cm from the target line, and the explanation indicates that the chosen distance resulted in a high proportion of the balls in the sample landing in the target band.

Part (c) is partially correct (P) if the catapult is placed at an acceptable distance from the target line, but the explanation is incomplete or incorrect.

Part (c) is incorrect (I) if the catapult is placed less than 137.5 centimeters or more than 139.5 centimeters from the target line.

Notes:
- Simply saying "because it's the median (or mean)" is an incomplete explanation.
- Some students may confuse the 5 cm band as meaning 5 cm on either side of the target line. If the student chooses the median (or mean) and satisfactorily addresses why the median (or mean) was selected OR chooses a value of 137-140 cm and the explanation indicates that the chosen distance resulted in a high proportion of the balls in the sample landing in the target band, score the response as partially correct.
- If a student gives the distance from the catapult to the front or back of the shaded band rather than the distance to the target line, but gives an otherwise correct response, score part (c) as partially correct.
- If a student picks catapult A in part (b) and follows through correctly in part (c), then part (c) should be scored as essentially correct.

4   **Complete Response**

         All three parts essentially correct

3   **Substantial Response**

         Two parts essentially correct and one part partially correct

2   **Developing Response**

         Two parts essentially correct and no parts partially correct

OR

         One part essentially correct and two parts partially correct

OR

         Three parts partially correct

1   **Minimal Response**

         One part essentially correct and either zero or one part partially correct

OR

         No parts essentially correct and two parts partially correct

Review Questions for Topic 3—Gathering Data

Multiple Choice

1. The Women's Health Study randomly assigned nearly 40,000 women over the age of 45 to receive either aspirin or a placebo for over 10 years to examine the effect of aspirin on cancer risk to healthy women. This long-term trial was best conducted as
   (A) a census.
   (B) an observational study.
   (C) a randomized comparative experiment.
   (D) a single-blind randomized comparative experiment.
   (E) a double-blind randomized comparative experiment.

2. In the previous problem, suppose that the researchers suspected that women over the age of 55 may respond differently to the treatment. Given that a random sample of 40,000 women over the age of 45 has already been chosen, the study would have been improved by
   (A) a stratified sample, with strata determined by age.
   (B) a stratified sample, with strata determined by gender.
   (C) a block design, with blocks determined by age.
   (D) a block design, with blocks determined by gender.
   (E) a double-blind completely randomized design.

3. Researchers were interested to know whether internal vehicle temperatures vary by outside temperatures. To evaluate this, temperature rise was measured continuously over a 60-minute period in a dark sedan on 16 different clear, sunny days with outside temperatures ranging from 72°F to 96°F. The researchers' method of analysis is best described as
   (A) a census.
   (B) a survey.
   (C) an observational study.
   (D) a randomized comparative experiment.
   (E) a single-blind randomized comparative experiment.

4. Respondents to a randomly distributed questionnaire answered the question, "Do you agree that nuclear weapons should never be used because they are immoral?" The study that uses the results of this questionnaire will *most* likely suffer from which type(s) of bias?
   (A) undercoverage
   (B) voluntary response
   (C) response
   (D) nonresponse
   (E) all of the above

5. A statistics teacher decides to compare this year's students to all those she has taught over the history of the course. She will use the students' examination performances as the method of comparison. Which of the following is true in this context?
   (A) The mean performance of this year's students is a parameter.
   (B) The mean performance of this year's students is a statistic.
   (C) The mean of all except this year's students is a parameter.
   (D) The mean of all students is a statistic.
   (E) None of these is true.

Use the information below for questions 6–8.

Left-handedness is the preference for the left hand over the right for everyday activities such as writing. Studies indicate that about 93% of the population may be characterized as right-handed. Researchers would like to conduct a study to determine if left-handed teens differ significantly in their ability to memorize facts than right-handed teens.

6. The explanatory variable in this study is
   (A) ability to memorize facts.
   (B) age.
   (C) handedness.
   (D) researchers.
   (E) none of the above.

7. A random-number table is used to simulate the number of teens selected before a left-handed teen is found. A proper assignment of digits is
   (A) let the digits 0–6 represent a left-handed teen and 7–99 represent a right-handed teen; ignore repeats of numbers 7–99 until you get a number from 0–6.
   (B) let the digits 0–6 represent a left-handed teen and 7–99 represent a right-handed teen; count repeats of numbers 7–99 and continue until you get a number from 0–6.
   (C) let the digits 00–06 represent a left-handed teen and 07–99 represent a right-handed teen; ignore repeats of numbers 07–99 until you get a number from 00–06.
   (D) let the digits 00–06 represent a left-handed teen and 07–99 represent a right-handed teen; count repeats of numbers 07–99 and continue until you get a number from 00–06.

(E) let the digits 01–07 represent a left-handed teen and 08–100 represent a right-handed teen; count repeats of numbers 08–100 and continue until you get a number from 01–07.

8. Suppose the researchers concluded that the average number of words memorized by left-handed teens was statistically significantly higher than the average number memorized by right-handed teens. In this context, statistically significant means that
   (A) the number of words memorized by left-handed students exceeded the number memorized by right-handed students.
   (B) the average number of words memorized by left-handed students exceeded the average number memorized by right-handed students.
   (C) right-handed students tend to not do as well as left-handed students in memorizing words.
   (D) it would be unlikely to observe an average difference as large as was observed by chance variation.
   (E) a randomized, controlled experiment was conducted.

9. Pollsters are interested in conducting a survey to determine if students who belonged to an online social network during college continued with it after graduation. They plan to randomly select 50 people from a list of registered voters. A problem with the proposed plan is that the
   (A) sampling frame may be different from the population of interest.
   (B) sample size is too small.
   (C) sample is a systematic random sample.
   (D) sample was not stratified by political party.
   (E) sampling design should incorporate multistage sampling.

10. The purpose of randomization in an experiment is to
   (A) make the experiment seem fair.
   (B) make the treatment groups as similar as possible.
   (C) make the treatment groups as different as possible.
   (D) create the blocks in an experiment.
   (E) reduce variability within treatment groups.

Free Response

1. A political party's position on a ballot is often determined by random selection. For the last 10 years, the Innovators have never had the coveted first position on the ballot; this position has always gone to either the Old-timers or Preservationists. The Innovators have cried foul. Use the random-number table below to test the likelihood of the disputed result. Clearly communicate your scheme, conduct 10 trials, and report your results.

32813	90372	59627	94240	12957
11832	26220	79684	53312	26114

Probability Review 2

LICENSE PLATES For the given configuration, determine how many different license plates are possible if (a) digits and letters can be repeated, and (b) digits and letters cannot be repeated.

(13.) 4 letters followed by 2 digits

$$26^4 \cdot 10^2$$

$$45697600$$

PERMUTATIONS WITH REPETITION Find the number of distinguishable permutations of the letters in the word.

47. GRAVEL   **48.** PANAMA

$$\frac{6!}{} = 720 \qquad \frac{6!}{3!} = 120$$

(65.) **SCHOOL CLUBS** A Spanish club is electing a president, vice president, and secretary. The club has 9 members who are eligible for these offices. How many different ways can the 3 offices be filled?

$$_9P_3 = 504$$

CARD HANDS Find the number of possible 5-card hands that contain the cards specified. The cards are taken from a standard 52-card deck.

15. 1 ace and 4 cards that are not aces   **16.** 5 hearts or 5 diamonds

$$_{54}C_1 \cdot _{48}C_4 = 778320 \qquad 2 \cdot \frac{13}{52} \qquad 2574$$

39. **STUDENT COUNCIL** Five representatives from a senior class of 280 students are to be chosen for the student council. In how many ways can students be chosen to represent the senior class on the student council?

$$_{280}C_5 = 1.38 \times 10^{10}$$

DISJOINT EVENTS Events A and B are disjoint. Find $P(A$ or $B)$.

4. $P(A) = 0.55, P(B) = 0.2$

$$.55 + .2 = .75$$

OVERLAPPING EVENTS Find the indicated probability.

10. $P(A) = 0.6, P(B) = 0.2$
$P(A$ or $B) = 0.7$
$P(A$ and $B) = $ ___?___

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$
$$.7 = .6 + .2 - x$$
$$\boxed{.1}$$

FINDING PROBABILITIES Find the indicated probability. State whether A and B are disjoint events.

29. $P(A) = 0.6$
$P(B) = 0.32$
$P(A \text{ or } B) = $ __?__
$P(A \text{ and } B) = 0.25$

.6 + .32 − .25

.92 − .25

.67

8 2
−2 5
6 7

43. **BIOLOGY** You are performing an experiment to determine how well plants grow under different light sources. Out of the 30 plants in the experiment, 12 receive visible light, 15 receive ultraviolet light, and 6 receive both visible and ultraviolet light. What is the probability that a plant in the experiment receives either visible light or ultraviolet light?

$12 + 15 − 6 = \frac{21}{30}$

9 in the dark?

49. **PET STORE** A pet store has 8 black Labrador retriever puppies (5 females and 3 males) and 12 yellow Labrador retriever puppies (4 females and 8 males). You randomly choose one of the Labrador retriever puppies. What is the probability that it is a female or a yellow Labrador retriever?

8 12

$\frac{9 + 8}{20} = \frac{17}{20} = .85$

INDEPENDENT EVENTS Events A and B are independent. Find the indicated probability.

4. $P(A) = 0.3$
$P(B) = 0.4$
$P(A \text{ and } B) = $ __?__

.12

7. $P(A) = $ __?__
$P(B) = 0.8$
$P(A \text{ and } B) = 0.6$

$.6 = .8 \cdot x$

$x = .75$

DEPENDENT EVENTS Events A and B are dependent. Find the indicated probability.

17. $P(A) = 0.7$
$P(B|A) = 0.5$
$P(A \text{ and } B) = $ __?__

20. $P(A) = $ __?__
$P(B|A) = 0.4$
$P(A \text{ and } B) = 0.2$

$P(B) = P(\ )$

$P(A \text{ and } B) = P(A)(P(B|A))$
$= .7 \cdot .5 = .35$

$.2 = x (.4)$

.5

40. **ACCIDENT REENACTMENT** You are a juror for a trial involving a nighttime car accident in a certain city. Use the tree diagram and the facts below to determine the probability that the car involved in the accident was blue.

- The make of the car is known. Of the cars in the city matching this make, 85% are green and 15% are blue.

- A witness of the accident identified the car as blue.

- In reenactments of the accident, the witness correctly reported the color of the car 80% of the time.

$\frac{.12}{.29} = .41$