

Directions: Circle the correct answer for each *multiple choice* question. You may write on this test for your scratch work.

1. John recently scored 113 on a particular standardized achievement test. The scores on the test are distributed with a mean of 100 and a standard deviation of 10. His cousin, Brandon, took a different standardized test and scored 263. The scores on Brandon's test have a mean of 250 and a standard deviation of 25. Which student did relatively better on his particular test?
- (A) John did better on his test.
(B) Brandon did better on his test.
(C) They both performed equally as well on their respective tests.
(D) It is impossible to tell since they did not take the same test.
(E) It is impossible to tell since the number of students taking the test is unknown.
2. Who makes more mistakes on their income tax forms: accountants or taxpayers who prepare the forms themselves? A random sample of income tax forms that were prepared by accountants was drawn from IRS records. An equal number of forms that were self-prepared by taxpayers was also drawn. The average number of errors per form was compared to determine if one group tends to make more mistakes than the other. What type of study is this?
- (A) Census
(B) Experiment
(C) Voluntary response survey
(D) Observational study
(E) Matched-pairs study
3. The baggage handling services of On-Time Airlines is interested in how many baggage handlers they need on duty at various times of the day to ensure passengers do not wait an unreasonable amount of time for their baggage. An airport executive performed a study and found there is a correlation between the number of passengers arriving at given times and the number of baggage handlers needed. She sampled various times during the day and different days of the week including weekend. She recorded the number of passengers arriving within any one-hour time block. Computer output from the regression equation analysis is shown below.

Predicted Baggage Handlers = $2.86 + 0.00408$ (number of passengers)

Predictor	Coef	StDev	T	P
Constant	2.860	1.324	2.16	0.083
Passengers	0.004081	0.001168	3.49	0.017

S = 1.562

R-sq = 70.9%

R-Sq(adj) = 65.1%

What is the value of the correlation coefficient for number of baggage handlers and number of arriving passengers?

- (A) -0.842 (B) 0.651 (C) 0.709 (D) 0.842 (E) 1.562

4. A dance club holds a raffle at the end of each dance. Five dancers are selected at random to each draw one numbered tag from a hat without replacement. There are 50 tags in the hat numbered from 1 to 50. Drawing a tag from 1 through 5 wins \$20, tags from 6 through 25 wins \$10, and tags from 26 through 50 wins \$5. In order to determine the average amount of money paid out, a simulation will be conducted using a random number table. Which of the following assignments of random numbers to tag values is most appropriate for the simulation?
- (A) Using single-digit numbers, assign 0 to represent a \$20 prize, 1–4 to represent a \$10 prize, and 5–9 to represent a \$5 prize.
- (B) Using single-digit numbers, assign 0 to represent a \$20 prize, 1 to represent a \$10 prize, and 2 to represent a \$5 prize. Numbers 3–9 are ignored.
- (C) Using two-digit numbers, assign 20 to represent a \$20 prize, 10 to represent a \$10 prize, and 05 to represent a \$5 prize. Numbers 00–04, 06–09, 11–19, 21–99 are ignored.
- (D) Using two-digit numbers, assign 01–05 to represent a \$20 prize, 06–25 to represent a \$10 prize, and 26–50 to represent a \$5 prize. Numbers 51–99 and 00 are ignored.
- (E) Using two-digit numbers, assign 01–10 to represent a \$20 prize, 11–40 to represent a \$10 prize, and 41–99 and 00 to represent a \$5 prize.
5. The student council wants to survey students at the school to see what brands of soda pop they want in the school machines. They randomly sampled 30 Freshmen, 30 Sophomores, 30 Juniors, and 30 Seniors. The sampling method they used is a
- (A) simple random sample.
- (B) stratified random sample.
- (C) cluster sample.
- (D) systematic random sample.
- (E) convenience sample.
6. Which of the following is a legitimate probability distribution?

x	3	4	5	6
$P(x)$	0.1	0.1	0.1	0.1

(A)

x	0	1	2	3	4	5
$P(x)$	0.2	0.1	0.1	0.2	0.2	0.3

(B)

x	-5	0	5	10	15
$P(x)$	0.6	0.1	0.2	0.1	0

(C)

x	1	3	5	7
$P(x)$	0.3	0.2	0.2	0.4

(D)

x	1	1.5	2	2.5	3	3.5	4	4.5
$P(x)$	0.05	0.08	0.07	0.20	0.30	0.10	0.21	-0.01

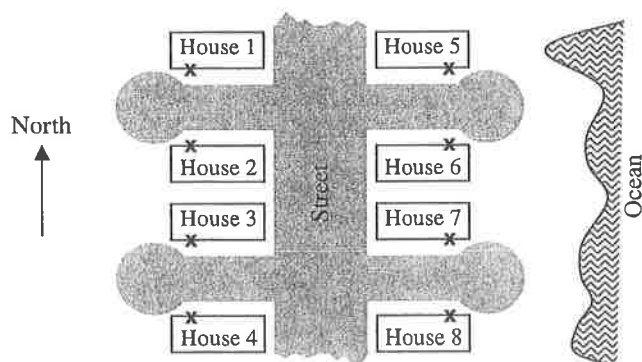
(E)

7. Some A.P. Statistics students were interested in finding out if there was a relationship between the number of hours of study for a chapter test and the score on that test. Based on the number of hours their classmates studied for the chapter 3 test and the scores on the test (out of 100%), the least-squares regression line calculated was $\hat{y} = 72.53 + 5.88x$, where x is the number of hours studied and \hat{y} is the predicted score on the test. Which statement correctly interprets the meaning of the slope of this regression line?
- (A) For each additional hour studied, the predicted score on the test increases by approximately 73 percent.
 - (B) For each additional hour studies, the predicted score on the test increases by approximately 6 percent.
 - (C) For each additional percent of increase on the test, the predicted score on the test increases approximately 73 percent.
 - (D) For each additional percent of increase on the test, the predicted score on the test increases approximately 6 percent.
 - (E) We cannot use this regression equation, since cause-effect has not been proven.
9. A baseball coach wants to compare the number of hits by two groups of batters each using a different type of bat. Which type of graphical display would NOT be appropriate?
- (A) Parallel boxplots
 - (B) Dotplots drawn on the same scale
 - (C) Back-to-back stemplots
 - (D) Histograms drawn on the same scale
 - (E) Scatterplot

10. A distribution of scores has a mean of 60 and a standard deviation of 18. If each score is doubled, then 5 is subtracted from that result, what will be the mean and standard deviation of the new scores?
- (A) mean = 115, standard deviation = 31
 - (B) mean = 115, standard deviation = 36
 - (C) mean = 120, standard deviation = 6
 - (D) mean = 120, standard deviation = 31
 - (E) mean = 120, standard deviation = 36
11. A new medication has been developed to cure a certain disease. The disease progresses in three stages, stages I, II, and III, each progressively worse than the one before it. Ninety volunteers are gathered to test the new medication, thirty in each of the three stages of the disease. The medication will be administered to subjects daily in one of three dosages: 100 mg for each subject in stage I of the disease, 200 mg to each subject in stage II, and 400 mg to each subject in stage III. After eight weeks, the proportion of subjects cured of the disease will be recorded. Why is this not a good experimental design?
- I. Because experiments of this type should only use one dosage level of medication.
 - II. Because disease stage is potentially confounded with dosage level.
 - III. Because the experiment lacks a control group.
- (A) I only
 - (B) II only
 - (C) I and II only
 - (D) II and III only
 - (E) I, II, and III
12. A certain variety of table grapes has fruit diameters that are distributed normally with mean 13 mm and standard deviation 2 mm. Approximately what proportion of grapes have diameters between 12 mm and 16 mm?
- (A) 0.134
 - (B) 0.378
 - (C) 0.500
 - (D) 0.625
 - (E) 0.683

14. In a recent high school basketball tournament where over 750 games were played, the mean team score was 68 points and the standard deviation was 13 points. The scores were approximately normally distributed. A coach was overheard saying that his team scored 95 points in one game. About what proportion of teams' scores during the tournament were more than 95 points?
- (A) 0.0035
 - (B) 0.0190
 - (C) 0.05
 - (D) 0.9810
 - (E) 2.07
15. The traffic safety officer of a local police force was trying to see if there was an association between the number of cars that did not use a main intersection in town because of the traffic light and the number of tickets written for speeding on the alternate route. The correlation between these two variables was found to be 0.58. Which of the following statements is true?
- (A) About 58% of the variation in the number of speeding tickets can be explained by the linear relationship between the number of speeding tickets issued and the number of cars that did not use the main intersection in town.
 - (B) Any potential linear relationship between the number of cars not using the main intersection in town and the number of speeding tickets written on an alternate route would be positive.
 - (C) If one uses the main intersection through town, one is 58% more likely to receive a ticket than using the alternate route.
 - (D) Since the correlation is not close to 1, there cannot be a linear relationship between the number of cars not using the main intersection in town and the number of speeding tickets written on an alternate route.
 - (E) Getting a speeding ticket is a direct cause of taking the alternate route.
16. A major automobile manufacturer is trying to improve its customer service at its dealerships across the United States. A survey of 200 customers in Arizona who recently purchased a vehicle from this manufacturer were asked if they were satisfied with the customer service at the dealership. Is it reasonable to generalize the conclusion to the population of all customers in the U.S. that purchased from this manufacturer?
- (A) No, because customers were only sampled in one state.
 - (B) No, because 200 is not a large enough sample.
 - (C) No, because only one sample was taken.
 - (D) Yes, because the sample size is more than 30.
 - (E) Yes, because the population of all new vehicle owners by this manufacturer is more than 2000.

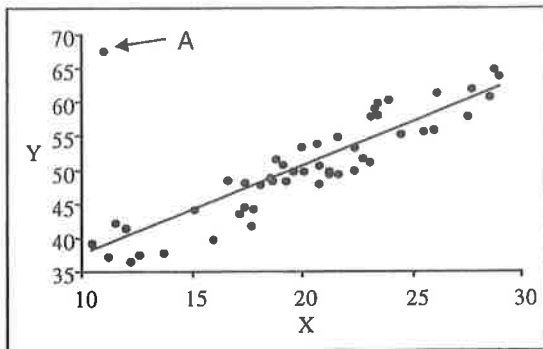
17. A politician is considering running for public office. He wants to measure his name recognition by doing a survey of voters in his district. Which of the following survey methods would produce unbiased results?
- (A) The politician stands in front of a grocery store in his district on Saturday morning and asking each person entering the store if he recognizes the politician's name.
 - (B) Placing pollsters in front of every grocery store in his district on Saturday morning and asking each person entering the store if he recognizes the politician's name.
 - (C) Sending a survey card to all registered voters in the district asking them to call a phone number to state whether or not they recognize the politician's name.
 - (D) Calling people from his district listed in the phone book and asking each if he recognizes the politician's name.
 - (E) None of these methods would produce unbiased results.
18. A garage door manufacturer has developed a new type of door for houses in the Southeast part of the United States. Doors in this area of the country are particularly susceptible to damage from salty ocean spray and the sun's rays, which tend to shine mainly on the south side of the house. An experiment will test the new type of garage door against the existing type of door on 8 houses in a particular residential area. An overhead view of the area is shown below. The location of the garage door on each home is marked with an "X."



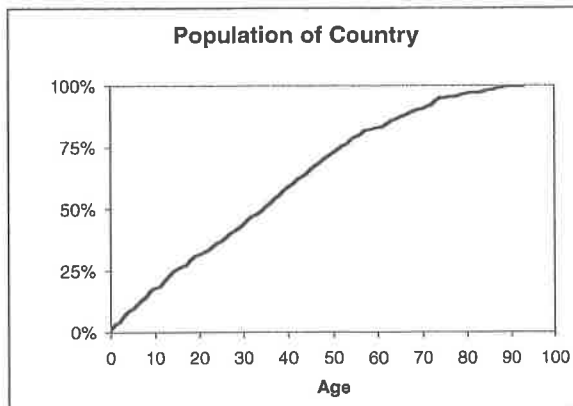
Which of the following blocking schemes is most appropriate to account for variables in this study other than the type of door?

- (A) Form the houses into two blocks: {1, 2, 3, 4} and {5, 6, 7, 8}.
- (B) Form the houses into two blocks: {1, 3, 5, 7} and {2, 4, 6, 8}.
- (C) Form the houses into four blocks: {1, 5}, {2, 6}, {3, 7}, and {4, 8}.
- (D) Form the houses into four blocks: {1, 3}, {2, 4}, {5, 7}, and {6, 8}.
- (E) No blocking is necessary in this experiment.

19. What is the approximate residual of the data point "A" on the scatterplot with least squares regression line shown below?



- (A) 11
(B) 29
(C) 39
(D) 58
(E) 68
20. The distribution of population by age in a particular country is represented by the given cumulative relative frequency plot.



Which of the following statements about the population is true?

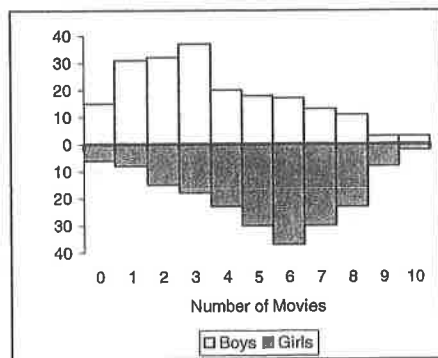
- (A) The median age is about 46 years.
(B) The interquartile range of ages is about 27 years.
(C) The mean age is greater than the median age.
(D) There are more people older than 65 years of age than there are people older than 30 years of age.
(E) Seventy-five percent of people are older than 50 years of age.

23. A statistics class randomly surveyed two hundred boys and two hundred girls at their school and asked each respondent how many movies they saw in theaters over summer break. The results are shown in the frequency table below. (Example: 15 boys saw no movies.)

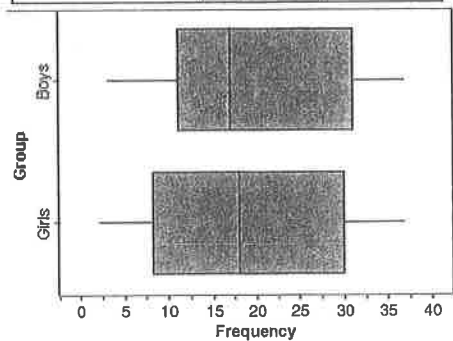
Number of Movies	Frequency of Boys	Frequency of Girls
0	15	6
1	31	8
2	32	15
3	37	18
4	20	23
5	18	30
6	17	37
7	13	30
8	11	23
9	3	8
10	3	2

Which of the following graphs is appropriate to compare the number of movies seen by boys and girls over summer break?

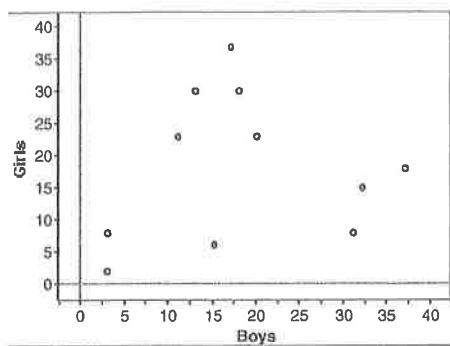
(A)



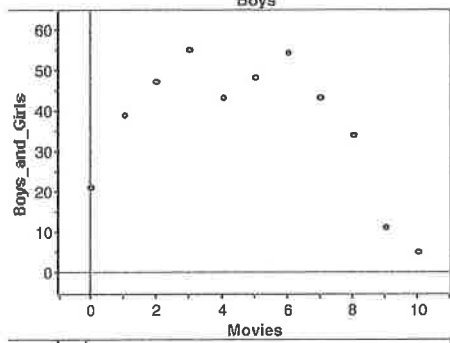
(B)



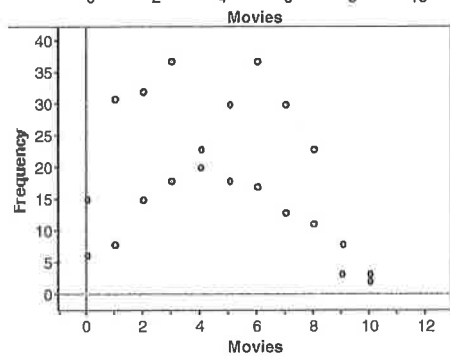
(C)



(D)



(E)



Free Response Section

Directions: Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy of your results and explanation.

1. Male and Female heights (in cm) are given for two high school basketball teams.

Male	169	170	170	173	174	175	178	180	183	185	185	191	195
Female	163	168	172	173	175	175	178	180	180	183	187	188	

- a) Construct parallel box-and-whisker plots of the data below.
- b) Construct back-to-back stem and leaf plots of the data.
- c) Compare the distribution of heights of male and female basketball players.
- d) What information is more readily available to you using the box-and-whisker plot that is not with the stem-and-leaf plot?
- e) What information is more readily available to you using the stem-and-leaf plot that is not with the box-and-whisker plot?

2. Die A has four 9's and two 0's on its faces. Die B has four 3's and two 11's on its faces. When either of these dice is rolled, each face has an equal chance of landing on top. Two players are going to play a game. The first player selects a die and rolls it. The second player rolls the remaining die. The winner is the player whose die has the higher number on top.

a) Suppose you are the first player and you want to win the game. Which die would you select? Justify your answer.

b) Suppose the player using Die A receives 45 tokens each time he or she wins the game. How many tokens must the player using Die B receive each time he or she wins in order for this to be a fair game? Explain how you found your answer.

(A fair game is one in which the player using die A and the player using die B both end up with the same number of tokens in the long run).

3. A dentist has just returned from a dental convention where a new formula of toothpaste with additional fluoride was announced. The maker of the toothpaste claims a reduction in cavities over the existing formula. Before recommending this new toothpaste to her patients, the dentist decides to conduct an experiment of her own. She asks 80 patients to volunteer to be part of the experiment. The experiment will last for one year.

a) Design the experiment that will allow the dentist to determine if the new formula reduces cavities as compared to the old formula.

b) Explain "blinding" in the context of this situation and why it is important for the experiment.

4. Airlines routinely overbook flights because they expect a certain number of no-shows. An airline runs a 5 P.M. commuter flight from Washington, D.C. to New York City on a plane that holds 38 passengers. Past experience has shown that if 41 tickets are sold for the flight, then the probability distribution for the number who actually show up for the flight is as shown in the table below.

Number who actually show up	36	37	38	39	40	41
Probability	0.46	0.30	0.16	0.05	0.02	0.01

Assume that 41 tickets are sold for each flight.

- a) There are 38 passenger seats on the flight. What is the probability that all passengers who show up for this flight will get a seat? Show work or explain your solution.

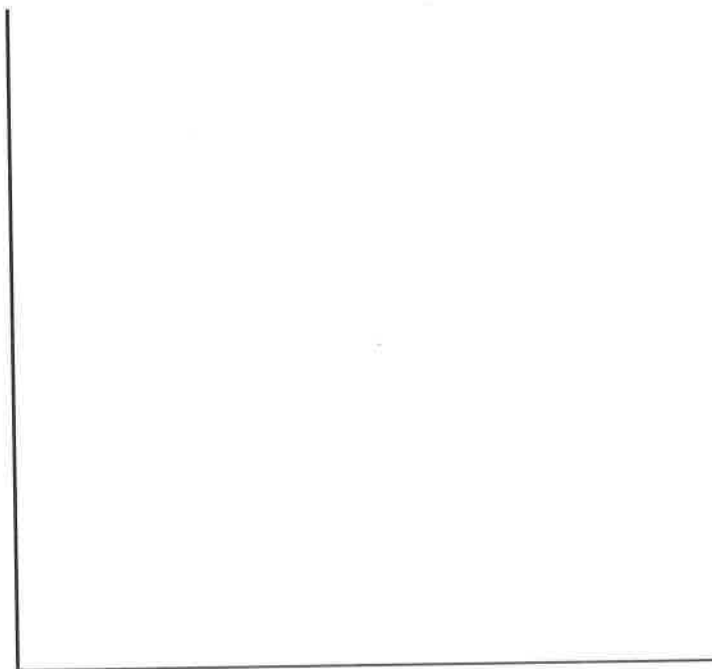
- b) What is the expected number of no-shows for this flight? Show work.

- c) Given that not all passenger seats are filled on a flight, what is the probability that only 36 passengers showed up for the flight? Show work.

5. Animal-waste lagoons and spray fields near aquatic environments may significantly degrade water quality and endanger health. The National Atmospheric Deposition Program has monitored the atmospheric ammonia at swine farms since 1978. The data on the swine population size (in thousands) and atmospheric ammonia (in parts per million) for one decade are given below.

Year	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
Swine Population	0.38	0.50	0.60	0.75	0.95	1.20	1.40	1.65	1.80	1.85
Atmospheric Ammonia	0.13	0.21	0.29	0.22	0.19	0.26	0.36	0.37	0.33	0.38

- (a) Construct a scatterplot for these data.



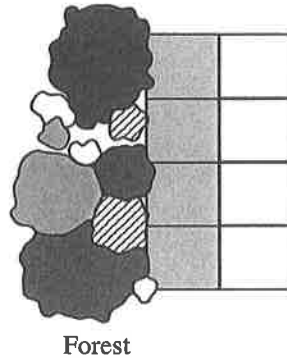
- (b) The value for the correlation coefficient for these data is 0.85. Interpret this value.
- (c) Based on the scatterplot in part (a) and the value of the correlation coefficient in part (b), does it appear that the amount of atmospheric ammonia is linearly related to the swine population size?
- Explain.
- (d) What percent of the variability in atmospheric ammonia can be explained by swine population size?

2001 AP® STATISTICS FREE-RESPONSE QUESTIONS

4. Students are designing an experiment to compare the productivity of two varieties of dwarf fruit trees. The site for the experiment is a field that is bordered by a densely forested area on the west (left) side. The field has been divided into eight plots of approximately the same area. The students have decided that the test plots should be blocked. Four trees, two of each of the two varieties, will be assigned at random to the four plots within each block, with one tree planted in each plot.

The two blocking schemes shown below are under consideration. For each scheme, one block is indicated by the white region and the other block is indicated by the gray region in the figures.

Blocking Scheme A



Forest

Key

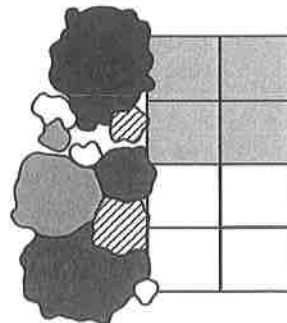


Block 1



Block 2

Blocking Scheme B



Forest

- (a) Which of the blocking schemes, A or B, is better for this experiment? Explain your answer.
- (b) Even though the students have decided to block, they must randomly assign the varieties of trees to the plots within each block. What is the purpose of this randomization in the context of this experiment?