**AP Statistics Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Midterm Exam Review**

**Topic I. Sampling and Experimentation: Planning and Conducting a Study**

1. Define these terms:

|  |  |
| --- | --- |
| Census  Population  Simple Random Sample (SRS)  Bias in a sample  Undercoverage bias  Non-response bias  Response bias  Confounding  Stratified random sample | Cluster Sample  Convenience Sample  Voluntary response sample  Systematic Sample  Completely Randomized design  Block design  Matched Pairs design  Experiment  Observational study |

1. For each of the situations described, state whether the sampling procedure is simple random sampling, stratified random sampling, cluster sampling, systematic sampling, or convenience sampling.
   1. All first-year students at a university are enrolled in 1 of 30 sections of a seminar course. To select a sample of freshmen at this university, a researcher selects four sections of the seminar course at random from the 30 sections and all students in the four selected sections are included in the sample.
   2. To obtain a sample of students, faculty, and staff at a university, a researcher randomly selects 50 faculty members from a list of faculty, 100 students from a list of students, and 30 staff members from a list of staff.
   3. A university researcher obtains a sample of students at his university by using the 85 students enrolled in his Psychology 101 class.
   4. To obtain a sample of the seniors at a particular high school, a researcher writes the name of each senior on a slip of paper, places the slips in a box and mixes them, and then selects 10 slips. The students whose names are on the selected slips of paper are included in the sample.
   5. To obtain a sample of those attending a basketball game, a researcher selects the 24th person through the door. Then, every 50th person after that is also included in the sample.
2. Can the sample described in part (b) above result in a simple random sample? Explain why or why not.
3. The Ministry of Health in the Canadian Province of Ontario wants to know whether the national health care system is achieving its goals in the province. Much information about health care comes from patient records but that source doesn’t allow us to compare people who use health services with those who don’t. So the Ministry of Health conducted the Ontario Health Survey, which interviewed a random sample of 61,239 people who live in the Province of Ontario.
   1. What is the population for this sample survey? What is the sample?
   2. The survey found the 76% of males and 86% of females in the sample had visited a general practitioner at least once in the past year. Do you think these estimates are close to the truth about the entire population? Why or why not?
   3. Is this an experiment or an observational study? How can you tell?
4. What are the characteristics of a well-designed and well-conducted study?
5. Can aspirin help prevent heart attacks? The Physicians’ Health Study, a large medical experiment involving 22,000 male physicians, attempted to answer this question. One group of about 11,000 physicians took an aspirin every second day, while the rest took a placebo. After several years the study found that subjects in the aspirin group had significantly fewer heart attacks than the subjects in the placebo group.
   * + - 1. Identify the experimental units, the factor(s), the levels, and the response variable in the health study.
         2. Describe a completely randomized design for the health study.
6. A mortgage lender routinely places advertisements in a local newspaper. The advertisements are of three different types: one focusing on low interest rates, one featuring low fees for first-time buyers, and one appealing to people who may want to refinance their homes. The lender would like to determine which advertisement format is most successful in attracting customers to call for more information.
   1. Describe an experiment that would provide the information needed to make this determination. Be sure to consider extraneous factors such as the day of the week that the advertisement appears in the paper, the section of the paper in which the advertisement appear, daily fluctuations of the interest rate and so forth.
   2. What role does randomization play in your design?

**Multiple Choice**

1. You want to take an SRS of 50 of the 816 students who live in a dormitory on a college campus. You label the students 001 to 816 in alphabetical order. In the table of random digits, you read the entries

95592 94007 69769 33547 72450 16632 81194 14873

The first three students in your sample have labels

* 1. 955, 929, 400
  2. 400, 769, 769
  3. 559, 294, 007
  4. 929, 400, 769
  5. 400, 769, 335

1. You are planning an experiment to determine the effect of the brand of gasoline and the weight of a car on gas mileage measured in miles per gallon. You will use a single test car, adding weights so that its total weight is 3000, 3500, or 4000 pounds. The car will drive on a test track at each weight using each of Amoco, Marathon, and Speedway gasoline. Which is the best way to organize the study?
   1. Start with 3000 pounds and Amoco and run the car on the test track. Then do 3500 and 4000 pounds. Change to Marathon and go through the three weights in order. Then change to Speedway and do the three weights in order once more.
   2. Start with 3000 pounds and Amoco and run the car on the test track. Then change to Marathon and then to Speedway without changing the weight. Then add weights to get 3500 pounds and go through the three gasoline brands in the same order. Then change to 4000 pounds and do the three gasoline brands in order again.
   3. Choose a gasoline at random, and run the car with this gasoline at 3000, 3500, and 4000 pounds in order. Choose one of the two remaining gasoline brands at random and again run the car at 3000, then 3500, then 4000 pounds. Do the same with the last gasoline.
   4. There are nine combinations of weight and gasoline. Run the car several times using each of these combinations. Make all these runs in random order.
   5. Randomly select an amount of weight and brand of gasoline, and run the car on the test track. Repeat this process a total of 30 times.
2. You work for an advertising agency that is preparing a new television commercial to appeal to women. You have been asked to design an experiment to compare the effectiveness of three versions of the commercial. Each subject will be shown one of the three versions and then asked about her attitude toward the product. You think there may be large differences between women who are employed and those who are not. Because of these differences, you should use
   1. A block design, but not a matched pairs design
   2. A completely randomized design
   3. A matched pairs design
   4. A simple random sample
   5. A stratified random sample
3. In order to estimate the proportion of students at a college who spend more than 2 hours per day on Facebook, a random sample of students at the college is selected and each student is interviewed about his or her use of Facebook. The students conducting the survey are worried that people who spend a lot of time on Facebook might be embarrassed to admit it and that their responses to the survey might not be honest. What type of bias are the students conducting the survey worried about?
   1. Undercoverage bias
   2. Nonresponse bias
   3. Response bias
   4. Bias due to confounding
   5. They shouldn’t worry – there is no obvious source of bias
4. A researcher wants to compare the effect of a new type of shampoo on hair condition. The researcher believes that men and women may react to the shampoo differently. Additionally, the researcher believes that the shampoo will react differently on hair that is dyed. The subjects are split into four groups: men who dye their hair; men who do not dye their hair; women who dye their hair; women who do not dye their hair. Subjects in each group are randomly assigned to the new shampoo or the old shampoo. This experiment
   1. is completely randomized.
   2. has three factors (shampoo type, gender, whether hair is dyed).
   3. has two factors (gender and whether hair is dyed) blocked by shampoo type.
   4. has two factors (shampoo type and whether hair is dyed) blocked by gender.
   5. has one factor (shampoo type), blocked by gender and whether hair is dyed.
5. For a certain experiment, the available experimental units are eight rats, of which four are female (F1, F2, F3, F4) and four are male (M1, M2, M3, M4). There are to be four treatment groups, A, B, C, and D. If a randomized block design is used, with the experimental units blocked by gender, which of the following assignments of treatments is impossible?
6. A → (F1, M1), B → (F2, M2), C → (F3, M3), D → (F4, M4)
7. A → (F1, M2), B → (F2, M3), C → (F3, M4), D → (F4, M1)
8. A → (F1, M2), B → (F3, F2), C → (F4, M1), D → (M3, M4)
9. A → (F4, M1), B → (F2, M3), C → (F3, M2), D → (F1, M4)
10. A → (F4, M1), B → (F1, M4), C → (F3, M2), D → (F2, M3)
11. We wish to compare the average ages of the math and science teachers at your high school. Which is the best way to collect the data?
    1. Census
    2. Sample survey
    3. Observational study
    4. Experiment
    5. Simulation

**Use the following to answer questions 15and 16.**

One hundred volunteer subjects participated in a study to determine if room temperature affects people’s ability to concentrate. Female volunteers were given 10 minutes to try to memorize the words on a list of 50 nonsense words. The room temperature was controlled at 65 degrees (a cold room) while they complete the task. Male volunteers were also given 10 minutes to try to memorize the same list of words, but for the males, room temperature was controlled at 85 degrees (a hot room). At the end of the 10 minutes, each subject was asked to list as many of the words as he or she could remember, and the number correct was recorded. The resulting data were then used to determine if the mean number of words differed for the cold room and hot room conditions.

1. Which of the following is a confounding variable – that is, which of the following is confounded with the treatments?
   1. Room temperature (cold or hot)
   2. Gender (male or female)
   3. Number of nonsense words recalled
   4. Length of time given to memorize the words
   5. Can’t tell because volunteers were used
2. The poor design of this experiment results in a variable that is confounded with the treatments. Which of the following changes to the design would be effective in eliminating this confounding?
   1. Use only one room temperature.
   2. Use only male subjects in the study and assign the males to one of the two room-temperature conditions at random.
   3. Assign the volunteers at random to one of the two room-temperature conditions.
   4. Create two blocks by putting all of the females in one block and all of the males in the other block. Then, within each block, assign subjects at random to one of the room-temperature conditions.
   5. B, C, and D are all strategies that would be effective.

# Topic II. Exploratory Analysis of Data

1. The following quiz scores are from 2 different classes for an AP Stats test in chapter 1.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 4th Hour | 48 | 76 | 82 | 96 | 92 | 84 | 100 | 98 | 96 | 76 | 92 | 72 | 88 | 82 | 66 |
| 58 | 78 | 81 | 78 | 78 | 92 | 92 | 78 | 84 | 52 | 70 | 84 | 88 | 92 | 84 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5th Hour | 90 | 96 | 78 | 94 | 94 | 88 | 86 | 96 | 86 | 82 | 90 | 87 | 88 | 76 | 92 |
| 94 | 80 | 82 | 88 | 84 | 86 | 80 | 86 | 72 | 96 | 90 |  |  |  |  |

1. Create back-to-back box-plots (on the same scale).
2. Compare the distribution of quiz scores from the 4th hour and quiz scores from the 5th hour.
3. Which student performed better, the student in the 4th hour who made a 76 or the student in the 5th hour who made an 80? Justify your answer.
4. The data in the chart below shows the survival times in days for guinea pigs after they were injected with tubercle bacilli in a medical experiment.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 43 | 45 | 53 | 56 | 56 | 57 | 58 | 66 | 57 | 73 | 74 | 79 | 80 | 80 | 81 | 81 | 81 | 82 | 82 | 83 |
| 83 | 84 | 88 | 89 | 91 | 91 | 92 | 97 | 99 | 99 | 100 | 101 | 102 | 102 | 102 | 103 | 104 | 107 | 108 | 109 |
| 113 | 114 | 118 | 121 | 123 | 126 | 128 | 137 | 138 | 139 | 144 | 147 | 156 | 162 | 174 | 178 | 179 | 184 | 191 | 198 |
| 211 | 214 | 243 | 249 | 329 | 380 | 403 | 511 | 522 | 508 | 510 | 514 | 520 | 520 | 521 | 530 | 530 | 533 | 540 | 541 |

1. Create the following charts and graphs for the data in the chart above:
   1. Frequency Table
   2. Histogram
   3. Stemplot
2. Find the following values for the data.

Measures of **Center**: Median, Mean

Measures of **Spread**: Range, IQR, Standard Deviation, Variance

Measures of **Position**: Q1, Q2, Q3, Min, Max, The 70 Percentile

1. List the 5-number summary and create the modified box plot for the data.
2. Identify any outliers by using the IQR method.
3. Describe the data.
4. If the data were changed in the following ways, which one of the summary measures would change and how would they change?

Change the max days to 1000

Trim the data by 10%

Change the unit of measures by dividing every piece of data by 100

**Multiple Choice**

1. The five-number summary for a data set is given by min = 5, Q1 = 18, med = 20, Q3 = 40, max = 75. If you wanted to construct a modified boxplot for the data set (that is, one that would show outliers, if any existed), what would be the maximum possible length of the right-side “whisker”?
   1. 33
   2. 35
   3. 45
   4. 53
   5. 55
2. The frequency table below summarizes the times in the last month that patients at the emergency room of a small-city hospital waited to receive medical attention.



Which of the following represents possible values for the median and mean waiting times for the emergency room last month?

1. median = 27 minutes and mean = 24 minutes
2. median = 28 minutes and mean = 30 minutes
3. median = 31 minutes and mean = 35 minutes
4. median = 35 minutes and mean = 39 minutes
5. median = 45 minutes and mean = 46 minutes
6. Boxplots of two data sets are shown below.



Based on the boxplots, which statement below is true?

1. The spread of both plots is about the same.
2. The means of both plots are approximately equal.
3. Plot 2 contains more data points than Plot 1.
4. The medians are approximately equal.
5. Plot 1 is more symmetric than Plot 2.

Here is a dotplot of the adult literacy rates in 177 countries in 2008, according to the United Nations. For example, the lowest literacy rate was 23.6%, in the African country of Burkina Faso.



1. The overall shape of this distribution is
   1. clearly skewed to the right
   2. clearly skewed to the left
   3. roughly symmetric
   4. uniform
   5. there is no clear shape
2. The mean of this distribution (*don’t* try to find it) will be
   1. very close to the mode
   2. greater than the median
   3. less than the median
   4. you can’t say, because the median is random
   5. you can’t say, because the mean is random
3. Based on the shape of this distribution, what numerical measures would best describe it?
   1. The five-number summary
   2. The mean and standard deviation
   3. The mean and the quartiles
   4. The median and the standard deviation
   5. It is not possible to determine which numerical values to use
4. Which one of the following would be a correct interpretation if you have a z-score of +2.0 on an exam?
   1. It means that you missed two questions on the exam.
   2. It means that you got twice as many questions correct as the average student.
   3. It means that your grade was two points higher than the mean grade on this exam.
   4. It means that your grade was in the upper 2% of all grades on this exam.
   5. It means that your grade is two standard deviations above the mean for this exam.
5. A data set consists of five numbers: -6.0, -4.5, 0, 5.0, and an unknown 5th number. For these five data points, which of the following statistics can never be greater than zero?
   1. The mean
   2. The median
   3. The standard deviation
   4. The interquartile range
   5. All of the above could possibly be greater than zero depending on the 5th number
6. A large company is interested in improving the efficiency of its customer service and decides to examine the length of the business phone calls made to clients by its sales staff. A cumulative relative frequency graph is shown below from data collected over the past year. According to the graph, the shortest 80% of calls will take how long to complete?



* 1. Less than 10 minutes.
  2. At least 10 minutes.
  3. Exactly 10 minutes.
  4. At least 5.5 minutes.
  5. Less than 5.5 minutes.

1. Which of the following statements are true?
   * 1. It is impossible for a data set to have a standard deviation that is larger than the mean.
     2. If two data sets have different ranges, the data set with the larger range will always have the larger standard deviation.
     3. The interquartile range of a data set can never be larger than the range.
2. i only
3. ii only
4. iii only
5. ii and iii only
6. i, ii, and iii

# Topic III Anticipating Patterns: Probability and Simulation

1. Elaine is enrolled in a self-paced course that allows three attempts to pass an examination on the material. She does not study and has 2 out of 10 chances of passing on any one attempt by pure luck. What is Elaine’s likelihood of passing, provided that she will have three attempts to pass the exam? (Assume the attempts are independent because she takes a different exam at each attempt.)
   1. Explain how you would use a random digit table to simulate Elaine’s attempts at the exam. Elaine will of course stop taking the exam as soon as she passes.
   2. Simulate 10 repetitions using the random digits below. What is your estimate of Elaine’s likelihood of passing the course?

59636 88804 04634 71197 19352 73089 84898 45785

62568 70206 40325 03699 71080 22553 11486 11776

* 1. A more realistic model for Elaine’s attempts to pass an exam would be as follows: On the first try she has a probability 0.2 of passing. If she fails on the first try, her probability on the second try increases to 0.3 because she learned something from the first try. If she fails on the first 2 attempts, the probability of passing on the third attempt is 0.4. She will stop as soon as she passes. The course rules force her to stop after three attempts. Explain how to simulate one repetition of Elaine’s tries on the exam with this new approach.
  2. Simulate 10 repetitions and estimate the probability that Elaine eventually passes the exam with the approach in part c.

13873 81598 95052 90908 73592 75186 87136 95761

54580 81507 27102 56027 55892 33063 41842 81868

1. Probability is a measure of how likely an event is to occur. Match one of the probabilities that follow with each statement about an event.

0 0.01 0.3 0.6 0.99 1.00

1. The sun will rise in the west in the morning.
2. New Year’s Eve will be on December 31st next year.
3. An event is very unlikely, but it will occur vary rarely.
4. The event will occur most of the time. Very rarely will it not occur.
5. Give an example of a context where the 2 remaining probabilities might occur.
6. What is the formula used for each of the following probability calculations? Are there any special conditions required to use any of these formulas?
   1. Addition Rule (general & disjoint)
   2. Multiplication Rule (general & independent)
   3. Conditional Probability
7. The following table reports Census Bureau data on undergraduate students in U.S. colleges and universities in the fall of 1991.

UnderGraduate College enrollment by age of students – Fall 1991 (thousands of students)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Age | 2-yr full-time | 2-yr part-time | 4-yr full time | 4-yr part-time | Totals |
| 15-17 | 44 | 4 | 79 | 0 |  |
| 18-21 | 1345 | 456 | 3869 | 159 |  |
| 22-29 | 489 | 690 | 1358 | 494 |  |
| 30-44 | 287 | 704 | 289 | 627 |  |
| 45+ | 49 | 209 | 62 | 160 |  |
| Totals |  |  |  |  |  |

* 1. Fill in the “totals” in the table above. What is the grand total (GT) of students who were enrolled in colleges and universities in the fall of 1991?
  2. What percent of all undergraduate students were 18-21 years old in the fall of the 1991?
  3. Find the percent of the undergraduates enrolled in each of the four types of programs who were 18-21 years old. Make a bar chart to compare these percents.
  4. The 18-21 group is the “traditional” age group for college students. Briefly summarize what you have learned from the data about the extent to which this group predominates in different kinds of college programs.

1. The type of medical care a patient receives may vary with the age of the patient. A large study of women who had a breast lump investigated whether or not each woman received tests (a mammogram and a biopsy) when the lump was discovered. 44.5% of the women are under 65 years old. The probability that a patient is under 65 years of age and the tests were done is 0.321. For those 65 and older, the probability of having tests done is 0.658.
2. What is the probability that a patient in this study is under 65 and did not have tests done?
3. What is the probability that a patient in this study did not have tests done?
4. A patient in this study had tests done. What is the probability that this patient is under 65?
5. Are the events A = (patient was 65 or older) and B= (the tests were done) independent? Were the tests omitted on older patients more or less frequently than would be the case if testing were independent of age?
6. Here are the counts (in thousands) of earned degrees in the United States in a recent year, classified by level and by the sex of the degree recipient:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Bachelor’s | Master’s | Professional | Doctorate | Total |
| Female | 616 | 194 | 30 | 16 |  |
| Male | 529 | 171 | 44 | 26 |  |
| Total |  |  |  |  |  |

1. If you choose a degree recipient at random, what is the probability that the person you choose is a woman?
2. What is the probability that you choose a woman, given that that person chosen received a professional degree?
3. Are the events “choose a woman” and “choose a professional degree recipient” independent? How do you know?
4. Consolidated Builders has bid on two large construction projects. The company president believes that the probability of winning the first contract (event A) is 0.6, that the probability of winning the second (event B) is 0.4 and the probability of winning both jobs is 0.2 .
5. Draw the Venn diagram that illustrates the relationship between events A and B.
6. Find the following probabilities:

P(A or B) P(A and B) P(A, and Not B)

P(Not A, and B) P(not A and not B)

1. What is the difference between discrete and continuous random variables?
2. Let X be the number of courses for which a randomly selected student at a certain university is registered. The probability distribution of X appears in the accompanying table.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| P(X) | 0.02 | 0.03 | 0.09 | 0.25 | 0.40 | 0.16 | 0.05 |

* 1. What is P(X = 4)?
  2. What is P(X ≤ 4)?
  3. What is the probability that the selected student is taking at most five courses?
  4. What is the probability that the selected student is taking at least five courses?
  5. Calculate P(3 ≤ X ≤ 6) and P(3< X < 6). Explain why the two probabilities are different.
  6. Find the mean, standard deviation and variance of the random variable x.

1. You have two scales for measuring weights in a chemistry lab. Both scales give answers that vary a bit in repeated weightings of the same item. If the true weight of a compound is 2.00 grams, the first scale produces readings X that have mean 2.000 grams and standard deviations 0.002 grams. The second scale’s readings Y have mean 2.001 grams and standard deviation of 0.001 grams.
2. What are the mean and standard deviation of the difference Y – X between the readings? (The readings X and Y and independent.)
3. You measure once with each scale and average the readings. Your result is Z = (X + Y)/2. What are the mean and standard deviation of Z?
4. Among employed women, 25% have never been married. You select 10 employed women at random and identify their marriage status.
5. Determine the random variable X and describe its probability distribution. Explain your choice.
6. Create a probability distribution table and a probability histogram for this data.
7. What is the probability that exactly 2 of the 10 women in your sample have never been married?
8. What is the probability that 2 or fewer have never been married?
9. What are the mean and standard deviation for this distribution?
10. A basketball player makes 80% of his free throws. We put him on the free throw line and ask him to shoot free throws until he misses one. Let X = number of free throws the player takes until he misses.
11. What is the appropriate probability model? What assumptions do you need to make in order for this model to apply? Verify that X has the identified distribution. What actions constitute “success” in this context?
12. What is the probability that the player will make 5 shots before he misses?
13. What is the probability that he will make at most 5 shots before he misses?
14. What is the mean of this distribution?
15. The area under the curve for a normal distribution is represented by a bell-shaped curve. What are the properties of a normal distribution? Sketch a normal curve.
16. A certain population of whooping cranes that migrate between Wisconsin and Florida every year has a SRS taken. The sample of 15 male cranes was weighed before they left Wisconsin to begin their trip. The mean weight of the 15 males was found to be 22.7 pounds with a standard deviation of 2.3 pounds. It is generally accepted that the weights of animal species are approximately normally distributed.
17. What is the probability that a random selected male crane weights less than 20 pounds?
18. What is the probability that a random selected male crane weighs more than 25 pounds?
19. What is the probability that a random selected male crane weighs between 21 and 26 pounds?
20. When these cranes reach Florida, another random sample of 25 male cranes is weighed and measured. The mean weight is recorded at 19.5 pounds with a standard deviation of 1.7 pounds. Assuming these sample statistics are equal to the population parameters, what is the probability that the mean of a new random sample of 25 male cranes will be between 15 and 22 pounds?
21. What is the probability that the new sample of 25 cranes would have a mean greater than 23 pounds?
22. What is the probability that a randomly selected crane would be less than 18 pounds at the end of the trip?
23. The Helsinki Heart Study asks whether the anti-cholesterol drug gemfibrozil will reduce heart attacks. In planning such an experiment, the researchers must be confident that the sample sizes are large enough to enable them to observe enough heart attacks. The Helsinki study plans to give gemfibrozil to 2000 men and a placebo to another 2000 men. The probability of a heart attack during the 5-year period of the study for men this age is about 0.04. We can think of the study participants as an SRS from a large population, of which the proportion p = 0.04 will have heart attacks.
24. What is the expected number of heart attacks that the study will find in one group of 2000 men (provided that the treatment doesn’t change the probability of 0.04)?
25. What is the probability that the group will suffer at least 75 heart attacks?
26. Children in kindergarten are sometimes given the Ravin Progressive Matrices Test (RPMT) to assess their readiness for learning. Experience at Southward Elementary School suggests that the RPMT scores for its kindergarten pupils have a mean of 13.6 with a standard deviation of 3.1. The distribution is close to normal. Mr. Brown has 22 children in his kindergarten class this year.
27. What is the probability that class’s mean score will be less than 12.0?
28. Mr. Brown suspects that the class RPMT scores will be unusually low because the test was interrupted by a fire drill. He wants to find the level L such that there is only a probability of 0.05 that the mean score of his class fall below L. What is this value of L. (Hint: this requires you to find the z-score and then convert to the x-score.)
29. Explain what is meant by the Law of Large Numbers. How does this law apply to sampling distributions?
30. What is the Central Limit Theorem? How is the CLT used in sampling distributions?

**Multiple Choice**

1. A child is 40 inches tall, which places her at the 90th percentile of all children of a similar age. The heights for children of this age form an approximately Normal distribution with a mean of 38 inches. Based on this information, what is the standard deviation of the heights of all children of this age?
   1. 0.20 inches
   2. 0.31 inches
   3. 0.65 inches
   4. 1.21 inches
   5. 1.56 inches
2. A large set of test scores has mean 60 and standard deviation 18. If each score is doubled, and then 5 is subtracted from the result, the mean and standard deviation of the new scores are
   1. Mean 115, std. dev. 31
   2. Mean 115, std. dev. 36
   3. Mean 120, std. dev. 6
   4. Mean 120, std. dev. 31
   5. Mean 120, std. dev. 36
3. The figure below shows a Normal density curve. Which of the following gives the best estimates for the mean and standard deviation of this Normal distribution?



* 1. µ = 200, σ = 50
  2. µ = 200, σ = 25
  3. µ = 225, σ = 50
  4. µ = 225, σ = 25
  5. µ = 225, σ = 275

1. The probability distribution for the number of heads in four tosses of a coin is given by:



The probability of getting at least one tail in four tosses of this coin is

* 1. 0.2500
  2. 0.3125
  3. 0.6875
  4. 0.9375
  5. None of these

1. In a certain large population of adults, the distribution of IQ scores is strongly left-skewed with a mean of 122 and a standard deviation of 5. Suppose 200 adults are randomly selected from this population for a market research study. The distribution of the sample mean of IQ scores is
   1. Left-skewed with mean 122 and standard deviation 0.35
   2. Exactly normal with mean 122 and standard deviation 5
   3. Exactly normal with mean 122 and standard deviation 0.35
   4. Approximately normal with mean 122 and standard deviation 5.
   5. Approximately normal with mean 122 and standard deviation 0.35.
2. Suppose we roll a fair die four times. The probability that a 6 occurs on exactly one of the rolls is
   1. 
   2. 
   3. 
   4. 
   5. 
3. The number of unbroken charcoal briquets in a twenty-pound bag filled at the factory follows a Normal distribution with a mean of 450 briquets and a standard deviation of 20 briquets. The company expects that a certain number of the bags will be underfilled, so the company will replace for free the 5% of bags that have too few briquets. What is the minimum number of unbroken briquets the bag would have to contain for the company to avoid having to replace the bag for free?
   1. 404
   2. 418
   3. 448
   4. 411
   5. 425
4. Suppose that you have torn a tendon and are facing surgery to repair it. The orthopedic surgeon explains the risks to you. Infection occurs in 3% of such operations, the repair fails in 14%, and both infection and failure occur together 1% of the time. What is the probability that the operation is successful for someone who has an operation that is free from infection?
   1. 0.0767
   2. 0.8342
   3. 0.8400
   4. 0.8600
   5. 0.9900
5. According to the U.S. Census, the proportion of adults in a certain county who owned their own home was 0.71. An SRS of 100 adults in a certain section of the county found that 65 owned their home. Which one of the following represents the approximate probability of obtaining a sample of 100 adults in which fewer than 65 own their home, assuming that this section of the county has the same overall proportion of adults who own their home as does the entire county?
   1.  d. 
   2.  e. 
   3. 
6. Records from a random sample of dairy farms yielded the information below on the number of male and female calves born at various times of the day.



What is the probability that a randomly selected calf was born in the night or was a female?

* 1. 
  2. 
  3. 
  4. 
  5. 

1. When people order books from a popular online source, they are shipped in standard-sized boxes. Suppose that the mean weight of the boxes is 1.5 pounds with a standard deviation of 0.3 pounds, the mean weight of the packaging material is 0.5 pounds with a standard deviation of 0.1 pounds, and the mean weight of the books shipped is 12 pounds with a standard deviation of 3 pounds. Assuming that the weights are independent, what is the standard deviation of the total weight of the boxes that are shipped from this source?
   1. 1.84
   2. 2.60
   3. 3.02
   4. 3.40
   5. 9.10
2. A grocery chain runs a prize game by giving each customer a ticket that ma win a prize when the box is scratched off. Printed on the ticket is a dollar value ($500, $100, $10) or the statement, “This ticket is not a winner.” Monetary prizes can be redeemed for groceries at the store. Here are the distribution of the prize values and the associated probabilities for each prize:



Which of the following are the mean and standard deviation, respectively, of the winnings?

* 1. $15.00, $2900.00
  2. $15.00, $53.85
  3. $15.00, $26.93
  4. $156.25, $53.85
  5. $156.25, $26.93

1. Suppose the probability that a softball player gets a hit in any single at-bat is .300. Assuming that her chance of getting a hit on a particular time at bat is independent of her other times at bat, what is the probability that she will not get a hit until her fourth time at bat in a game?
   1.  d. 
   2.  e. 
   3. 
2. The probability that Color Me Dandy wins a horse race at Batavia Downs given good track conditions is 0.60. The probability of good track conditions on any given day is 0.85. What is the probability that Color Me Dandy wins or the track conditions are good?
   1. 0.94
   2. 0.51
   3. 0.49
   4. 0.06
   5. The answer cannot be determined from the given information.
3. Let X represent the score when a fair six-sided die is rolled. For this random variable,  and . If the die is rolled 100 times, what is the approximate probability that the total score is at least 375?
   1. 0.0000
   2. 0.0017
   3. 0.0721
   4. 0.4420
   5. 0.9279
4. Commute distances for employees of a large company have a distribution that is quite skewed. Suppose the mean commute distance is 20 miles and that the standard deviation is 10 miles. What can be said about the mean commute distance for a random sample of size 4 employees from this plant?
   1. The distribution of  will be approximately normal with mean 20 and standard deviation 10.
   2. The distribution of  will be approximately normal with mean 20 and standard deviation 5.
   3. The mean and standard deviation of  will be 20 and 10, but the sampling distribution of  will not be approximately normal.
   4. The mean and standard deviation of  will be 20 and 5, but the sampling distribution of  will not be approximately normal.
   5.  will have a standard normal distribution.
5. Sixty percent of the students at an urban university carry more than one credit card. Selecting a student at random from the students at this university can be viewed as an experiment with two possible outcomes: *has more than one credit card* and *does not have more than one credit card*. Which of the following would be a correct assignment of random digits to the outcomes in order to simulate selecting a student at random from the students enrolled at the university?
   1. 0 represents more than one credit card, 1 represents not more than one credit card, ignore all other digits.
   2. Odd digits represent more than one credit card, and even digits represent not more than one credit card.
   3. 0, 1, 2, and 3 represent more than one credit card, and 4, 5, 6, 7, 8, and 9 represent not more than one credit card.
   4. 0, 1, 2, 3, 4, and 5 represent more than one credit card, and 6, 7, 8, and 9 represent not more than one credit card.
   5. C and D both represent correct assignments.
6. In order for the sampling distribution of the sample mean to be approximately normal, which of the following must be true?
7. The sample size must be large.
8. The population distribution must be normal or approximately normal.
   1. I only
   2. II only
   3. Either I or II
   4. Both I and II
   5. Neither I nor II is necessary
9. A and B are events with P(A) = 0.6, P(B) = 0.7, and P(A ∩ B) = 0.4. Which of the following is true?
   1. A and B are mutually exclusive.
   2. A and B are independent.
   3. P(A U B) = 0.6 + 0.7
   4. P(A|B) = P(A)
   5. None of the above is true.
10. Ten percent of the students at a particular university smoke. Consider the chance experiment that consists of selecting a random sample of 20 students from this university. Let X = the number of students in the sample who smoke. The random variable X has which of the following probability distributions?
    1. A binomial distribution
    2. A geometric distribution
    3. A normal distribution
    4. A uniform distribution
    5. None of the above
11. For a normally distributed population with mean 0 and standard deviation 1.0, the population interquartile range is closest to which of the following values?
    1. 0.50
    2. 1.28
    3. 1.349
    4. 1.645
    5. 1.96
12. Which of the following is not a property of a binomial experiment?
    1. It consists of a fixed number of trials, *n*.
    2. Outcomes of different trials are independent.
    3. Each trial can result in one of several different outcomes.
    4. Observations consist of the number of successes for each trial of the experiment.
    5. The probability of success is constant for each trial.
13. Suppose that X is a random variable that has a uniform distribution over the interval from 0 to 10. Which of the following probabilities is largest?
    1. P(X > 8)
    2. P(X ≥ 8)
    3. P(X = 10)
    4. P(3 < X < 6)
    5. P(X < 4)
14. Which of the following best describes what is meant by the term *sampling variability*?
    1. There are many different methods for selecting a sample.
    2. Two different samples from the same population may have different values for a sample statistic.
    3. The sample size can sometimes vary from sample to sample.
    4. Convenience samples may not be representative of the population.
    5. Population characteristics vary from population to population.
15. A random sample is to be selected from a population. For which combination of *n* and *p* is it reasonable to assume that the sampling distribution of the sample proportion  will be approximately normal?
    1. n = 10 and p = .4
    2. n = 25 and p = .5
    3. n = 30 and p = .2
    4. n = 40 and p = .1
    5. n = 100 and p = .05
16. The histograms shown here are approximate sampling distributions of the sample mean. Each histogram is based on selecting 500 different samples, each of size n. All three histograms were constructed by sampling from the same population, but the sample sizes were different. Which histogram was based on samples with the smallest sample size, n?

|  |  |  |
| --- | --- | --- |
|  | midterm review picture 2.JPG | midterm review picture 3.JPG |

* 1. I
  2. II
  3. III
  4. The same sample size was used for all three histograms.
  5. It is not possible to tell by looking at the histograms.

1. For which of the following will the sample mean tend to differ least from sample to sample?
   1. Random samples of size 10
   2. Random samples of size 20
   3. Random samples of size 30
   4. Random samples of size 50
   5. Random samples of size 100
2. The Air Force receives 20% of its parachutes from Company C1, 55% from Company C2, and the rest from Company C3. The probability that a randomly selected parachute will fail to open is 0.0025, 0.002, and 0.0022, depending on whether it is from Company C1, C2, or C3, respectively. If a randomly chosen parachute fails to open, what is the probability it is from company C3?
   1. 0.00055
   2. 0.00215
   3. 0.0022
   4. 0.250
   5. 0.256
3. Which of the following is true?
4. The sampling distribution of  has a mean equal to *p*.
5. The sampling distribution of  has a standard deviation equal to .
6. The sampling distribution of  is always symmetric.
   1. I only
   2. II only
   3. III only
   4. I and II only
   5. I and III only
7. Suppose the true proportion of people who use public transportation to get to work in the Washington, D.C., area is 0.45. In a simple random sample of 250 people who work in Washington, what is the standard deviation of the sampling distribution of ?
   1. 0.4975
   2. 0.2475
   3. 0.0315
   4. 0.0009
   5. 0.617