

Exam 1 Review

- Exam 1 is Monday July 8 and covers Chapters 1-4.
- You may use one (regular size) page of notes with notes on one side in a font size not to exceed 12 points (the size of font on this sheet).
- Cell phones may not be used as calculators and must be silenced and put away.
- Study ideas – review vocabulary, homework problems, quizzes, in-class worksheets, do additional problems from the book – examples can be found in the text and more problems at the end of the chapter. Re-work homework problems, don't just look them over. Make sure you do and understand the review problems
- I will not ask you for definitions.

Topics

- Population vs. Sample
- Parameter vs. Statistics
- Types of data - qualitative vs. quantitative; discrete vs. continuous quantitative data
- Types of sampling – random, simple random, probability, stratified, cluster, convenience
- Bias in sampling and methods (occurs when a population is *systematically* over or under represented)
- Percentages
- Measures of Center
 - mean, median, mode
- Measures of Variation
 - range, standard deviation (sample vs. population), variance (sample vs. population)
- Measures of Relative Standing
 - z-scores, percentiles, quartiles, 5-number summary, interquartile range
- Graphs
 - box plots, histograms, line plots, scatter plots, bar charts, stem and leaf plots
- Outliers
- Tables
 - frequency and relative frequency charts
- Shapes of graphs
 - symmetric, normal, uniform, skewed left, skewed right
- Normal graph – the empirical rule (page 19)
- “unusual” values
- Interquartile range

- Probability
 - sample space
 - empirical vs. theoretical probability
 - independent events
 - mutually exclusive
 - complementary events
 - addition rule
 - multiplication rule
 - conditional probability
-

Practice Problems

Disclaimer: The following set of practice problems is NOT meant to be an exhaustive list of all types of problems that may be on your test.

1. A university has 5,000 students. Describe a method for choosing a sample of 50 students that would illustrate the following sampling methods:
 - a. cluster
 - b. stratified
 - c. convenience
 - d. systematic

2. Classify each of the following as quantitative or qualitative (categorical). If quantitative identify it as discrete or continuous type data:
 - a. The number of close friends a person has.
 - b. The time it takes a person to brush their teeth in the morning.
 - c. A person's phone number.
 - d. A person's hair color.

3. Consider the following set of 14 data points:
45, 76, 90, 57, 69, 55, 54, 49, 71, 57, 56, 60, 44, 10
 - a. Make a stem and leaf graph of the data.
 - b. Make a histogram of the data.
 - c. Find the 5-number summary of the data and make a box plot.
 - d. What is the interquartile range?
 - e. What is the range?
 - f. Are there any outliers? Explain.
 - g. Find the mean.
 - h. Find the median.

- i. Find the mode.
 - j. Find the standard deviation.
 - k. Are there any unusual values? Explain.
4. A new “quick test” for pneumonia was administered to 850 patients. A “positive” test result means that the test predicts that the person has pneumonia. The results are summarized in the following table.

Test Result	Did the person actually have pneumonia?	
	YES	NO
POSITIVE	450	60
NEGATIVE	20	320

- a. What is the probability that a randomly chosen person had pneumonia given that the test result was positive?
 - b. What is the probability that the test incorrectly predicted whether or not a person had pneumonia?
5. The probability that a phone will be answered when someone calls the “helpline” is 0.65. Assume that whether one call will be answered is INDEPENDENT of whether another will.
- a. Find the probability that all of the first three calls will all be answered.
 - b. Find the probability that at least one of the first four calls will be answered.
 - c. Find the probability that exactly one of the first four calls will be answered.
6. Two cards are dealt at random from a standard 52 card deck (WITHOUT REPLACEMENT). (A description of a deck of cards is on the back of this sheet.)
- a. Find the probability that the first card is a face card and the second is NOT a face card.
 - b. Find the probability that they are both face cards.
 - c. Find the probability that the second is a face card given the first is NOT a face card.
 - d. Find the probability that the second card is a face card.

7. A lie detector test was administered to 1000 people to determine the accuracy of the test. A “fail” test result means the test indicated that the person lied.

Test Result	Was the person lying?	
	YES	NO
PASS	70	610
FAIL	240	80

- a. What is the probability that a randomly chosen person was lying?
 - b. What is the probability that a person would PASS the test given that they were NOT lying?
 - c. What is the probability of a false positive (meaning here a PASS on the test given they were lying)?
 - d. What is the probability the test accurately predicts whether or not the person was lying?
8. A class consists of 100 women and 50 men. Consider the following three different sampling methods for collecting a sample of size 15:
- a. I randomly select 10 women and 5 men.
 - b. I randomly select 15 people from my sample.
 - c. I randomly select 5 women and 10 men.
- I. For each case, find the probability that a woman is selected $P(W)$, and the probability that a man is selected $P(M)$.
 - II. For each identify if the sampling is random, simple random, or probability. Circle all that apply
 - a. Random Simple Random Probability
 - b. Random Simple Random Probability
 - c. Random Simple Random Probability
 - III. Which of the above (a,b,c) will give us a stratified sample (according to our definition in class, not the book)?
9. Did the following numbers most likely come from data that was a parameter or statistic?
- a. The percentage of the licensed drivers in Oregon that are female is 52%.
 - b. 1 in 4 cats has green eyes.
 - c. 32% of people eat cheese every day.

10. Find the 40th and 70th percentiles of the following 12 data points:
10.2, 9.8, 11, 6.5, 13.1, 14.5, 18.2, 4.9, 5.1, 10.1, 16.9, 18
11. Suppose a set of data has a normal distribution with a mean of 50 and a standard deviation of 5. Fill in the following blanks:
68 % of the data in the set falls between _____ and _____.
_____% of the data falls between _____ 40 _____ and _____ 60 _____.
What percent of the data falls within 3 standard deviations of the mean? _____
12. Scores on a state test follow a normal distribution with mean 234 and standard deviation
- A student scores 240 on the test. What is the z-score?
 - A score above what value would be considered *unusually high* for this test?
13. Sketch a graph that is skewed right.
14. A pair of jeans cost \$100 and a skirt costs \$80. Two of the following statements are true. Two are false. Which are true?
- The skirt cost 25% less than the jeans.
 - The jeans cost 25% more than the skirt.
 - The jeans cost 20% more than the skirt.
 - The skirt cost 20% less than the jeans.
15. Suppose that a bag of marbles has 6 yellow, 5 red, and 3 green. I reach in and grab 2 marbles (without replacement).
- Write down the sample space for the set of outcomes of this experiment. Use correct set notation.
 - Are all outcomes in your sample space equally likely? Explain.
 - Find the probability that the second marble is yellow.
16. Suppose that every strand of holiday lights sold by EverLight has a probability of 0.25 of being defective. Assume independence between strands of lights being defective.
- If I purchase 3 strands of lights, find the probability that exactly one strand will be defective.
 - If I purchase 3 strands of lights what is the probability that *at least* one strand will be defective?
 - If I purchase 5 stands of lights, what is the probability that at most one strand will be defective? (Think carefully about what outcomes are in this event.)

Answers:

Disclaimer: Sometimes I type the wrong answer – if your answer differs from mine and you cannot figure out why, email me to make sure there is not a typo.

1. (other answers are ok, but the general idea should be the same)
 - a. Cluster: We could select entire classes and use everyone in the class. Choose enough so we have 50 students.
 - b. Stratified. We choose a sample of 50 freshman, sophomores, juniors and seniors in the same proportion as they appear in the University population
 - c. convenience – we stand in the dining hall and ask the first 50 people we see.
 - d. Systematic: We alphabetize all students, then we take the first person, and every 10th person after that until we have 50.
2.
 - a. Quantitative – discrete
 - b. Quantitative – continuous
 - c. Qualitative
 - d. Qualitative
3.
 - a. See page 40 in your book for how to make a stem and leaf graph
 - b. See page 38 in your book for how to make a histogram
 - c. 5-number: (10,49,56.5, 69,90). See page 22 of your book for how to make a box plot
 - d. 20
 - e. 80
 - f. Yes, 10 is an outlier because it is more than 1.5 times the Interquartile range below the left edge of the box
 - g. 56.6
 - h. 56.5
 - i. 57
 - j. 17.7 (population) or 18.4 (sample)
 - k. Yes, 10 and 99 are both more than 2 standard deviations away from the mean.
4.
 - a. $450/510 = 0.882$
 - b. $80/850 = 0.094$
5.
 - a. 0.275
 - b. 0.985
 - c. 0.111
6.
 - a. $480/2652 = 0.181$

MTH 243 Introduction to Probability and Statistics

- b. $132/2652 = 0.05$
 - c. $12/51 = 0.235$
 - d. $612/2652 = 0.231$
- 7.
- a. $310/1000 = 0.310$
 - b. $610/690 = 0.884$
 - c. $70/310 = 0.226$
 - d. $850/1000 = 0.850$
8. Do not to I,II, the answer to III is (a)
- 9.
- a. Parameter (they could get that information from the DMV)
 - b. Statistic
 - c. Statistic
10. 40th percentile is 10.1; 70th percentile is 14.5
11. 44 and 55, 95%, 99.7%
- 12.
- a. z-score is 0.4
 - b. Above 265 is unusually high
13. the tail should be on the right
14. False, True, False, True
- 15.
- a. {YY, RR, GG, YR,RY, YG, GY, RG, GR}
 - b. No, they are not equally likely. There are a different number of each color, making some combinations more likely than others.
 - c. $78/182 = 0.429$
- 16.
- a. 0.422
 - b. 0.578
 - c. 0.633