

## Review Questions for Consideration

### Chapter 1

1. What are the various ways we can use to graph univariate data? I count 4 major ways. Give a brief explanation of the process involved in for each way.
2. What are the two competing quantitative methods used to describe univariate data? Give a brief definition or explanation of each parameter within the two competing methods.
3. What are the main things we look for in graphs of distributions? What additional things should we also consider in these graphs?
4. How do you know if a data set is skewed left? Skewed right? Be specific. How would a boxplot look that was skewed in each direction?
5. What is meant by resistant? Non-resistant? Which of the univariate parameters that we have studied would fall into each category?
6. What is the formula for:
  - a. mean of a sample?
  - b. standard deviation of a sample?
  - c. What symbols are used for the mean and standard deviation of a sample?
  - d. What symbols (don't worry about formulas) are used for the mean and standard deviation of a population?
  - e. How are the variance and standard deviation of a sample (or population) related? Be sure to include a unit of measurement relationship for the last question.

### Chapter 2

1. What happens to the mean, median, standard deviation and IQR of a data set if you add a fixed number to each member of a data set? What happens to these same measures if you take a data set and *multiply* each value by a fixed number?
2. What conclusions can you make about a distribution if the mean = median? Mean > median? Mean < median? Briefly explain why you can make these conclusions.
3. What does a density curve represent? What are the two basic requirements for a density curve?

4. What does the empirical rule say about all normal distributions? How do you standardize observations from a normal distribution with mean  $\mu_x$  and standard deviation  $\sigma_x$  into one that is normally distributed with mean 0 and standard deviation 1?
5. Describe how you would use Table A in the book in two ways. First, to determine the proportions of observations within two given values in a normally distributed population. Second, to determine the observation values that would account for a particular proportion of the data in a normally distributed population.
6. What is a percentile score?
7. Briefly describe how you would use a histogram of a data set to assess normality. Would using a dotplot instead of a histogram be a more accurate assessment? Why or why not? Yes, I agree using a dotplot may be more tedious, but that is not the question
8. Describe how using a normal probability plot can be used to assess normality. Be sure to address what you are plotting on the y-axis and the x-axis on the plot. Also be sure to briefly explain what you are looking for in this plot.

### Chapter 3 -

1. What are the qualitative things you should look for in a scatterplot of bivariate data?
2. How can you add a categorical variable to a scatterplot? WHY would you want to add a categorical variable to a scatterplot? Give an example.
3. How do you calculate r for a data set? What does this have to do with Z-scores? What is the interpretation of this r value? What are the units of an r value? What are the units of a z-score?
4. What are the properties of r?
5. What is the difference between a correlation coefficient and a coefficient of determination?
6. In addition to the LSRL ( $\hat{y} = a + bx$ ), r and  $r^2$  what things should you report with your analysis of bivariate data? HINT: It's a graph.
7. What is our GOAL in obtaining a LSRL? That is, what have we optimized or minimized? Would the LSRL be different if we switched the explanatory and the response variable? What about r and  $r^2$ ?
8. What is the relationship between the slope of the line and the correlation coefficient? How can you find the y-intercept of the LSRL

- given only the correlation coefficient and the mean-based statistics for both  $x$  and  $y$ ?
9. What  $(x,y)$  can you GUARANTEE will always be on the LSRL?
  10. What the heck is an outlier? An influential point? Can a point be one and not the other?
  11. How do you calculate residual values? What is the sum of the residuals? What is the sum of the squares of the residuals and what does it have to do with the LSRL?
  12. When analyzing a residual plot, what should you look for?
  13. What is the interpretation of  $r^2$ ?

#### Chapter 4 –

1. A well-designed survey will have three properties. For each of the following explain who, what, etc. that is appropriate.
  - Representative
  - Random
  - Large
2. What is a SRS? Does each individual have an equal chance of selection in an SRS? What about each combination of individuals?
3. How do you use a table of digits to get a random sample?
4. What is the difference between cluster and stratified sampling? What is multi-stage sampling?
5. What is the difference between stratified sampling and blocking? Hint: While they are analogous, they are used in different situations.
6. Give a brief explanation of each of the four main principles of experimental design including their purpose and how you can ensure they are addressed.
  - a. CONTROL
  - b. RANDOMIZATION
  - c. REPLICATION
  - d. BLOCKING
7. What is meant by statistically significant?
8. Describe each of the following types of bias
  - a. Voluntary response
  - b. Convenience sampling
  - c. Undercoverage
  - d. Nonresponse (please pay special attention to the difference between this and undercoverage).
  - e. Positive Response
  - f. Wording of questions

9. What distinguishes an experiment from an observational study? What do you hope to determine with an experiment that you cannot determine with an observational study?
10. Give an example of a basic blocking design in an experiment.
11. You have three factors in a liquid experiment - temperature, salt content and color. For temperature you have 3 choices, salt content you have 4 choices and color you have 2 choices. How many treatment groups will be necessary for your experiment?
12. You wish to simulate a poll result cited by FOX news discussing the 2004 democratic primaries. They estimate that 22.3 % of Iowa voters support Gephardt, 24.6 % support Dean and the remaining support another candidate. How will you simulate this? How many digits will you need?

### Chapter 5 –

1. What are the two major components of a probability model?
2. What is a sample space?
3. What is the multiplication principle and how would you use it?
4. IF DISJOINT –  $P(A \text{ or } B) = ?$
5. IF NOT DISJOINT –  $P(A \text{ or } B) = ?$
6. IF INDEPENDENT –  $P(A \text{ and } B) = ?$
7. IF NOT INDEPENDENT  $P(A \text{ and } B) = ?$
8. IF NOT DISJOINT and also INDEPENDENT,  $P(A \text{ or } B) =$
9. Draw a Venn Diagram to represent the following situation:  $P(A) = .3$ ,  $P(B) = .6$ ,  $P(A \text{ and } B) = .2$ 
  - Are events A and B independent? How do you know?
  - Are events A and B disjoint? How do you know?
  - Find the following:
    - $P(A \text{ and } B)$
    - $P(A \text{ and } B^c)$
    - $P(A^c \text{ and } B)$

- $P(\text{Neither A nor B})$
- $P(\text{Not A and B})$

10. The probability of guessing a correct answer on a multiple choice test is .2 (assuming each question has 5 choices). You are going to take a 6 question multiple choice test and guess the answers on each. Find the following (expressions are fine):
- $P(\text{getting all correct})$
  - $P(\text{getting all incorrect})$
  - $P(\text{not getting all correct})$ ...note this is not the same as  $P(\text{getting all incorrect})$
  - $P(\text{not getting all incorrect})$ ...note this is not the same as  $P(\text{getting all correct})$
  - $P(\text{getting at least one correct})$ ...aha...this is the same as one of the above ☺
  - $P(\text{getting at least one incorrect})$ ...aha...this is the same as one of the above ☺
11. If the  $P(A \text{ and } B) = .4$  and the  $P(A) = .7$ , what is the  $P(B \text{ given } A)$ ?
12. If the  $P(A \text{ and } B) = .3$  and the  $P(A \text{ Given } B) = .6$ , which can you find, the  $P(A)$  or the  $P(B)$ ? Find it.
13. If  $P(A \text{ given } B) = 1/3$  and the  $P(B) = .3$ , what is the  $P(A \text{ and } B)$ ?
14. If  $P(A \text{ or } B) = .8$ ,  $P(B \text{ Given } A) = .3$  and  $P(A) = .5$ , what is the  $P(B)$ ?  
Hint: You cannot find directly – first find  $P(A \text{ and } B)$  from given, then find  $P(B)$  from  $P(A \text{ or } B)$  given. I absolutely LOVE this question.

## Chapter 6 –

1. What is the definition of a Random Variable? What two things will you need in order to DEFINE your random variable?
2. What is the difference between a discrete and a continuous RV? What are the two requirements for the probabilities of a discrete RV? Of a continuous RV?
3. What is a uniform probability distribution? How would the function be defined if the continuous RV can take on the values from 0 to n inclusive?
4. How do you determine (by formula) the mean, variance and standard deviation of a discrete RV? What is the difference between a mean of a RV and an expected value of a RV? [Note – recall that FINDING the mean and the variance of a continuous RV is beyond the scope of this class and requires methods from calculus.]
5. What does the law of large numbers state? Are there any assumptions about the observations from the population that are crucial to this law?
6. Complete the following rules for means and variances of RV's
  - i.  $\mu_{X+Y} =$
  - ii.  $\mu_{X-Y} =$
  - iii.  $\mu_{a+bX} =$
  - iv.  $\text{variance}(a + bX) =$
  - v.  $\text{variance}(X + Y) =$
  - vi.  $\text{variance}(X - Y) =$
  - vii.  $\text{standard deviation}(X - Y) =$
7. What are the four conditions that need to be satisfied to for a situation to be a binomial setting?
8. How is the RV X defined in a binomial distribution? What values can X take? What are the two parameters that you need in order to determine a probability such as  $P(X = 3)$ .
9. What is the difference between a pdf (probability distribution function) and a cdf (cumulative distribution function)? If X is a RV that is  $B(10, .3)$  give an example of how the two functions are related.

10. Explain what the calculator commands  $\text{binomlpdf}(50, .43, 23)$  and  $\text{binomlcdf}(50, .43, 23)$  would represent in words. You may put it into the context of a problem if you wish.
11. What is the formula (not calculator command) for determining the probability of  $(X = k)$  in a  $B(n,p)$ ?
12. How does a *geometric* random variable differ from a binomial random variable? How are they similar? What values can a geometric RV take?
13. What is mean and variance and standard deviation of a random variable that is  $B(n,p)$ ?
14. What is the mean of a geometric RV that had a probability of success =  $p$ ? [Note: the variance of a Geometric Dist is given by  $(1-p)/p^2$  if you care but is not given in the text]
15. If Mr. G has a 24% chance of making a free throw and shoots each independently confirm the following probabilities:
  - a.  $P(\text{It takes Mr. G Exactly 8 attempts to make his first}) = .035$
  - b.  $P(\text{It takes Mr. G more than 8 attempts to make his first shot}) = .111$
  - c.  $P(\text{It takes Mr. G less than 8 attempts to make his first shot}) = .853$

## Chapter 7

1. Describe what a sampling distribution represents.
2. How will the spread of the sampling distribution of  $\bar{x}$  change as you increase the sample size?
3. How will the spread of the sampling distribution of  $\hat{p}$  change as you increase the sample size?
4. How will the spread of the sampling distribution of  $\bar{x}$  change as you increase the number of samples?
5. What is the difference between a parameter and a statistic?
6. What does it mean for a statistic to be unbiased?
7. Describe in words what  $\hat{p}$  would represent.
8. What does the Central Limit Theorem say?
9. What does the Law of Large Numbers say? NOTE: These last two get confused, but of course you wouldn't let that happen to you.
10. Under what conditions can you assume that the sampling distribution of  $\bar{x}$  is  $N(\mu, \sigma/\sqrt{n})$ ?
11. Under what conditions can you assume that the sampling distribution of  $\hat{p}$  is  $N(p, \sqrt{p(1-p)/n})$ ?
12. Mr. G's Geometry class has grades that are normally distributed with a mean of 84.4% and a standard deviation of 3.3%.
  - a. What is the probability that a randomly chosen student will have a grade that is below 80%? What assumptions/conditions are

- necessary in order to ensure this probability is an accurate calculation?
- What is the probability that a randomly chosen group of 5 students will have a mean grade that is greater than 90%? What assumptions/conditions are necessary in order to ensure this probability is an accurate calculation?
  - What is the probability that a randomly chosen group of 30 students has a mean grade that is less than 82%?
  - Which of the answers above would change if the population of student's grades were distinctly NONNORMAL?

## Chapter 8

### CONFIDENCE INTERVALS

- What is meant by a critical value ( $z^*$ )?
- What are we assuming about  $\sigma$  (the standard deviation of the population) throughout this chapter?
- What is the formula for  $\sigma_{\bar{X}}$ , the standard deviation of the sampling distribution of  $\bar{X}$ ?
- What is the formula for the margin of error for the statistic  $\bar{X}$ ?
- How do you find a confidence interval from  $\bar{X}$  and your margin of error?
- What is the *interpretation* of a confidence interval? Yes, I am asking for those two magic sentences!
- You weigh 30 bags of M&M peanut bags and get a mean weight for those 30 bags to be 12.4 oz. You are told that the standard deviation for bag weights,  $\sigma$ , is known to be .3 oz<sup>2</sup>. We are interested in a 94% confidence interval. Find the following
  - $z^*$
  - $\sigma_{\bar{X}}$
  - margin of error
  - 94% confidence interval
  - What is the interpretation of this interval?
- Suppose you would like to cut the margin of error in 7c in half. How many total bags should you weigh in order to do this?
- How many bags should you weigh in 7 in order to ensure you have a margin of error no more than .03 oz at a confidence interval of 94%?
- How would you get the margin or error for a statistic from the confidence interval only?
- Name three cautions you have for someone who picks up a calculator and calculates a confidence interval. (see pages 487-488)



[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

[REDACTED]

Chapter 12 stuiff

Chapter 4 –

- ✓ What are the three main properties of logarithms?
- ✓ If you have a suspected (exponential) relationship of  $y = AB^x$ , how do you use the properties of logarithms to obtain a linear relationship? What graph would you suspect to be a straight line? That is, what would you plot on the dependent and independent axes? How do you then complete the inverse transformation to put the model back into the context of the problem?
- ✓ Same as above question but for power function.
- ✓ What is an initial test for exponential growth? How can you tell by looking at an equation if you have exponential growth or exponential decay?
- ✓ What is a marginal distribution? What is a conditional distribution? If you have a two-way table that measures “n” of one characteristic and “m” of another characteristic, how many marginal distributions will there be? How many conditional distributions will there be?
- ✓ Part 1 - What is extrapolation? Why should you care? Part 2 – What warnings do you have for someone performing linear regression on a data set of *averages*? Why do you make this warning?
- ✓ Describe Simpson’s paradox? What is the main lesson learned from Simpson’s in relation to analysis of statistical data?
- ✓ What is a lurking variable? A common response variable? A confounding variable? Which of the above is a subset of another?

