Name: ______________________________________ Period: __________ Date: __________

Statistics Review

MM4D1. Using simulation, students will develop the idea of the central limit theorem.
MM4D2. Using student-generated data from random samples of at least 30 members, students will determine the margin of error and confidence interval for a specified level of confidence.
MM4D3. Students will use confidence intervals and margin of error to make inferences from data about a population. Technology is used to evaluate confidence intervals, but students will be aware of the ideas involved.

1. Suppose we are interested in the average weight of chickens in America.

a) What would be the population?

b) What would be the variable of interest?

c) Find the 95% confidence interval for the average weight of chickens when a sample of 25 chickens produced an average weight of 4.6 pounds. Assume $\sigma = 2.5$. Answer with a complete sentence!

2. What is the critical value (z*) that corresponds to a confidence level of 90%?

3. Find a 99% confidence interval for the mean diameter of cabbage if we sampled 36 heads of cabbage from a physics class which had a sample average of 15 cm. Assume that the population standard deviation is 1.8 cm. Answer with a complete sentence!

a) What is the margin of error?

4. A study of the career paths of hotel managers sent questionnaires to an SRS of hotels belonging to major US hotel chains. There were 114 responses. The average response of these 114 general managers had spent with their current company was 11.78 years. Construct and interpret the 99% confidence interval for the mean number of years general managers of major-chain hotels have spent with their current company. Assume the standard deviation of this variable is 3.2 years.

a) How large would the sample need to be to ensure that the error does not exceed 0.5 years?

5. Mr. Cook is interested in knowing the average class size for all classes at WHS. Because each teacher has 6 classes this makes for quite a bit of data! To lighten his workload he decides to take a random sample of 10 classes and go from there. Use the following data:
33, 28, 24, 18, 24, 24, 26, 27, 31, 30

a) Identify the population.

b) Identify the variable of interest.

c) Construct a 90% confidence interval for the average class size at WHS. Assume that the standard deviation of class sizes is 3.2 students. Answer with a complete sentence.

d) How large of a sample would he need to take to reduce the error to plus or minus 1 student?

6. I want to know about the proportion of students at WHS who like school lunch. I randomly choose 5 homerooms and distribute a survey to everyone in each of those classes.

a) Identify the population.

b) What sampling technique was used?

c) After collecting the surveys, I see that 68 out of 97 students respond that they do NOT like school lunch. Make a 95% confidence interval for the proportion of students who like school lunch.

7. Find the ESTIMATE that was used to make the confidence interval (1.4, 3.6).

8. What is the confidence level for an interval whose margin of error is 2.965 when $\sigma = 14$ and $n = 121$.

9. A random sample of 56 candy corns produces an average weight of 3.8 grams. If $\sigma = 0.01$, find the 99% confidence interval for the weight of the candy corn.

10. A randomly distributed survey shows that 17 out of 63 people LOVE the color purple. Find the 90% confidence interval for the proportion of people who feel the same way.

11. I would like to learn more about raccoons. Specifically I am interested in knowing the average number of teeth they have. I take a sample of 32 raccoons and find that they have an average of 22 teeth. If I know that $\sigma = 4.2$ teeth. Answer the following questions:

a) What is the population?

b) What is the variable of interest?

c) Find the 90% confidence interval for the average number of teeth for all raccoons.

d) If I wanted to make my interval more precise, I can change the confidence level OR I can change the sample size. How many **additional** raccoons would I need in my sample to reduce my margin of error to $\pm$ 1 tooth?

12. I would now like to learn even more about raccoons. Specifically, I am interested in knowing the proportion of them that actually eat from trashcans. I distribute a survey to 46 randomly chosen raccoons asking them about their dining preferences and find that 31 of the 46 would NEVER eat from a trashcan. Answer the following questions:

a) Is my sample sufficiently large for the Central Limit Theorem to apply? What is "sufficiently large"?

b) Find the 95% confidence interval for the proportion of all raccoons that enjoy dining from trashcans.

c) If I want to make my interval more accurate (meaning that there is a smaller chance I will be wrong) should I *increase* or *decrease* my confidence level?

13. In 2004, ACT, Inc. reported that 363 of 505 randomly selected college freshmen in public colleges returned to college the next year. They also found that 853 of 1139 randomly selected college freshmen in private colleges returned to college the next year.

a) Estimate the national freshman-to-sophomore retention rate in **public** colleges with a 95% confidence interval.

b) Estimate the national freshman-to-sophomore retention rate in **private** colleges with a 95% confidence interval.

c) Choose the correct answer to the following statement: If we calculated 98% confidence intervals instead of 95% confidence intervals, the 98% CI would be
   i) narrower or ii) wider

d) If we wished to conduct our own survey of the freshman-to-sophomore retention rate p in public colleges, how many freshmen should be included in the sample if we want to estimate p to within 95% confidence? (Use 0.72 as the preliminary estimate of p).

e) Choose the correct interpretation of the 95% confidence interval in part b:
   i. In 95% of all random samples of private colleges, the freshman-to-sophomore retention rate will be 74.9%.
   ii. In 95% of all random samples of private colleges, the freshman-to-sophomore retention rate will be between 72.4 % and 77.4%.
   iii. There is a 95% chance that the interval (72.4%, 77.4%) contains the true freshman-to-sophomore retention rate for private colleges.
   iv. There is a 95% chance that the true freshman-to-sophomore retention rate of private colleges is in the interval (72.4%, 77.4%).
   v. If many random samples are selected, each sample with 1139 private college freshmen, 95% of the sample freshman-to-sophomore retention rates of p will be in the interval (0.724, 0.774).

14. Sociologists and sports psychologists have speculated that athletes who make it to the professional level in a sport where the older participants (that is, their birthdate is early in the "sport year") in the sport as a youth. In youth baseball leagues the "sport year" begins on August 1, so children in August are the older participants. In a random sample of 250 major league baseball players born since 1975, 35 were born in the month of August. National demographic statistics show that 9% of all births occur in August. Is there a higher percentage of professional baseball players born in August than in the general population? Use a 95% confidence interval to answer the question.

15. A study of the health of teenagers plans to measure the blood cholesterol level of an SRS of youths aged 13 to 16. The researchers will report the mean **x** from their sample as an estimate of the mean cholesterol level μ in this population.

a. Explain what it means to say that x is an unbiased estimator of μ.

b. The sample result of x is an unbiased estimator of the population mean μ no matter what size SRS the study chooses. Explain why a large sample gives more trustworthy results than a small sample.

16. Your mail-order company advertises that it ships 90% of its orders within three working days. You select an SRS of 100 of the 5000 orders received in the past week for an audit. The audit reveals that 86 of these orders were shipped on time.

a. What is the sample proportion of orders shipped on time?

b. If the company really ships 90% of its orders on time, what is the probability that the proportion in an SRS of 100 orders is as small as the proportion in your samples or smaller?

c. A critic says, "Aha! You claim 90%, but in your sample the on-time percent is lower than that. So the 90% claim is wrong." Explain why your probability calculation in *b* shows that the result of the sample does not refute the 90% claim.

17. The Gallup Poll once asked a random sample of 1540 adults, "Do you happen to jog?" Suppose that in fact 15% of all adults jog.

a. Find the mean and standard deviation of the proportion of the sample who jog (Assume the sample is an SRS).

b. Explain why you can use the formula for the standard deviation of **p** in this setting.

c. Check that you can use the Normal approximation for the distribution of **p**.

d. Find the probability that between 13% and 17% of the sample jog.

e. What sample size would be required to reduce the standard deviation of the sample proportion to one-third the value you found in *a*?

18. The Harvard College Alcohol Study finds that 67% of college students support efforts to "crack down on underage drinking." The study took a random sample of almost 15,000 students. The administration of a local college surveys an SRS of 100 students and finds that 62 support a crackdown on underage drinking.

a. What is the sample proportion who supports a crackdown on underage drinking?

b. If in fact the proportion of all students attending this college who support a crackdown is the same as the national 67%, what is the probability that the proportion in a SRS of 100 students is as small as or smaller than the result of the administration's sample?

c. A writer in the college's student paper says that "support for a crackdown is lower at our school than nationally." Why does the survey not support his conclusion?

19. According to a market research firm 52% of all residential telephone numbers in LA are unlisted. A telephone sales firm uses random digit dialing equipment that dials residential numbers at random, whether or not they are listed in the telephone directory. The firm calls 500 numbers.

a. What are the mean and standard deviation of the proportion of unlisted numbers in the sample?

b. What is the probability that at least half the numbers dialed are unlisted?

20. A bottling company uses a filling machine to fill plastic bottles with cola. The bottles are supposed to contain 300 ml. In fact, the contents vary according to a Normal distribution with μ = 298 ml and σ = 3ml.

a. What is the probability that an individual bottle contains less than 295 ml?

b. What is the probability that the mean contents of the bottles in a six-pack is less than 295 ml?

21. The idea of insurance is that we all face risks that are unlikely but carry high cost. Think of a fire destroying your home. So we form a group to share the risk: we all pay a small amount and the insurance policy pays a large amount to those few of us whose homes burn down. An insurance company sees that in the entire population of homeowners the mean loss from fire is $250 and the standard deviation is $300.

a. What type of distribution would this represent?

b. If the company sells 10,000 policies can we use the CLT to create an approximately normal distribution? Why/why not?

c. What is the approximate probability that the average loss of the 10,000 policies will be greater than $260?

22. The number of traffic accidents per week at an intersection during a year varies with mean 2.2 and standard deviation 1.4. The number of accidents in a week must be a whole number, so the population distribution is not Normal.

a. Let **x** be the mean number of accidents per week at the intersection during a year (52 weeks). What is the approximate distribution of **x** according to the central limit theorem?

b. What is the approximated probability that **x** is less than 2?

c. What is the approximate probability that there are fewer than 100 accidents at the intersection in a year? (Hint: Restate this event in terms of **x**).

23. The Wechsler Adult Intelligence Scale (WAIS) is a common "IQ test" for adults. The distribution of WAIS scores for persons over 16 years of age is approximately Normal with mean 100 and standard deviation 15.

a. What is the probability that a randomly chosen individual has a WAIS score of 105 or higher?

b. What are the mean and standard deviation of the sampling distribution of the average WAIS score for an SRS of 60 people?

c. What is the probability that the average WAIS score of an SRS of 60 people is 105 or higher?

d. Would your answers to any of *a*, *b*, or *c* be affected if the distribution of WAIS scores in the adult population were distinctly non-Normal?