Midterm II

BIOS:4120 **Score** ______ / 100

Introduction to Biostatistics

Knute Carter

Name (Last, First) ________________________________

Teaching Assistant ________________________________ **Lab Section** ____________

1. (24 points total) For females between 18 and 24 years of age, systolic blood pressures are normally distributed with a mean of 114 mm Hg and a standard deviation of 12 mm Hg. Thus, if the random variable X is defined as the systolic blood pressure of a randomly selected 18-24 year old female, then X is a normal random variable with a mean of 114 mm Hg and a standard deviation of 12 mm Hg.

   a) (8 points) Hypertension is defined as a systolic blood pressure over 140 mm Hg or a diastolic blood pressure over 90 mm Hg. Find the probability of a randomly selected 18-24 year old female being hypertensive based on her systolic blood pressure exceeding 140 mm Hg. *(Show work!)*

   For parts (b) through (d), consider a random sample of 16 women between 18 and 24. Let $\bar{X}$ represent the average systolic blood pressure of these 16 women.

   b) (3 points) What is the mean of $\bar{X}$? The standard deviation (or standard error) of $\bar{X}$?

c) (3 points) Sketch a plot that illustrates the general shape of the sampling distribution curve for $\bar{X}$. Label the points 108, 114, and 120 along the horizontal axis of the plot. Specify the percentage of the total area under the sampling distribution curve that lies between 108 and 120.

d) (10 points) Find the probability that $\bar{X}$ will assume a value less than 116 mm Hg. (*Show work!*)

2. (24 points total) A dietician is interested in studying the mean daily caloric intake of Big Ten male collegiate athletes in two sports: football and basketball. To collect data for her study, she enlists the participation of a random sample of 61 football players and 7 basketball players. She asks each athlete to keep a dietary diary over a month during his season, from which she determines a daily caloric intake value for each player. Assume that similar studies have led the dietician to believe that such intake values are normally distributed within each of the player populations.

a) (10 points) For the sample of football players, the dietician obtains a mean of 3020 calories with a standard deviation of 440 calories. Use these statistics to construct a 90% two-sided confidence interval for the mean daily caloric intake of Big Ten football players. Interpret this interval in the context of the problem. (*Show work!*)

b) (8 points) For the sample of basketball players, the dietician obtains a mean of 2480 calories with a standard deviation of 240 calories. Use these statistics to construct a 90% upper confidence bound for the mean daily caloric intake of Big Ten basketball players. Interpret this interval in the context of the problem. (*Show work!*)

c) (3 points) Based on the results of (a) and (b), can the dietician safely assert that the mean dietary caloric intake of Big Ten football players exceeds that of Big Ten basketball players? Why or why not?

d) (3 points) In this application, it is assumed that normality holds in each of the player populations. Is this assumption crucial for the result of part (a)? Is this assumption crucial for the result of part (b)? Briefly explain.

3. (10 points total) Quitline Iowa is a statewide smoking cessation telephone counseling hotline. For staffing purposes, administrators are interested in characterizing the likelihood of obtaining a certain number of phone calls during an hourly period. Based on records of past hotline activity, it is known that the average number of phone calls received in an hour is 2.2. What is the probability of obtaining 2 or 3 phone calls during a specific hourly period? (*Show work!*)

4. (18 points total) Tay-Sachs disease (abbreviated TSD) is a genetic disorder common in Eastern European people of Jewish descent (Ashkenazi Jews). The disease occurs when harmful quantities of a fatty acid derivative accumulate in the nerve cells of the brain.

   In a certain Jewish population, suppose that 1 in 18 infants are born with TSD. Consider a sample of 20 infants randomly selected from this Jewish population.

   a) (6 points) What is the probability exactly two of these 20 infants will have TSD? (*Show work!*)

   b) (12 points) What is the probability at least two of these 20 infants will have TSD? (*Show work!*)

5. (24 points total) Multiple choice. Choose one answer per question. (3 points each)

   (i) Suppose that three statisticians - John, Paul, and George - all plan to construct two-sided confidence intervals for an unknown population mean μ based on the same set of data. John plans to use a 99% confidence level, Paul plans to use a 95% level, and George plans to use a 90% level. Who will have the widest interval?

      (a) John
      (b) Paul
      (c) George
      (d) All intervals should be of roughly the same width, since all intervals are being based on the same set of data.
      (e) It is not possible to answer the question without having the numerical results for the three intervals.

   (ii) Suppose that three statisticians - Jimmy, Robert, and Bonzo - all plan to construct 95% two-sided confidence intervals for an unknown population mean μ based on independently drawn samples. Jimmy plans to use a sample size of 50, Robert plans to use a sample size of 500, and Bonzo plans to use a sample size of 5,000. Who should have the widest interval?

      (a) Jimmy
      (b) Robert
      (c) Bonzo
      (d) All intervals should be of roughly the same width since all are being constructed using the same confidence level.
      (e) It is not possible to answer the question; each of the three intervals has the same chance of being the widest.

   (iii) Let Z be a standard normal random variable. Which of the following statements is **false**?

      (a) $P(Z > 0) = 0.5$
      (b) $P(Z > z_{0.05}) = 0.05$
      (c) $P(-z_{0.10} < Z < +z_{0.10}) = 0.90$
      (d) $P(Z \leq -z_{0.05}) = 0.05$.
      (e) $P(-5.00 < Z < 8.00) \approx 1.00$.

(iv) Which of the following is a **false** statement in regard to the margin of error?

(a) When estimating a population mean μ based on a large sample size, the margin of error is defined as ME $= z_{\alpha/2}(\sigma/\sqrt{n})$ (or ME $= z_{\alpha/2}(s/\sqrt{n})$ if σ is unknown).

(b) When estimating a population mean μ, the margin of error represents the width of a two-sided confidence interval for μ.

(c) In the media, the margin of error is generally evaluated using a 95% confidence level.

(d) The general form of a two-sided confidence interval is as follows:
point estimate ± margin of error

(v) Let X = the pregnancy duration for a randomly selected expectant adult female. Medical records indicate that X is a normal random variable with a mean of $\mu = 266$ days and a standard deviation of $\sigma = 16$ days. What is the probability of X assuming a value greater than 250 days?

(a) 0.841

(b) 0.159

(c) 1.000

(d) 0.682

(e) None of the above.

(vi) Which of the following is **not** a criterion for a Poisson process?

(a) An event occurs periodically over time (or space).

(b) The probability of an occurrence of the event in an interval of time (or space) is proportional to the length of the interval.

(c) Within a single interval of time (or space), only a finite number of occurrences of the event are possible.

(d) The events occur independently, both within the same interval and between consecutive intervals.

(vii) In a medical study, a physician wishes to estimate the mean systolic blood pressure of a cohort of adult patients who are moderately hypertensive. She wishes to find the sample size needed to ensure that the width of a 95% confidence level does not exceed 5 mm Hg. Based on a pilot study, she estimates the standard deviation of her target population to be 20 mm Hg. What sample size should the physician use?

(a) 62

(b) 44

(c) 246

(d) 174

(e) 16

(viii) Suppose a sample of size n is randomly drawn from a population in which the underlying data has a mean of μ, a standard deviation of σ, and a highly left-skewed distribution. Which of the following is a valid statement regarding the sample mean $\bar{X}$?

(a) The mean or expected value of $\bar{X}$ is μ.

(b) The standerd deviation or standard error of $\bar{X}$ is $\sigma/\sqrt{n}$.

(c) If the sample size is large, the sampling distribution of $\bar{X}$ will be close to a normal distribution.

(d) If the sample size is large, the sampling distribution of

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

will be close to a standard normal distribution.

(e) The sample data will be roughly normally distributed.

(f) The population data will be roughly normally distributed.

(g) Each of (a), (b), (c) and (d) are valid.

(h) Each of (a), (b), (c), (d), (e) and (f) are valid.