CLARK UNIVERSITY

Name: ____________________________

Math 218 Mathematical Statistics

First Test

February 2016

You may refer to one sheet of notes on this test, and you may use a calculator. You may leave your answers as expressions such as $\binom{8}{4}\frac{e^{1/3}}{\sqrt{2\pi}}$ if you like.

1. [15; 3 points each part] Data variables come in various types. Some are categorical (also called qualitative) while others are numerical. Those that are qualitative are either nominal or ordinal. Those that are numerical are either continuous or discrete. For each of the following examples, decide whether it is nominal, ordinal, continuous, or discrete.

_______ **a.** Grade of meat: prime, choice, good.

_______ **b.** Birthweight of newborns in grams.

_______ **c.** Car model (e.g. Ford Escort, or Mitsubishi Galant).

_______ **d.** Time until a lightbulb burns out.

_______ **e.** Days absent from class.

2. [16; 4 points each part] A study was done to evaluate the benefits of an exercise program to reduce mortality among patients who have survived a heart attack. Mortality among volunteers who enrolled in an exercise program was compared with that among controls selected from medical records of eligible patients who chose not to join the exercise probram. Reasons given for not exercising included physician disapproval, shift work, and lack of interest. The study results showed that the patients in the exercise program had half as many deaths as those in the control group.

a. What is the response variable in this study?

b. What is the explanatory variable in this study?

c. Name at least one confounding variable.

d. Use the statement above to give at least one explanation why the chosen control group is not appropriate and how it could influence study results.

3. [15; 3 points each part] True or false. Write the whole word "true" or the whole word "false."

______ **a.** A normal plot of data indicates that the data is normally distributed if the plot approximates a straight line.

______ **b.** Whereas treatment factors are controlled in an experiment, nuisance factors, also called noise factors, are all the other factors that might affect the response variable.

______ **c.** Student's T-distribution is the primary distribution used to study Poisson processes.

______ **d.** To standardize a random variable, subtract its mean and divide by its standard deviation, $Y = (X - \mu_X)/\sigma_X$, for then $\mu_Y = 0$ and $\sigma_Y = 1$.

______ **e.** Simple random sampling is where the population can be divided into homogeneous subpopulations and a small sample is drawn from each subpopulation resulting in a sample that is representative of the population.

4. [20; 5 points each part] We have used the term "variance" in several ways.

a. Define the variance σ^2 of a random variable X.

b. Define the variance S^2 of a sample $X_1, X_2, \ldots, X_n$.

c. How is S^2 related to σ^2?

d. How is the variance $Var(\overline{X})$ of the sample mean related to either σ^2 or to S^2?

5. [20; 4 points each part] On maximum likelihood functions. Consider the family of exponential distributions parametrized by θ, a positive real number. The density function is

$$f(x \mid \theta) = \theta e^{-\theta x} \qquad \text{for } x \text{ positive.}$$

a. What is the joint density function $f(x_1, x_2, \ldots, x_n | \theta)$ for a random sample $X_1, X_2, \ldots, X_n$ from that distribution? (Assume all the x_i's are positive.)

b. Write down the likelihood function $L(\theta | x_1, x_2, \ldots, x_n)$?

c. Express the log of the likelihood function, and simplify it

d. Take $\frac{d}{d\theta}$ of the log of the likelihood function.

e. Use the work you've done so far to determine is the maximum likelihood estimator $\hat{\theta}$ for θ, that is, find the value of θ maximizes this likelihood for given $x_1, x_2, \ldots, x_n$?

6. [20] On estimators. Recall the geometric distribution with parameter p. Recall that when independent Bernoulli trials are repeated, each with probability p of success, then the number of trials X it takes to get the first success has a geometric distribution. The probability mass function for the geometric distribution is $f(x) = q^{x-1}p$, for $x = 1, 2, \ldots$. The mean of this geometric distribution is $\mu = 1/p$, and its variance is $\sigma^2 = (1-p)/p^2$.

a. [4] Given a random sample $X_1, X_2, \ldots, X_n$ from a geometric distribution with an unknown parameter p, let our estimator $\hat{\mu}$ for μ be the sample mean $\hat{\mu} = \overline{x}$. Determine the expectation $E(\hat{\mu})$ of this estimator.

b. [4] Is this $\hat{\mu}$ an unbiased estimator for μ?

c. [4] Determine the variance $Var(\hat{\mu})$ of this estimator.

d. [8] Determine the mean squared error MSE$(\hat{\mu})$ of this estimator.

#1.[15]	
#2.[16]	
#3.[15]	
#4.[20]	
#5.[20]	
#6.[20]	
Total	