STATISTICS 641 - EXAM I

Student's Name ____

Student's Email Address ____

INSTRUCTIONS FOR STUDENTS:

(1) The exam consists of 6 pages including this cover page and 13 pages of Tables.

(2) You have exactly **60 minutes** to complete the exam.

(3) Show *ALL* your work on the exam pages.

(4) Do not discuss or provide information to anyone concerning the questions on this exam or your solutions until I post the solutions to the exam.

(5) You may use the following:

- Calculator - Your device cannot facilitate a connection to the internet or to send text messages
- Summary Sheets - (**2-pages,** 8.5"**x**11"**, write on both sides of the two sheets**)
- The attached materials.

(6) Do not use any other written material except for your summary sheets and the attachments to the exam.

(7) Do not use a computer, cell phone, or any other electronic device (other than a calculator).

I attest that I spent no more than 60 minutes to complete the exam. I used only the materials described above. I did not receive assistance from anyone during the taking of this exam.

Student's Signature________________________________

Problem I. (60 points) CIRCLE ONE of (A, B, C, D, or E) corresponding to the **BEST** answer to the question. You **Do Not** have to provide justification for your answer.

However, for those problems involving calculations, show your work on the test and I will give partial credit.

(1.) The Government Accountability Office (GAO) was investigating the problems associated with FEMA's handling of the destruction caused by the many hurricanes in 2017. A random sample of twenty counties was taken from each of the five states impacted by a hurricane in 2017. A polling agency was hired to select a random sample of 150 residents in each of the 100 counties. From their responses to prepared questions, a measure of displeasure with FEMA was computed for each of the selected residents. These measures were then summarized into an overall average level of displeasure for the public as a whole. This type of study is an example of

A. a simple random sample.

B. a simple cluster sample.

C. a stratified simple random sample.

D. a stratified multistage cluster random sample.

E. a multistage cluster random sample.

(2.) A cardiologist has determined that the number of failures in heart pacemakers can be modeled as a Poisson process with an average rate of 0.4 failures per year. She wants to assess the possible number of failures, F, that may occur in a 5 year period after the pacemaker is placed in a patient. Let a randomly generated observation from a Uniform on (0,1) distribution be $U = 0.682$. The corresponding random observation from F is

A. 0

B. 1

C. 2

D. 3

E. 4

(3.) In order for a beam to be certified as a bridge support beam, its tensile strength must exceed 2500 units. A metallurgist examines 30 randomly selected beams and records $M_1,\ M_2,\ \ldots,\ M_{30}$, the tensile strength of the 30 beams. She then determines N, the number of beams out of the 30 beams which met the 2500 specification. The distribution which best represents the distribution for N is

A. binomial distribution

B. hypergeometric distribution

C. negative binomial distribution

D. Poisson distribution

E. Weibull distribution

(4.) A process engineer is studying the increase in the proportion of defective parts, D, after power outages in the production facility. The engineer determines that the cdf of D is given by

$$F(d) = \begin{cases} 0 & \text{if} \quad d \le 0 \\ 8d & \text{if} \quad 0 < d < 0.1 \\ 0.8 & \text{if} \quad 0.1 \le d < 0.2 \\ 0.75 + 0.25d & \text{if} \quad 0.2 \le d < 1 \\ 1 & \text{if} \quad d \ge 1 \end{cases}$$

Provide the engineer with a constant τ such that approximately 5% of all power outages will have a defective increase greater than τ.

A. $\tau = 0.8$

B. $\tau = 0.11875$

C. $0.1 \le \tau < 0.2$

D. $\tau = 0.00625$

E. cannot be determined with the given information

Use the following information for Questions 5 and 6:

A researcher is studying the time to failure of knee replacements for those individuals who had run a large number of 10K races. The researcher randomly selects 250 people who had run at least twenty 10K races over the past 10 years. Based on this data, she determines that a distribution with pdf

$$f(t) = \frac{\theta_2}{\theta_1^{\theta_2}} (t - \theta_3)^{\theta_2 - 1} e^{-((t-\theta_3)/\theta_1)^{\theta_2}} \quad \text{for} \quad t \ge \theta_3 \qquad \text{and} \qquad f(t) = 0 \quad \text{for} \quad t < \theta_3$$

(5.) The value of the pdf at $t = .25$, $f(.25)$, can be best described as

A. the proportion of knee replacements that will fail prior to the lower quartile of all knee replacements

B. the probability that a randomly selected runner will have a knee replacement failure prior to .25 years

C. the proportion of knee replacements that will fail prior to the 25

D. the probability that a randomly selected knee replacement will have completed 25

E. None of the above

(6.) The parameters, $(\theta_1,\ \theta_2,\ \theta_3)$, in the pdf, $f(t)$, given above are

A. $(\theta_1,\ \theta_2,\ \theta_3) =$ (shape, scale, location) parameters

B. $(\theta_1,\ \theta_2,\ \theta_3) =$ (scale, shape, location) parameters

C. $(\theta_1,\ \theta_2,\ \theta_3) =$ (location, scale, shape) parameters

D. $(\theta_1,\ \theta_2,\ \theta_3) =$ (shape, scale, range) parameters

E. $(\theta_1,\ \theta_2,\ \theta_3) =$ (scale, shape, range) parameters

(7.) Let Y have a pdf f which is a member of a location-scale family with location parameter θ_1 and scale parameter θ_2. It can then be concluded that

A. $P[|Y - \theta_1| \leq k\theta_2] \ \geq \ 1 - \frac{1}{k^2}$

B. the moments of the distribution of Y, μ_k, are the same value for all choices of θ_1 and θ_2

C. $P[|Y - \theta_1| \leq k\theta_2]$ is the same value for all choices of θ_1 and θ_2

D. the quantiles of the distribution of Y, $Q(u)$, are the same value for all choices of θ_1 and θ_2

E. none of the above statements are true

(8.) When using a kernel density estimator to estimate a continuous pdf, the selection of the bandwidth

A. is a compromise between smoothing enough to remove insignificant bumps and not smoothing too much to remove real peaks in the estimator.

B. has very little impact on the estimator.

C. has very little impact on the estimator provided a Gaussian kernel is used.

D. affects the bias of the estimator but not its variance.

E. affects the variance of the estimator but not its bias.

(9.) Two procedures for estimating unknown parameters in a family of distributions are maximum likelihood (MLE) and method of moments (MOM). MLEs are generally preferred over MOMs because

A. MLEs involve more complex equations and hence must be an improvement over MOMs.

B. MOMs are based on just the moments of the distribution whereas MLEs are based on the likelihood function.

C. The computation of MLEs tend to be easier than the computations for MOMs.

D. The computation of MOMs tend to be easier than the computations for MLEs.

E. All of the above are good reasons.

(10.) The pair (μ, σ) is preferred to $(Median, MAD)$ as a measure of a population's location-scale parameters when the population distribution

A. has absolutely no outliers.

B. has normal-like tails.

C. has a right skewed distribution.

D. has a heavy-tailed distribution.

E. has a symmetric distribution.

(11.) In many applied journal articles, the data is summarized by providing only the sample mean and standard deviation. It has been suggested that the sample estimates of the four parameters: mean, standard deviation, skewness, and kurtosis, would provide a relatively complete description of the population distribution function, F. Which one of the following statements is TRUE?

A. If skewness=0 and kurtosis=3, then F is similar in shape to a normal distribution.

B. If skewness=0 and kurtosis=23, then F has considerably more probability mass in its tails than a normal distribution.

C. If skewness=0 and kurtosis=.3, then the shape of F is bimodal with a low concentration about $\mu \pm \sigma$.

D. Distributions of widely different shapes may have the same values for mean, standard deviation, skewness, and kurtosis.

E. All of the above statements are False.

(12.) A toxicologist conducts an experiment to study the toxicity of a new insecticide. He randomly selects 100 lab raised rats who had no prior exposure to the insecticide and injects a measured amount of the insecticide in the 100 rats. The amount of insecticide, I, in the rat's bladder is known to peak at 5 days after exposure and then decrease slowly over the remaining 25 days in the study. The value of I is recorded 30 days post injection. Several of the rats die prior to 30 days but after the 5 day peak. The values of I on these rats are taken at the time of their death. We would describe the data from this study as being

A. Type I censoring

B. Type II censoring

C. Random censoring

D. Right censoring

E. Left censoring

Problem II (40 points) An electronics firm is testing the insulation used on its electrical cable. A study was conducted to determine the voltage level, V, at which the insulation failed. The study consisted of 20 specimens and the failure voltages obtained are as follows: 32.0, 35.4, 36.2, 39.8, 41.2, 43.3, 45.5, 46.0, 46.2, 46.4, 46.5, 46.8, 47.3, 47.4, 47.6, 49.2, 50.4, 50.9, 52.4, 56.3

(A.) The following SAS output produced the MLE's for the parameters, γ and α, from a Weibull distribution:

$$f(y) = \frac{\gamma}{\alpha^\gamma} y^{\gamma-1} e^{-\left(\frac{y}{\alpha}\right)^\gamma} \quad \text{for} \quad y \geq 0 \qquad\qquad F(y) = 1 - e^{-\left(\frac{y}{\alpha}\right)^\gamma} \quad \text{for} \quad y \geq 0$$

```
                         Standard   95% Confidence      Chi-
Parameter      DF Estimate    Error       Limits        Square    Pr > ChiSq
Intercept       1   3.8667   0.0251   3.8176   3.9159   23792.0     <.0001
Scale           1   0.1065   0.0184   0.0759   0.1495
Weibull Scale   1  47.7866   1.1979  45.4954  50.1931
Weibull Shape   1   9.3863   1.6224   6.6891  13.1712
```

Compute the MLE estimate of the probability that a cable will have a failure voltage greater than 50

(B.) The researcher was uncertain about using a Weibull distribution in the analysis and decides to use a distribution-free method of estimating the survival function. The Kaplan-Meier estimator of the survival function, $\widehat{S}(v)$, is displayed in the output given below.

v	0.0	32.0	35.4	36.2	39.8	41.2	43.3	45.5	46.0	46.2	46.4
$\widehat{S}(v)$	1.00	0.95	0.90	0.85	0.80	0.75	0.70	0.65	0.60	0.55	0.50
v	46.5	46.8	47.3	47.4	47.6	49.2	50.4	50.9	52.4	56.3	
$\widehat{S}(v)$	0.45	0.40	0.35	0.30	0.25	0.20	0.15	0.10	0.05	0	

Compute both the distribution-free (DF) estimate and the MLE estimate of the voltage, V_{cr}, for which approximately 28% of cables will have a failure voltage less than V_{cr}