Statistics 533 Midterm Exam Name ____________________

Spring 2000 **You must show all of your work**

When asked to explain something, provide an explanation that could be understood by someone who does not have formal training in statistical methods. Your explanations should be clear, but concise. You do not need to do any complicated calculations. Set things up and show me that you know how to do the calculations. Leave requested numerical answers as fractions.

1. A random sample of $n = 50$ motors were put on test at a particular point in time and the test was run until $t_c = 2000$ hours at which point $r = 5$ units had been reported as failing. The reported failure times were 14, 88, 132, 165, 409 hours. Because of continuous monitoring, the reported failure times were close to exact.

   (a) Find a nonparametric estimate of $F(t)$, the cdf for the motors. Plot the estimate in an appropriate manner on the attached Weibull probability paper.

   (b) Find a large-sample approximate 95% nonparametric confidence interval for $F(500)$. There are several alternative methods. Just do something simple, not worrying about the adequacy of the approximation in this particular application.

   (c) Comment on the shape or other characteristics of the nonparametric estimate of $F(t)$ for the motors. Do you think that a Weibull distribution could be used to adequately describe the failure time distribution of these motors? Based on what you see in the data, and your other knowledge, what comments would you make about the failure time distribution of these motors and possible causes of failure?

2. Explain the proper interpretation for the confidence level associated with confidence intervals, like that computed in part 1b above.

3. Consider the failure-time distribution with cdf

$$F(t) = 1 - \exp\left[-\left(\frac{t-\mu}{\alpha}\right)^2\right], \quad t > \mu. \tag{1}$$

(a) Derive an expression of the p quantile of this distribution.

(b) Find an appropriate transformation that will cause the cdf to plot as a straight line and explain how you could use this transformation to make a probabilty plot for this distribution.

(c) Derive an expression for the hazard function corresponding to the cdf in equation (1). How would you describe the shape of this hazard function? Be as specific as possible.

(d) Write down an expression for the log likelihood for a sample of n observations that are reported as exact failures (i.e., use the density approximation), based on the model expressed in equation (1).

(e) Derive a simple expression for the ML estimator of α in equation (1), assuming that the value of μ is given.

4. Chapter 6 of Meeker and Escobar (1998) shows how to linearize a cdf. Why is this linearization useful? Why not just plot the usual "S-shaped" cdf instead?

5. Explain why it is that the observed information matrix is more commonly used than the Fisher information matrix when it is necessary to obtain an estimate of the variance-covariance matrix of ML estimates when analysing censored data.

6. A company keeps close tabs on warranty returns, monitoring for changes in reliability of product in the field. The equation

$$L(\theta) = \prod_{i=1}^{n} [F(t_i + .5; \boldsymbol{\theta}) - F(t_i - .5; \boldsymbol{\theta})]$$

gives the "correct likelihood" assuming that reported failure times t_i were obtained by rounding to the nearest integer value of weeks of service (i.e., if 29 is reported, we know that the failure occurred between 28.5 and 29.5 weeks of service).

   (a) How should the "correct likelihood" be written if the reported value of 29 indicates instead that the failure occurred in the 29th week (note that recorded time is discrete)? Draw a simple picture to illustrate.

   (b) For which recording method is using the simple density approximation $f(29; \boldsymbol{\theta})$ more accurate? Explain.

   (c) Suppose that $f(t)$ is very easy to compute relative to $F(t)$. Suggest a "refined" density approximation that might be used in such cases where the simple density approximation should not be used. Draw a simple picture to illustrate.

7. Refer to Figure 10.6 from Meeker and Escobar (a copy is attached). A life test is planned to run for 1000 hours and the responsible engineers feel that life can be described by a lognormal distribution with $\sigma^{\square} = .5$ with approximately 7% failing at 1000 hours cycles. Management has asked for an estimate and a 95% confidence interval for $t_{.1}$ for this distribution. In planning the test, it is thought that the sample size should be large enough such that the confidence interval upper endpoint should be approximately 50% larger than the ML estimate of $t_{.1}$.

   (a) Provide an expression for the planning value for $t_{.1}$ as a function of the information given above?

   (b) Provide an expression for $\text{Avar}(\widehat{t}_{.1})$, as a function of the elements of the approximate large sample covariance matrix for the ML estimates $\widehat{\mu}$ and $\widehat{\sigma}$.

   (c) What is the approximate sample size that is needed to estimate $t_{.1}$ for this distribution, according to the specified criterion, assuming that both μ and σ will be estimated using the method of maximum likelihood?

   (d) What could you recommend to the responsible engineer who was concerned about the large number of units needed for testing. Make a list of suggestions.

8. Consider the following system diagram. Time to failure for block (component) i has a cdf $F_i = F_i(t)$ and the times to failure of the individual blocks are independent. Provide an expression for the cdf of the system as a function of F_1, F_2, F_3, and F_4.