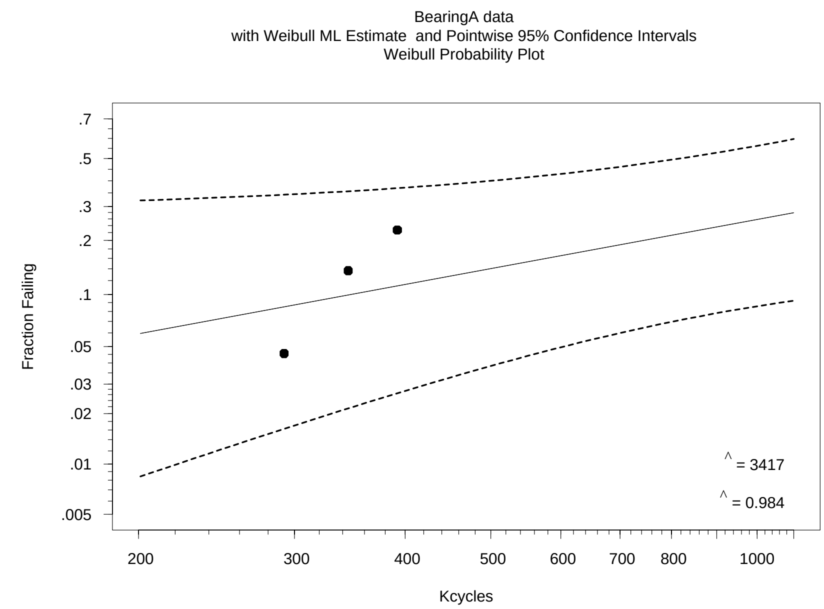
Statistics 533 Midterm Exam Name ________________________

Spring 2004 **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. Twelve motors were put on test at the same time. One of the motors was removed from test after 200 kcycles (kcycles is thousands of cycles). Failures were observed at 292, 345, 392 kcycles and the remaining eight units ran until 1100 kcycles, when the test ended. The following figure is a Weibull probability plot of the data. The line is the ML estimate of the Weibull $F(t)$. An approximate 95% confidence interval for the Weibull shape parameter β is $[0.3456, \quad 2.801]$.



(a) Compute a nonparametric estimate of the $F(t)$ from these data.

(b) Explain how the probability plotting positions in the figure were computed from the nonparametric estimate of the $F(t)$.

(c) Why is it that the ML line does not go through the points in the probability plot?

(d) The Weibull distribution does not fit these data very well. In light of this, but based on the existing data, what can you say about the hazard function of the motors?

2. Explain reasons why a likelihood-based confidence interval would be expected to provide a better approximate procedure than a normal (Wald) approximation method.

3. In a life test with Type II censoring (also known as "failure censoring"), units are put on test and run until a pre-specified number r of the units fail. Let $\widehat{t}_p$ denote the ML estimate of the p quantile of a life distribution. Assuming a log-location-scale distribution), the quantity

$$Z_{\log(\widehat{t}_p)} = \frac{\log(\widehat{t}_p) - \log(t_p)}{\widehat{\text{se}}_{\log(\widehat{t}_p)}}. \quad (1)$$

is "pivotal."

(a) Briefly explain what it means when we say that a quantity like that in (**??**) is "pivotal."

(b) Briefly explain why it is useful for such a quantity to be pivotal.

(c) Show how the knowledge of the distribution of (**??**) can be used to define a procedure to compute a confidence interval for t_p.

4. Suppose that failure time T of a unit selected at random from a product population has the cdf

$$\Pr(T \leq t) = F(t) = \exp\left[-\left(\frac{\alpha}{t-\mu}\right)\right], \quad t > \mu, \quad \alpha > 0. \tag{2}$$

(a) Derive an expression for the quantile of the distribution of T.

(b) Is α a location parameter, a scale parameter, or a shape parameter? Explain why.

(c) Given a set of multiply right-censored data, list the steps that you would use to construct some kind of plot to assess the adequacy of this distribution as a model for the data.

(d) Derive an expression for the hazard function corresponding to the cdf in equation (**??**).

(e) How would you describe the shape of this hazard function? What does this suggest about the cause of failures in the population.

(f) Write down an expression for the log likelihood for a sample of n observations that are reported as exact failures (i.e., there is no censoring and you can use the density approximation for the individual log likelihood terms), based on the model expressed in equation (**??**).

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

5. Suppose that you have an expression for $\mathrm{Var}[\widehat{\theta}]$, where $\widehat{\theta}$ is the ML estimator of a scalar parameter θ.

   (a) Use the delta method to get an expression for the approximate (large-sample) variance of $\sqrt{(\widehat{\theta})}$.

   (b) Draw a picture and explain briefly why this approximation tends to be better when the sample size is larger.