Statistics 533 Final Examination 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. Each question or part of a question is worth 5 points for a total of 100 points.

1. Failure is caused by a chemical reaction, the rate of which can be described by the Arrhenius relationship. Suppose that time to failure $T \sim \text{LOGNOR}(\mu, \sigma)$ at an given level of temperature.

   (a) Derive an expression for the time acceleration factor relating life at two different levels of temperature.

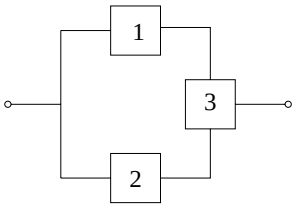
   (b) Show why it is that the simple linear regression relating log life and reciprocal Kelvin temperature, implied by the Arrhenius rate reaction model, will have constant σ.

2. The manufacturer of computer hard disks reports in its promotional literature a "MTBF" figure. This figure is obtained by taking a sample of units from each days production, putting the units on test for a period t_c (perhaps one week), the pooling the available data over many months, and computing

$$\text{MTBF} = \frac{TTT}{r}$$

   where TTT is the total time on test of all units that were tested and r is the observed number of failures (typically a very small number of drives). Reported figures are typically numbers like MTBF=100 years. Comment on the usefulness and validity of the use of this figure for characterizing the reliability of disk drives.

3. Consider the following system diagram.



(a) Derive an expression for the cdf $F_T(t)$ of the system as a function of F_1, F_2, and F_3, the cdfs for the individual components. Assume independence of the failure times of the individual components.

(b) Suppose that each component follows a Weibull distribution with its own scale and shape parameters and that these have been estimated from separate studies, providing ML estimates of these parameters and the corresponding covariance matrix. These estimates could be used with the result in part 3a to obtain the ML estimate of $F_T(t)$. Give a simple high-level formula that shows how one could obtain an estimate of the standard error for this estimate.

(c) Suppose instead, that the failure times of components 1 and 2 are positively correlated but independent of component 3, with marginal distributions F_1, F_2, and F_3. If one incorrectly used the independence assumption to compute the cdf $F_T(t)$ of the system would answer be too big or too small? Draw a picture to give an intuitive explanation for why.

(d) Give an example of a system in which one might expect component life times to be positively correlated.

4. Suppose that you have an expression for $\text{Var}(\widehat{\theta})$, 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 $\log(\widehat{\theta})$.

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

5. Consider the use of a Weibull distribution to describe a time-to-failure process.

   (a) Explain why doing a Bayesian analysis with a specified prior distribution on the unknown parameters μ and σ of a Weibull distribution is not a model that implies that μ and σ vary from unit to unit in the population.

   (b) Explain how you could set up a model that allows μ and σ vary from unit to unit in the population.

6. Maximum likelihood is the standard method of fitting a distribution to censored data. An alternative that has been suggested is to use ordinary least squares to fit a line to the points on a probability plot. Give two reasons why this is a bad thing to do, in general.

7. Probability plots are important tools for helping to find a suitable parametric distribution and for presenting the results of an analysis. The 3-parameter Weibull distribution cdf can be expressed as

$$F(t;\mu,\sigma,\gamma) = \Phi_{\text{sev}}\left[\frac{\log(t-\gamma)-\mu}{\sigma}\right], \quad t > \gamma.$$

   (a) Assuming that $\gamma = 0$, show how to make, starting with plain graph paper, a Weibull probability plot of a complete sample of 10 observations reported as exact failures.

   (b) Show why any Weibull cdf will plot as a straight line on Weibull probability plot in part 7a.

   (c) Assuming that σ is specified, explain how how to make, starting with plain graph paper, an alternative Weibull probability plot that would allow graphical estimation of γ and μ.

8. An company maintains a fleet of trucks and would like to do a retrospective study to compare two different brands of head light bulbs. The fleet contains 55 trucks. All trucks started with head light bulb Brand A, but during the life of the trucks, when a head light failed, it was replaced with Brand B. The available data show the date of purchase of each truck and the date of replacement for head lights that were replaced, and whether it was a replacement of the left or the right bulb.

   (a) How would you organize the data to treat them as repairable system data and what kind of questions could you answer from an analysis focusing on the mean cumulative number of failures.

   (b) Under what conditions would it be appropriate to use the times between failures of the head lights to estimate a time to failure distribution for head light life?

   (c) Give an example of a situation where it would not be appropriate to use the times between failures of the head lights to estimate a time to failure distribution for head light life.

   (d) What graphical tool could you use to check to see if there is correlation between the times to failure between the left and right bulb failure times.

   (e) An engineer has suggested that when one headlight is replaced, that it would be good practice to replace both light bulbs. Explain the logic of this suggestion and under what assumptions this would *and* would not be a wise thing to do.