Stat 565 First Exam Name ____________________
Fall 2003

Instructions: Do not write your answers on this exam. Use other paper. Show your work. You may receive partial credit for incomplete answers. Do not be overly concerned about obtaining numerical values for final solutions. You will receive full credit if you show how to solve the problem by presenting appropriate formulas or descriptions, even when you do not complete the calculations. Put your name on all sheets of paper that you turn in.

1. In a study of the effects of two drugs, called drug A and drug B, for treating patients diagnosed with myelomatosis, 16 patients were randomly assigned to treatment, with 8 subjects assigned to each drug. Results on survival times are shown below. The variable **Time** gives the time in days from the start of the treatment until either death or censoring (due to loss of follow-up period). The variable **Status** is coded 1 for those who died and 0 for those who were censored.

Results for Drug A			Results for Drug B		
Patient	Time	Status	Patient	Time	Status
1	8	1	9	45	1
2	132	1	10	216	1
3	52	1	11	192	1
4	220	1	12	145	0
5	132	0	13	205	1
6	18	1	14	149	1
7	18	1	15	360	0
8	43	1	16	221	1

a) Is this a case-control study, a cohort study, a randomized clinical trial, or some other type of study? Explain.

b) Identify the experimental unit, any treatment factor, and any blocking (or matching) factor. If there are none, just say so.

c) Survival status at 120 days is summarized in the following table.

	Alive at 120 Days	Dead before 120 Days
Drug A	3	5
Drug B	7	1

i. Estimate the relative risk of dying before 120 days for treatment these two drugs.

ii. Is there a significant difference in survival rates at 120 days for these two drugs? Show how you performed your test and clearly state your conclusion.

d) Compute the Kaplan-Meier estimator of the survivor function for treatment with Drug A. (Use only the data for Drug A.)

e) The value of the log-rank test statistic for the data shown above is 3.92. State the null and alternative hypotheses for this test. State the model assumptions on which this test is based. Can the null hypothesis be rejected at the $\alpha = .05$ level of significance?

2. Recall that the survivor function for a Weibull failure time distribution can be written as

$$S(t) = e^{-(\lambda t)^{\alpha}} \quad \text{for} \quad \lambda > 0,\ \alpha > 0,\ \ t \geq 0$$

a) Give a formula for the hazard function for this parameterization of the Weibull distribution.

b) Suppose we wanted to fit a Weibull model to the survival time data for Drug A in Problem 1. Assuming that the subjects respond independently, write out the likelihood function for those data.

c) Using Weibull distributions for the survival time distributions for the two drugs, explain how you could test the null hypothesis that the survival distributions are the same for the two drugs. Outline how you would do the test, but do not perform any calculations.

3. In a case-control study of low infant birth weights, data on 120 women who had low birth weight babies were extracted from records of several hospitals in Boston. Each of these records was matched by age and hospital to a record of a women who gave birth to a normal weight baby. The risk factor of primary interest was smoking status during pregnancy.

a) The data are displayed in the following table:

		Case	
		Smoked	Did not Smoke
Control	Smoked	12	9
	Did not Smoke	28	71

Test the null hypothesis that the proportion of women who smoke during pregnancy is the same for women who have normal and low birth weight babies. Show how you performed your test and state your conclusion.

b) Using the word "case" to identify a women who gave birth to a low birth weight baby and the word "control" to identify a women who gave birth to a normal birth weight baby

$$Y_{1i} = \begin{cases} 1 & \text{if the case smoked during pregnacy} \\ 0 & \text{if the case did not smoke during pregnancy} \end{cases}$$

$$Y_{2i} = \begin{cases} 1 & \text{if the control smoked during pregnacy} \\ 0 & \text{if the control did not smoke during pregnancy} \end{cases}$$

Consider the logistic regression model given by the conditional probabilities

$$P(Y_{1i} = 1) = \frac{\exp(\alpha_i + \beta)}{1 + \exp(\alpha_i + \beta)} \quad \text{for the i-th case}$$

$$P(Y_{2i} = 1) = \frac{\exp(\alpha_i)}{1 + \exp(\alpha_i)} \quad \text{for the i-th control}$$

Then

$$\log\left(\frac{P(Y_{ji} = 1)}{P(Y_{ji} = 0)}\right) = \alpha_i + \beta Z_j \quad \text{i=1,…,n and j=1,2}$$

where $Z_1 = 1$ and $Z_2 = 0$, and

$$e^{\beta} = \frac{\text{odds that a case smokes during pregnancy}}{\text{odds that a control smokes during pregnancy}}$$

for each matched pair. Show that e^{β} is also

$$\frac{\text{odds that a smoker is a case}}{\text{odds that a non-smoker is a case}}$$

c) Write out the conditional likelihood function for (Y_{1i}, Y_{2i}) i=1,…,n. Clearly define any additional notation that you use.

d) The maximum conditional likelihood estimate for β is $\hat{\beta} = 1.135$ with standard error 0.383. Construct an approximate 95% confidence interval for the relative risk of having a low birth weight baby if a woman smokes during pregnancy.

4. The results shown below were obtained from fitting the Cox proportional hazards model to the survival data for 65 patients who received heart transplants. One objective of this study was to examine the effects of some factors that could help to explain why some patients survived longer than others. The explanatory variables are:

SURG	Coded 1 if the patient had open heart surgery prior to acceptance into the heart transplant program, 0 otherwise
M2	Coded 1 if there was a donor-recipient mismatch on the HLA-A2 antigen, 0 otherwise.
AGE	Age (in years) at the time of the heart transplant

The survival time is the number of days from the date of the heart transplant until death. Some survival times were right censored due to loss to follow up or surviving beyond the end of the follow up period. The hazard function for the i-*th* patient has the form

$$h_i(t) = h_0(t)\, e^{[\beta_1(\mathrm{SURG_i}) + \beta_2(\mathrm{M2}_i) + \beta_3(\mathrm{AGE_i}) + \beta_4(\mathrm{M2_i})(\mathrm{AGE_i})]}$$

The estimated coefficients are as follows:

	Estimated Coefficient $\hat{\beta}_j$	Standard Error of $\hat{\beta}_j$	Chi-square Test	p-value	$e^{\hat{\beta}_j}$
SURG	-0.770	0.297	6.72	.009	0.463
M2	0.620	0.313	3.92	.047	1.859
AGE	0.049	0.018	7.41	.006	1.050
AGExM2	0.032	0.012	7.11	.008	1.033

The large sample estimate of the covariance matrix for the parameter estimates is

$$\begin{array}{c} \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{array}
\begin{array}{c} \begin{array}{cccc} \beta_1 & \beta_2 & \beta_3 & \beta_4 \end{array} \\
\begin{bmatrix} .08821 & -.02975 & .00241 & -.00057 \\ -.02975 & .09797 & .00152 & .00165 \\ -.00241 & .00152 & .00032 & .00008 \\ -.00057 & .00165 & .00008 & .00014 \end{bmatrix} \end{array}$$

a) Give an interpretation of the quantity e^{β_1} for the prior surgery factor in the context of this heart transplant study.

b) Suppose the estimated baseline survivor function, $\hat{S}_0(t)$, is equal to 0.8 at t=20 days. Estimate the probability that a 60 year-old subject who had previous open-heart surgery and did not match the donor on the HLA-A2 antigen survives at least 20 days. Explain how you would obtain a standard error for this estimate, but do not compute a value for the standard error.

c) How would you interpret the effect of age at time of transplant on survival for a patient who had previous open-heart surgery and did not match the donor on the HLA-A2 antigen? Give a 95% confidence interval for an appropriate hazard ratio.

d) The first three deaths occurred at 5, 14, and 17 days, and there were no other deaths or censored times before 20 days. Suppose the person entering the data incorrectly entered 9, 18 and 19 instead of 5, 14 and 17, respectively. How would the parameter estimates computed from the partial likelihood evaluated with these incorrect data differ from the parameter estimates computed from the partial likelihood evaluated with the correct data? If you do not have enough information to reach a conclusion, please indicate what other information you would need.

Exam Score _______________