Practice Exam Questions for STA 442/2101: December, 2009
Some of these questions will be on the exam

1. Problem 8.9.20 from Davison

2. Samples of the same material are sent to four laboratories for chemical analysis as part of a study to determine whether laboratories give the same results. The results for laboratories A–D, with the row means, are:

A	58.7	61.4	70.9	59.1	58.2	59.66
B	62.7	64.5	63.1	59.2	60.3	61.96
C	55.9	56.1	57.3	55.2	58.1	56.52
D	60.7	60.3	60.9	61.4	62.3	61.12

   (a) Write out the linear model used to fit this data, explaining all notation and giving needed assumptions.

   (b) Explain how to assess the consistency of the data with the null hypothesis that there is no difference among the laboratories.

3. Problem 9.6.3 of Davison

4. Problem 9.6.2 of Davison

5. One way to model binary data is to suppose the existence of an underlying continuous variable, W_i, say, where $W_i = x_i^T\beta + \epsilon_i$, and to define Y_i by

$$\begin{aligned} Y_i &= 1 \quad \text{if} \quad W_i > 0, \\ &= 0 \quad \text{if} \quad W_i \leq 0. \end{aligned}$$

   Show that if $\epsilon_i \sim N(0,1)$ that this leads to a binary model for Y_i in which $\Phi^{-1}(p_i) = x_i^T\beta$, where $p_i = \Pr(Y_i = 1)$, whereas if ϵ_i follows a logistic distribution

$$F(\epsilon_i) = \frac{1}{1+\exp(-\epsilon_i)}, \quad -\infty < \epsilon_i < \infty,$$

   that this leads to the logistic regression model. The former is called *probit* regression.

6. The following data set was contrived to illustrate some aspects of regression with survival data:

```
  time status x sex
1    4      1 0   0
2    3      1 2   0
3    1      1 1   0
4    1      0 1   0
5    2      1 1   1
6    2      1 0   1
7    3      0 0   1
```

Here `time` is the days to failure or censoring, `status` indicates death (1) or censoring (0), and `x` and `sex` are covariates. The model fit to the data was

$$h(t;\beta,x,\text{sex}) = h_0(t)\exp(\beta_0 + \beta_1 \text{x} + \beta_2 \text{sex}),$$

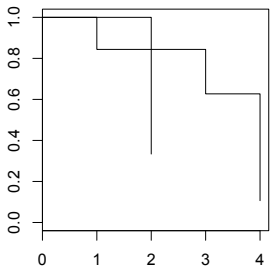
where $h(t;\beta,\text{x},\text{sex})$ is the hazard function for the distribution of time to failure.

The estimates and estimated standard errors for β_0, β_1 and β_2 are given in the following excerpt of `R` code:

```
> Surv(test1$time,test1$status)
[1] 4  3  1  1+ 2  2  3+

> summary(coxph(Surv(time, status) ~ x + factor(sex), test1))
Call:
coxph(formula = Surv(time, status) ~ x + factor(sex), data = test1)

  n= 7

                coef exp(coef) se(coef)     z Pr(>|z|)
x             0.7812    2.1841   0.7976 0.979    0.327
factor(sex)1 0.9338    2.5441   1.4081 0.663    0.507

             exp(coef) exp(-coef) lower .95 upper .95
x                2.184     0.4579    0.4575     10.43
factor(sex)1     2.544     0.3931    0.1610     40.19

Rsquare= 0.15   (max possible= 0.822 )
Likelihood ratio test= 1.13  on 2 df,   p=0.567
Wald test            = 0.96  on 2 df,   p=0.619
Score (logrank) test = 1.05  on 2 df,   p=0.5927
```

(a) Explain in non-technical language the inference implied by the confidence interval for `x`.

(b) A plot of the estimated survivor functions for the data with `sex` = 0 and `sex` = 1 (ignoring `x`) are shown below. Which of the two lines corresponds to `sex` = 1? Does it appear that the proportional hazards assumption is valid?



7. Problem 10.10.2 of Davison

8. Problem 10.10.9 of Davison

9. You should also review Examples H and J, and all the questions from homework.