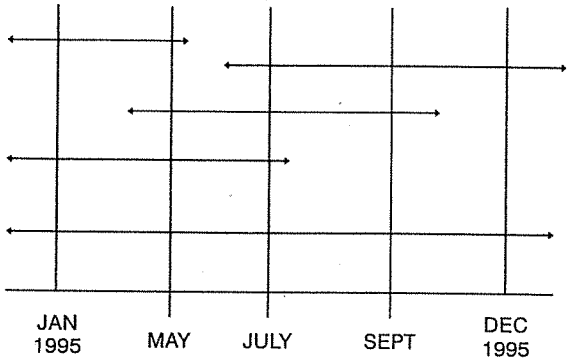
STAT 425/525 MidTerm Exam

(Due by 5:00 pm - Wednesday, October 19, 2016)

1. The figure below shows cases of a disease in a defined population of 50 individuals. There are 5 cases of a disease. The disease is fatal so the length of the lines shows the survival time.



   (a) What is the incidence proportion over the specified period of time.

   (b) What is the point prevalence in July?

   (c) What is the point prevalence in December?

2. The data are described in Problems 9.1 on pages 144-145 and the problem continues as Problem 10.3 on page 163. Refer to the specific questions below. R code to load these data into R is attached.

   (a) Calculate the ratio of the odds of TB given exposure to biomass fuel over the odds of TB given no exposure for the pooled data. Also give an approximate 95% confidence interval and interpret in terms of the problem.

   (b) Calculate the stratum specific (< 1000, ≥ 1000 peso income) odds ratios. Again provide me with the approximate 95% intervals.

   (c) Using the Cochran-Mantel-Haenzel approach provde me with an estimate of the common odds ratio adjustd for income along with an approximate 95% interval. Use the `mantelhaen.test` function.

   (d) A key assumption of the CMH procedure is that there is no interaction. Is there any evidence of an interaction involving income and biomass fuel exposure? You can use the Breslow-Day test (`breslowday.test` which I will provide you with).

(e) Based on the above which odds ratio would you present in, say, a final report and why? Filling in a table like the one below may help with your answer.

	OR	CI
Pooled		
CMH		
< 1000		
≥ 1000		

3. Refer to Figure 8.6 on page 106 in the text. Assume that access to medical care has no influence on health condition.

   (a) Draw a new causal graph reflecting this assumption.

   (b) Is there any confounding of the relationship between vaccination and health condition? If so identify which of the 3 variables are confounders: socioecomic status, access to medical care, and/or family history and explain why the confounding exists. If no confounding exists, explain why.

   (c) An investigator adjusts for access to medical care. Will their analysis of the relationship between vacccination and health condition be confounded? If so identify which of the remining 2 are confounders: socioecomic status and/or family history and explain why the confounding exists. If no confounding exists, explain why.

4. A retrospective study of lung cancer and tobacco smoking in several English hospitals resulted in the following data.

	LCYes	LCNo
None	7	61
<5	55	129
5-14	489	570
15-24	475	431
25-49	293	154
50+	38	12

   (a) Test for independence between lung cancer risk and smoking using the chi-squared test for independence. Interpret the results.

   (b) You should have found strong evidence of an associaton. Where does the independence break down and how?

   (c) Test for a trend in risk of lung cancer over smoking level. Interpret the results. Do the results of this test tell you anything about the direction of the trend?

   (d) Perform a goodness-of-fit test and give the results.

5. An experiment on the efficacy of a vaccine was carried out on 215 volunteers, 104 of which were randomly assigned to receive the vaccine with the rest receiving a placebo injection. The results were

	Caught Disease	Did Not Catch Disease	
Vaccine	10	94	104
Placebo	33	78	111
	43	172	

Estimate the Absolute Risk Reduction and the Number Needed to Treat. Give approximate 95% confidence intervals for these and interpret the results.

6. STAT 525: On Homework 4 you used the Delta Method to show that an approximation to the standard error of the MLE of the log of the odds

$$\log\left(\frac{\widehat{p}}{1-\widehat{p}}\right)$$

was given by

$$SE\left[\log\left(\frac{\widehat{p}}{1-\widehat{p}}\right)\right] = \sqrt{\frac{1}{n\widehat{p}(1-\widehat{p})}}$$

(a) Assuming approximate normality and approximate unbiasedness of the estimator (not unreasonable for the asymptotic behavior of an MLE) show that an approximate $100(1-\alpha)\%$ Wald CI for the log of the odds is given by

$$\log\left(\frac{\widehat{p}}{1-\widehat{p}}\right) \pm z_{1-\alpha/2}/\sqrt{n\widehat{p}(1-\widehat{p})}$$

(b) Show how to use this interval to obtain an interval for p itself.