Homework 4 STATS 449 (Due 10/18)

Instructions: Follow the instructions in each problem carefully. Late homework cannot be accepted.

1) A diagnostic test for a disease is said to be 90% accurate in that if a person has the disease, the test will detect it with probability 0.9. Also, if a person does not have the disease, the test will report that he or she does not have it with probability 0.9. Only 1% of the population has the disease in question. If a person is chosen at random from the population and the diagnostic test indicates that she has the disease, what is the conditional probability that she does, in fact, have the disease? Are you surprised by the answer?

Hint: Use Bayes Formula

2) The following data are based on a study on coronary artery disease. The sample is one of convenience since the patients studied were people who came to a clinic and requested an evaluation.

Investigators were interested in whether electrocardiogram (ECG) measurement was associated with coronary artery disease. Gender was thought to be associated with disease status, so investigators poststratified the data into male and female groups (gender is a potential confounder).

		Coronary Disease (Y)	
Gender (Z)	ECG (X)	Yes	No
Female	≥ 0.1 ST segment depression	21	21
	< 0.1 ST segment depression	9	29
Male	≥ 0.1 ST segment depression	37	10
	< 0.1 ST segment depression	20	24

Use SAS to answer parts a), b), and c) below. **Show** SAS code.

a) Conduct the Cochran-Mantel-Haenszel test of conditional independence. State the null hypothesis and give evidence to support your conclusions.

b) Conduct the Breslow-Day test for homogeneity of odds ratios. State the null hypothesis and give evidence to support your conclusions.

c) Does it seem reasonable to estimate a common odds ratio? If so, report the Mantel-Haenszel estimator $\hat{\theta}_{MH}$ and a 95% confidence interval. Interpret your findings.

3) Continuing with the data from Problem 2, fill in the symbols for the loglinear models that correspond to the given logit models in the table below. Use R to fit the corresponding loglinear models and fill in the degrees of freedom and G^2. Also fill in N and p for each model, respectively. Note that the values in the N columns do not change, but are different for the logit and loglinear models. **Show** your R code and the one line of output that shows G^2 and the degrees of freedom.

Logit Model	Logit Model N	Logit Model p	Loglinear Model	Loglinear Model N	Loglinear Model p	df	G^2
α							
$\alpha + \beta_i^X$							
$\alpha + \beta_k^Z$							
$\alpha + \beta_i^X + \beta_k^Z$							

a) Convince yourself that these logit and loglinear models are equivalent by fitting the corresponding logit models. For the logit model with only an intercept term, the R command (assuming "ymat" contains your two-column data matrix) is

```
m1 <- glm(ymat~1,.....)        # or just look at the Null Deviance
```

Show your R code for the logit models.

b) Now, compare these likelihood-based results to the Cochran-Mantel-Haenszel results found in Problem 2. That is,

1) Compare the values of the CMH and G^2 statistics that test conditional independence between X and Y.

2) Compare the values of the Breslow-Day and G^2 statistics that test for a common odds ratio.

3) Can you estimate a common odds ratio? How do you find it using the logit model? The loglinear model? Calculate it, along with a 95% confidence interval and interpret your findings.

c) In the association between X and Y, are any collapsibility conditions satisfied? Must provide supporting evidence.

Hint: This is a good example. Fit all 9 loglinear models. Check G^2 and df. See Lecture 12 for an example.