Homework 2 STATS 449 (Due 10-4)

Instructions: Answer the following questions using SAS or R as indicated. Perform hand calculations as required. Show SAS and R output including plots as requested. Late homework cannot be accepted. Neatness and presentation are important.

1) Nearsightedness (myopia) typically develops during the childhood years. Recent studies have explored whether there is an association between development of nearsightedness and the use of night lights with infants. Parents of children who were seen as outpatients in a university pediatric ophthalmology clinic were asked to participate in a study about their child's bedroom lighting at night before the age of 2 years. Results are shown in the table below. Use SAS to answer the following questions.

	Myopia (Y)	
Night Light(X)	Yes	No
Used	78	115
Not Used	18	114

a) Report the odds ratio and a 95% confidence interval. Interpret the results.

b) Report the relative risk and a 95% confidence interval. Interpret the results.

c) Conduct a test of statistical independence. Report the results from two tests that you know and interpret. Use $\alpha = 0.05$.

d) Give a probabilistic argument that explains why the odds ratio is greater than the relative risk. You may do this part by hand.

2) In an article about crime in the United States, *Newsweek* magazine quoted FBI statistics stating that of all blacks slain in 2010, 92% were slain by blacks, and of all whites slain in 2010, 85% were slain by whites. Let Y denote race of victim and X denote race of murderer.

a) Calculate and interpret the odds ratio between X and Y.

b) Given only the information above, can you calculate the relative risk (see the definition of RR in Lecture 7)? Why or why not?

3) The bootstrap is a useful tool for estimating the sampling distribution of an estimator. The estimated standard error of the bootstrap sample can be used when constructing confidence intervals. Assume that the data in the table below were generated from the multinomial distribution

$$n_{11}, n_{12}, n_{21}, n_{22} \sim \text{Multinomial}(n; \pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$$

	Disease	
Exposure	Yes	No
Yes	30	10
No	20	20

a) Use R to bootstrap a sample of size 1,000 from the approximate sampling distribution of the odds ratio, $\hat{\theta}$. Make a histogram and report the mean and standard error of the sample.

b) Repeat part a) except now approximate the sampling distribution of the log odds ratio, $\log\hat{\theta}$. Make the histogram and report the mean and standard error. Which histogram appears to follow a normal distribution? Use the results to construct a 95% confidence interval for θ.

c) Calculate the odds ratio, $\hat{\theta}$ by hand. Calculate the estimated standard error of the log odds ratio, $\hat{se}(\log\hat{\theta})$ by hand using the formula, and construct a 95% confidence interval for θ. Compare to the confidence interval calculated in part b).

4) A study was conducted to examine the association between cardiovascular disease (CVD) and obesity. Age is considered to be a potential confounder, so results are presented for those less than 50 years old, and those greater or equal to 50 years old.

		CVD (Y)	
Age (Z)	Obesity (X)	Yes	No
< 50	Yes	10	90
	No	35	465
≥ 50	Yes	36	164
	No	25	175
Total	Yes	46	254
	No	60	640

a) Calculate the $X - Y$ marginal odds ratio $\hat{\theta}_{XY}$ by hand.

b) Calculate the conditional $X - Y$ odds ratios $\hat{\theta}_{XY(1)}$ and $\hat{\theta}_{XY(2)}$ in the partial tables by hand.

Use SAS to answer parts c), d), and e). Assume $\alpha = 0.05$.

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

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

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