BS852: Mid Term Exam. March 2016.

Question 1   ….. out of 45pts

Question 2   ….. out of 20pts

Question 3   …..out of 35pts

Rules

- ***The paper must be handed in on March 21, at the beginning of class.***
- ***The paper should be fruit of your work.***

Question 1. [45pts]

In a paper published in the New England Journal of Medicine in 1972, Johnson and Johnson selected 85 Hodgkin's lymphoma patients who had a sibling of the same sex who was free of the disease and whose age was within 5 years of the patients. The goal of the study was to examine whether tonsillectomy is associated with the disease.

1. **What type of study is this?[5pts]**

They presented the data as:

	Tonsillectomy	None
Hodgkin's	44	41
Sib control	52	33

and computed the OR=0.68, with a chi-square 1.53 that showed no significant association between tonsillectomy and Hodgkin's lymphoma.

2. **What is problematic with this analysis? [3pts]**

3. **Complete the table below in which the data are summarized by matched pairs.[5pts]**

	Control	
Case	Tonsillectomy	None
Tonsillectomy	37	
None		26

4. **The OR for the association turns out to be OR =0.467, with a Chi-square test 2.901 on 1 degrees of freedom. Compute a 95% confidence interval for the OR. Use this result to discuss the conclusion of the study based on this analysis. [10pts]**

Investigators had kept a record of the sex of the matched pairs and decided to test the hypothesis that there is an interaction between sex and tonsillectomy on the chance for Hodgkin's disease. They were given this incomplete tables of matched pairs stratified by sex, and were told that 40 pairs included female siblings, and 45 pairs included male siblings.

	Control (male)		N. pairs		Control (Female)		N. pairs
Case (male)	Tonsillectomy	None		Case (female)	Tonsillectomy	None	
Tonsillectomy		2	30	Tonsillectomy			
None				None		13	
N. pairs				N. pairs			

5. **Complete the 2 tables of matched pairs and explain how the missing numbers were computed. Use the completed tables to test the hypothesis that there is an interaction between sex and tonsillectomy on the chance for Hodgkin's disease. [10pts]**

6. **Can you describe possible limitations of the analysis in part 5? [2pts]**

7. **Could the investigators use a similar approach to test whether there is an interaction between age and tonsillectomy and whether there is an interaction between race and tonsillectomy? [2pts]**

8. **Provide overall conclusions of the analysis. [8pts]**

Question 2

Investigators from the geriatric section at Boston Medical Center conducted a study to test whether there is an association between alcohol use and risk for fall at older ages. A sample of 700 subjects aged 65 and older who felt at least once in the past year and a sample of 2793 subjects age 65 and older who did not fall in the past year were enrolled in the study, and stratified by gender and whether they consume alcohol at least once per week. The data are displayed in the table below

Fall in past year	Males		Females	
	Exposed	Unexposed	Exposed	Unexposed
Yes	250	80	160	210
No	733	920	340	800

1. **What type of study is this? [3pts]**

The data were analyzed using the function epi2b2 in R, and below is part of the R script for analysis and the produced output:

```
data.q2 <- c(250, 80, 733, 920, 160, 210, 340, 800)
dim(data.q2) <- c(2,2,2)
dimnames(data.q2)[[1]] <- c("exposed", "unexposed");  dimnames(data.q2)[[2]] <- c("case", "control")
dimnames(data.q2)[[3]] <- c("Males", "Females")

library(epiR)
summary(epi.2by2(dat = data.q2, method = "case.control", conf.level = 0.95))
```

$OR.strata
```
      est        lower     upper
1 3.919551    2.976795   5.203186
2 1.792038    1.396387   2.298184
```

$OR.crude
```
     est        lower     upper
1 2.265749    1.908144   2.692919
```

$OR.mh
```
     est         se          lower      upper
1 2.612077    0.08990484    2.190074   3.115395
```

$chisq.strata
```
  test.statistic  df   p.value
1     108.59287  1    0.00000e+00
2     22.70972   1    1.88412e-06
```

$chisq.crude
```
  test.statistic  df   p.value
1     93.05741   1    0
```

$chisq.mh
```
  test.statistic  df   p.value
1     116.4953   1    3.702479e-27
```

$OR.homog
```
  test.statistic  df   p.value
1     18.16811   1    2.022366e-05
```

2. **Test the null hypothesis that there is no interaction between sex and exposure to regular alcohol use, using Breslow-Day test of interaction. [2pts]**

3. **Write down stratum-specific odds ratios, and 95% confidence intervals for each sex-stratum. Do the results show that the sex-specific associations between exposure to alcohol use and risk of falling are significant? Describe in words the effects of sex on the association between regular alcohol consumption and risk of falling. Be complete [5 pts]**

4. **Write down the sex adjusted odds ratio, the 95% confidence interval, and the Mantel-Haenszel χ^2 and use the results to show whether there is a significant association between exposure to alcohol and risk of falling, after adjusting for sex. [5pts]**

5. **What are the overall conclusions based on these analyses? [5pts]**

Question 3
Tuberculosis (TB) incidence rates in Moldova (in the former Soviet Union) are high and the risk of multidrug-resistant TB (MDR-TB) among TB cases is especially high. Investigators want to build a model that can best predict odds of MDR, given that an individual has TB and. They use surveillance data from Moldova for 2010, containing 2,283 TB cases, 1,425 of whom have multidrug resistance. The remaining 858 have drug susceptible TB. The investigators looked at the following potential risk factors: Age_lin (continuous variable), Male (1=male, 0=female), Homeless (1=yes, 0=no). The outcome variable "MDR" is coded 1 when the individual has MDR-TB and 0 when they have drug susceptible TB.

The investigators fitted three different logistic regression models to the data. The data were analyzed using R the glm statement and the output were as follows:

MODEL 1:
Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.153308	0.160779	-0.954	0.3403	
Age_lin	-0.014362	0.003498	-4.106	4.03e-05	***
Male	0.242044	0.108452	2.232	0.0256	*
Homeless	0.456893	0.171444	2.665	0.0077	**

Null deviance: 3022.6 on 2282 degrees of freedom
Residual deviance: 2996.0 on 2279 degrees of freedom

MODEL 2:
Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.132230	0.160291	-0.825	0.4094	
Age_lin	-0.014023	0.003485	-4.023	5.74e-05	***
Male	0.237223	0.108233	2.192	0.0284	*

Null deviance: 3022.6 on 2282 degrees of freedom
Residual deviance: 3003.0 on 2280 degrees of freedom

MODEL 3:
Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.68819	0.09639	-7.140	9.34e-13	***
Male	0.19215	0.10716	1.793	0.0730	.
Homeless	0.43174	0.17079	2.528	0.0115	*

Null deviance: 3022.6 on 2282 degrees of freedom
Residual deviance: 3013.1 on 2280 degrees of freedom

1. **Write down the formulas for each of the three models [6 pts]**

2. **Does age confound the relationship (a) between MDR and "male", (b) between MDR and "homeless"? Explain your answers. [6 pts]**

3. **Which model do you think the investigators should choose as their final model? Explain your answer, showing any workings you did. [6 pts]**

4. **Using model 1, report the factors that are significantly associated with having MDR-TB. List the odds ratios and confidence intervals and describe these associations in words [10 pts]**

The investigators then obtain information on education level attained by the individuals in the dataset. The variable is coded with 5 levels as follows: 1=Primary, 2=Secondary, 3= Specialized secondary, 4=Higher, 5=No education. They include this in Model 1 above and get the following results:

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	-0.235158	0.183570	-1.281	0.20018	
Age_lin	-0.013729	0.003525	-3.895	9.81e-05	***
Male	0.242635	0.109021	2.226	0.02604	*
Homeless	0.483545	0.173848	2.781	0.00541	**
education2	0.103848	0.102886	1.009	0.31281	
education3	-0.082985	0.153476	-0.541	0.58871	
education4	0.270061	0.247524	1.091	0.27525	
education5	-0.284711	0.413486	-0.689	0.49110	

Null deviance: 3022.6 on 2282 degrees of freedom
Residual deviance: 2992.1 on 2275 degrees of freedom

5. **Is this a better model than model 1? Show any workings. [7 pts]**