MATH 523 - EXERCISES 3

These exercises are not for assessment

1 The table below contains data collected to examine the relationship between blood pressure and heart disease. Investigate these data using suitable GLMs, treating blood pressure first as a factor predictor, secondly as a continuous covariate. Is there any variation of the incidence of heart disease with blood pressure ?

	Blood Pressure (millibars Hg)							
Heart Disease	<117	117-126	127-136	137-146	147-156	157-166	167-186	>186
Present	3	17	12	16	12	8	16	8
Absent	153	235	272	255	127	77	83	35

2 **The Mantel-Haenszel Procedure:** The table below summarizes 13 randomized controlled trials that compared two antihypertensive (blood pressure lowering) drugs.

	Treatment		Control	
Study	# Strokes	Total	# Strokes	Total
1	59	3903	88	3922
2	13	1721	22	1706
3	60	8700	109	8654
4	5	186	20	194
5	1	193	6	196
6	25	1048	36	1004
7	43	233	52	219
8	1	68	3	63
9	2	45	1	42
10	10	49	21	48
11	18	534	34	529
12	32	416	48	424
13	20	419	39	465

Consider these data as a $2 \times 2 \times 13$ table. Let ψ_k denote the odds ratio for study k, that is,

$$\psi_k = \frac{\Pr[\text{ Stroke} \mid \text{Treatment in Study } k\,]/(1-\Pr[\text{ Stroke} \mid \text{Treatment in Study } k])}{\Pr[\text{ Stroke} \mid \text{Control in Study } k]/(1-\Pr[\text{ Stroke} \mid \text{Control in Study } k])} = \frac{\pi_{11k}/\pi_{01k}}{\pi_{10k}/\pi_{00k}}$$

say, and let $\xi_k = \log \psi_k$ be the study-specific log-odds ratio, for $k = 1, ..., 13$.

(a) Obtain the mles and estimated standard errors of the parameters ξ_k, $\widehat{\xi}_k$ and s.e.$(\widehat{\xi}_k)$

(b) Carry out a test of the **heterogeneity** of the ξ_k, that is, attempt to reject the null hypothesis of **homogeneity**

$$H_0 : \xi_1 = \xi_2 = \ldots = \xi_{13}$$

against the general alternative using the following test procedure: define test statistic Q_ξ by

$$Q_\xi = \sum_{k=1}^{13} w_{\xi_k} \left(\widehat{\xi}_k - \overline{\xi}\right)^2 \qquad \text{where} \qquad \overline{\xi} = \frac{\sum_{k=1}^{13} w_{\xi_k}\widehat{\xi}_k}{W_\xi} \qquad w_{\xi_k} = \frac{1}{\left\{\text{s.e.}(\widehat{\xi}_k)\right\}^2} \qquad W_\xi = \sum_{k=1}^{13} w_{\xi_k}$$

which has, approximately, a Chi-squared distribution with $13 - 1 = 12$ degrees of freedom.

(c) If there is no evidence to reject the null hypothesis of homogeneity, the quantity $\overline{\xi}$ is an estimate of the overall treatment effect ξ, and the corresponding estimator has an approximate

$$\mathcal{N}\left(\xi, W_{\xi}^{-1}\right)$$

distribution. By considering this result, decide whether there is any evidence of a treatment effect in this study.

(d) Repeat (a) to (c) using a similar testing approach, but now for the **risk difference** parameter, $\delta_k = \pi_{11k} - \pi_{10k}$, using a similarly defined test statistic Q_δ

$$Q_\delta = \sum_{k=1}^{13} w_{\delta_k}\left(\widehat{\delta}_k - \overline{\delta}\right)^2 \qquad \text{where} \qquad w_{\delta_k} = \frac{1}{\left\{\text{s.e.}(\widehat{\delta}_k)\right\}^2}$$

and so on, which has the same approximate null distribution. Take care with s.e.$(\widehat{\delta}_k)$.

3 The table below gives data on the relationship between low birth weight (LBW) and smoking, stratified by race. Analyse the data using:

(a) a stratified analysis using the Mantel-Haenszel statistic, and

(b) logistic regression models,

		Race					
		White		Black		Other	
Smoking		Yes	No	Yes	No	Yes	No
LBW	Yes	19	4	6	5	5	20
	No	33	40	4	11	7	35
Total		52	44	10	16	12	55

4 The table below contains case-control data collected to investigate the relationship between myocardial infarction (MI) and recent oral contraceptive (OC) use (last use within the month before admission). Information on age and cigarette smoking were also available.

Cigarette	OC	Age Groups (years)									
		25–29		30–34		35–39		40–44		45-49	
Smoking	Use	MI	Control	MI	Control	MI	Control	MI	Control	MI	Control
None	Yes	0	25	0	13	0	8	1	4	3	2
	No	1	106	0	175	3	153	10	165	20	155
1-24	Yes	1	25	1	10	1	11	0	4	0	1
(per day)	No	0	79	5	142	11	119	21	130	42	96
$\geqslant$25	Yes	3	12	8	10	3	7	5	1	3	2
(per day)	No	1	39	7	73	19	58	34	67	31	50

Let

$$\pi_{1ijk} = \Pr(\text{MI}|\text{OC group } i, \text{age group } j, \text{smoking group } k)$$

where $i = 1, 2$ (OC use, non-OC use), $j = 1, 2, 3, 4, 5$ (25–29, 30–34, 35–39, 40–44,45–49), $k = 1, 2, 3$ (None, 1–24, $\geqslant$ 25). Investigate these data using logistic regression modelling. Is there an association between oral contraceptive use and myocardial infarction ? Justify your answer.

The following code in R can be used: please check for errors !

Code for Q2:

```
#Study Data

x<-c(59,13,60,5,1,25,43,1,2,10,18,32,20)
m<-c(3903,1721,8700,186,193,1048,233,68,45,49,534,416,419)
y<-c(88,22,109,20,6,36,52,3,1,21,34,48,39)
n<-c(3922,1706,8654,194,196,1004,219,63,42,48,529,424,465)

#Effects Estimates
pi1<-x/m
pi0<-y/n
rr<-pi1/pi0
delta<-pi1-pi0
xi<-log((pi1/(1-pi1))/(pi0/(1-pi0)))
round(cbind(pi1,pi0,rr,delta,xi),4)

#Aggregated Log-Odds Ratio Quantities
w.xi<-(1/x+1/(m-x)+1/y+1/(n-y))^(-1)
xi.hat<-sum(w.xi*xi)/sum(w.xi)
Q.xi<-sum(w.xi*(xi-xi.hat)^2)

Q.xi
qchisq(0.95,12)

#Z-test for fi
z.xi<-(xi.hat)*sqrt(sum(w.xi))

#Aggregated Risk Differences Quantities

w.delta<-(pi1*(1-pi1)/m+pi0*(1-pi0)/n)^(-1)
delta.hat<-sum(w.delta*delta)/sum(w.delta)
Q.delta<-sum(w.delta*(delta-delta.hat)^2)

Q.delta
qchisq(0.95,12)
```

Code for Q4:

```
case<-c(0,0,0,1,3,1,0,3,10,20,1,1,1,0,0,0,5,11,21,42,3,8,3,5,3,1,7,19,34,31)
control<-c(25,13,8,4,2,106,175,153,165,155,25,10,11,4,1,79,142,119,130,96,
12,10,7,1,2,39,73,58,67,50)
midata<-cbind(case,control)
mod0<-glm(midata~1,family=binomial)
a<-c(1:5)
age<-factor(rep(a,6))
s<-c(1:3)
i<-rep(10,3)
smoke<-factor(rep(s,i))
o<-c(1,1,1,1,1,2,2,2,2,2)
oc<-factor(rep(o,3))
mod1<-glm(midata~age,family=binomial)
mod2<-glm(midata~oc,family=binomial)
mod3<-glm(midata~smoke,family=binomial)
```