M3S12 BIOSTATISTICS - EXERCISES 4

1. The table below contains data collected to examine the relationship between blood pressure and heart disease. Investigate these data in SPLUS using a suitable GLM.

	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

Is there any variation of the incidence of heart disease with blood pressure ?

2. A case-control study of bladder cancer was conducted to test its putative association with dietary consumption of substance E. Cases and non-cases were selected randomly from white male hospital patients between 50 and 60 years of age residing in a large metropolitan area. The tables below summarizes the data by history of cigarette smoking. From previous investigations, it is known that smoking is a risk factor for bladder cancer. Note that consumption of substance E has been dichotomized into consumers (E) and non-consumers (E'). F denotes the occurrence of bladder cancer

SMOKERS

	E	E'	TOTAL
F	35	5	40
F'	20	10	30
TOTAL	55	15	70

NON-SMOKERS

	E	E'	TOTAL
F	10	15	25
F'	20	30	50
TOTAL	30	45	75

In answering the following questions, assume that the potential confounding effects of age have been appropriately controlled by restriction.

(a) What is the estimated crude **odds ratio** ψ comparing cases and non-cases with regard to previous consumption of substance E when the two tables are pooled ? Provide a 95% confidence interval for ψ.

(b) Compare the two estimated **smoking-specific** odds ratios and their confidence intervals to each other and to the crude odds ratio estimate. What do these results indicate with regard to confounding and interaction (that is, the impact of smoking status on the effect of E as a potential cause of bladder cancer, and *vice versa*) ?

(c) Based on all the information that you have, is a single summary odds ratio appropriate? If so, calculate it; if not, what is the best approach to summarize the data ?

(d) Calculate the crude and smoking-specific **relative risks** and **risk differences**. Comment on confounding and effect modification

For the remaining questions consider the results among SMOKERS ONLY

(e) Write down the logistic regression models for the crude analysis. Define ALL terms in the model. Estimate all parameters, and give appropriate interpretations.

(f) Assume that the 40 cases in this sample represent all cases in a population of smokers, but that the controls are a sample from a control population of 300,000. Fit a logistic regression model : Calculate

(i) the intercept coefficient that would have appeared in a cohort study, and

(ii) the RISK for someone who does not consume E and the relative risk for E. Compare to the odds ratio and discuss.

3. The table below contains 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. These data arise from a *case-control study.*

Cigarette Smoking	OC Use	Age Groups (years) 25–29 MI	Control	30–34 MI	Control	35–39 MI	Control	40–44 MI	Control	45-49 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
≥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

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

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, ≥ 25).

Investigate these data using logistic regression modelling.

Is there an association between oral contraceptive use and myocardial infarction ?. Justify your answer with the aid of the following SPLUS code; this code can be downloaded from

`stats.ma.ic.ac.uk/~das01/M3S12/Exercises/Ex04Q3.ssc`

```
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)
options(contrasts=c("contr.treatment","contr.poly"))
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")
```

4. **The Mantel-Haenszel Procedure:** Consider the data in the table below. It summarises 13 randomized controlled trials that compared two antihypertensive (blood pressure lowering) drugs, labelled as treatment E and control.E'

	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

In the usual notation, let $\pi_{1i} = P(F_i|E_i)$, $\pi_{0i} = P(F_i|E_i')$ and let $\xi_i = \log \psi_i$ be the study-specific log-odds ratio, for $i = 1, ..., 13$.

(a) Obtain the mles and estimated standard errors of the parameters ξ_i, $\hat{\xi}_i$ and $s.e.(\hat{\xi}_i)$

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

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

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

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

which has, approximately, a *Chisquared* 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

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

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

(d) Repeat (a)–(c) using a similar testing approach, but now for the **risk difference** parameter, $\delta_i = \pi_{1i} - \pi_{0i}$, using a similarly defined test statistic Q_δ,where

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

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

`stats.ma.ic.ac.uk/~das01/M3S12/Exercises/Ex04Q5.ssc`