Biostat/Stat 571 Exercise #8

Answer Key

Question 1:

Consider the logistic regression model:

$$\text{logit}(\mu_{ij}) \equiv \text{logit}(\mu_{ij}(\vec{\beta})) = \beta_0 + \beta_1 X_{1,ij} + \beta_2 X_{2,i},$$

where X_1 is a covariate that varies within a cluster and X_2 is a covariate that varies between clusters.

Assume:

- $X_{1,ij} = (j-3)/3$ for all clusters, i (i.e. a "time" variable).
- $X_{2,i} = 0$ for half the clusters and $X_{2,i} = 1$ for the other half (i.e. a "treatment" variable).

We can estimate $\vec{\beta} = (\beta_0, \beta_1, \beta_2)$ within a GEE framework via setting the following estimating function equal to zero:

$$\boxed{U(\vec{\beta}) = \sum_{i=1}^{N} D_i^T(\vec{\beta})\, V_i^{-1}(\vec{\beta}, \alpha)\, \{Y_i - \mu_i(\vec{\beta})\}}$$

$$\text{where} \quad D_i^T(\vec{\beta}) = \frac{\partial \vec{\mu}_i}{\partial \vec{\beta}}$$
$$V_i(\vec{\beta}, \alpha) = A_i(\vec{\beta})^{\frac{1}{2}}\, R_i(\alpha)\, A_i(\vec{\beta})^{\frac{1}{2}}.$$

We can note that for the above logistic regression problem we have:

$$\frac{\partial \mu_{ij}}{\partial \vec{\beta}} = \vec{x}_{ij}\mu_{ij}(1-\mu_{ij})$$
$$A_i(\vec{\beta}) = \text{diag}(\text{V}[Y_{ij}]) = \text{diag}(\mu_{ij}(1-\mu_{ij})).$$

The matrix $R_i(\alpha)$ is the assumed correlation matrix, for the vector of outcomes. The variance-covariance matrix for the resulting estimates, $\hat{\beta}$, is given by:

$$\text{V}[\hat{\beta}] = (\mathcal{I}_M)^{-1} \left\{ \sum_{i=1}^{N} D_i^T(\vec{\beta})\, V_i^{-1}(\vec{\beta}, \alpha)\, \text{COV}[\vec{Y}_i]\, V_i^{-1}(\vec{\beta}, \alpha)\, D_i^T(\vec{\beta}) \right\} (\mathcal{I}_M)^{-1},$$

$$\mathcal{I}_M = \sum_{i=1}^{N} D_i^T(\vec{\beta})\, V_i^{-1}(\vec{\beta}, \alpha)\, D_i^T(\vec{\beta}).$$

(a) Assume that the data are balanced with $n_i = 6$ observations per cluster. Suppose we assume that the true correlation structure is exchangeable (where α is the correlation coefficient):

$$R_i(\alpha) = \begin{bmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \cdots & \alpha & 1 \end{bmatrix}$$

Under this model, if we assume independence (i.e. $R_i(\alpha) = \mathrm{I}_6$, the identity matrix), then we have:

$$\begin{aligned} V_i(\vec{\beta}, \alpha) &= A_i(\vec{\beta})^{\frac{1}{2}}\ \mathrm{I}_6\ A_i(\vec{\beta})^{\frac{1}{2}} \\ &= \mathrm{diag}(\mu_{ij}(1-\mu_{ij})) \end{aligned}$$

$$\begin{aligned} \mathrm{COV}[\vec{Y}_i] &= A_i(\vec{\beta})^{\frac{1}{2}}\ R_i(\alpha)\ A_i(\vec{\beta})^{\frac{1}{2}} \\ &= \mathrm{diag}\left(\sqrt{\mu_{ij}(1-\mu_{ij})}\right)\ R_i(\alpha)\ \mathrm{diag}\left(\sqrt{\mu_{ij}(1-\mu_{ij})}\right) \end{aligned}$$

Under this model, if we assume exchangeable (i.e. the correct correlation matrix), then we have:

$$\begin{aligned} V_i(\vec{\beta}, \alpha) &= A_i(\vec{\beta})^{\frac{1}{2}}\ R_i(\alpha)\ A_i(\vec{\beta})^{\frac{1}{2}} \\ &= \mathrm{diag}\left(\sqrt{\mu_{ij}(1-\mu_{ij})}\right)\ R_i(\alpha)\ \mathrm{diag}\left(\sqrt{\mu_{ij}(1-\mu_{ij})}\right) \end{aligned}$$

$$\begin{aligned} \mathrm{COV}[\vec{Y}_i] &= A_i(\vec{\beta})^{\frac{1}{2}}\ R_i(\alpha)\ A_i(\vec{\beta})^{\frac{1}{2}} \\ &= \mathrm{diag}\left(\sqrt{\mu_{ij}(1-\mu_{ij})}\right)\ R_i(\alpha)\ \mathrm{diag}\left(\sqrt{\mu_{ij}(1-\mu_{ij})}\right) \end{aligned}$$

So, under the correct model we have $V_i(\vec{\beta}, \alpha) = \mathrm{COV}[\vec{Y}_i]$. We can therefore compute the variance-covariance matrix of the estimator that assumes independence ($\hat{\beta}^I$), and of the estimator that assumes exchangeable ($\hat{\beta}^E$), using the formula on the previous page.

Correlation (α)	ARE β_0	ARE β_1	ARE β_2
0.0	1.000	1.000	1.000
0.1	0.998	1.000	0.997
0.3	0.984	1.000	0.980
0.5	0.957	1.000	0.945
0.7	0.898	0.999	0.872
0.9	0.704	0.988	0.629

Table 1: ARE = $\mathrm{Var}[\hat{\beta}^E]$ / $\mathrm{Var}[\hat{\beta}^I]$; Complete data, truth = exchangeable

As the true variance model moves further from independence (i.e. α increases) then the independence estimator is doing worse for both the intercept and the between cluster covariate. Intuitively, one

can say that the amount of information in the data for a covariate that does not vary within a cluster is decreasing with increasing within cluster correlation. Although we have $n_i = 6$ observations contributing information from each cluster, because of the correlation the effective information that is coming from the cluster is equivalent to *fewer* than $n_i = 6$ independent observations.

We can see that the independence estimator for the within-cluster covariate coefficient does well, relative to the optimal exchangeable estimator. Regardless of the correlation structure there are still 6 pieces of information that are being used to estimate the slope β_1.

(b) Now we assume that half of the subjects only complete visits 1, 2 and 3, while the other half complete all 6 of the visits. Table 2 provides the resulting aymptotic relative efficiencies:

Correlation (α)	ARE β_0	ARE β_1	ARE β_2
0.0	1.000	1.000	1.000
0.1	0.987	0.985	0.984
0.3	0.935	0.895	0.921
0.5	0.864	0.752	0.835
0.7	0.751	0.565	0.699
0.9	0.499	0.324	0.397

Table 2: ARE = $\text{Var}[\hat{\beta}^E]$ / $\text{Var}[\hat{\beta}^I]$; Incomplete data, truth = exchangeable

In this scenario, we see that the performance of the independence estimator is even worse. For β_0 and β_2 we see a similar pattern, although the decline in relative efficiency is much steeper with α. For β_1 we find that the independence estimator does considerably worse. Consequently we can see that it would be very advantageous to ensure that the correlation structure is correctly (or at least approximately) specified.

(c) Here we assume that the true correlation matrix is autoregressive of lag 1, AR(1):

$$R_i(\alpha) = \begin{bmatrix} 1 & \alpha & \cdots & \alpha^5 \\ \alpha & 1 & \cdots & \alpha^4 \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^5 & \cdots & \alpha & 1 \end{bmatrix}$$

Table 3 below provides the asymptotic relative efficiencies assuming that we have complete data on all subjects. Here we find that even when there is very strong correlation ($\alpha = 0.9$) there is only a moderate decrease in efficiency associated with using the independence estimator. Getting the variance-covariance structure correct is not critical in this case.

Correlation (α)	ARE β_0	ARE β_1	ARE β_2
0.0	1.000	1.000	1.000
0.1	0.998	0.996	0.998
0.3	0.983	0.975	0.982
0.5	0.959	0.949	0.956
0.7	0.938	0.933	0.931
0.9	0.913	0.938	0.895

Table 3: ARE = Var$[\hat{\beta}^A]$ / Var$[\hat{\beta}^I]$; Complete data, truth = AR(1)

(d) Finally, we can examine the case where the truth is AR(1) but we don't have complete data. Using the same scenario as part (b), Table 4 provides the resulting asymptotic relative efficiencies:

Correlation (α)	ARE β_0	ARE β_1	ARE β_2
0.0	1.000	1.000	1.000
0.1	0.998	0.996	0.997
0.3	0.981	0.971	0.979
0.5	0.953	0.928	0.945
0.7	0.910	0.856	0.894
0.9	0.800	0.640	0.756

Table 4: ARE = Var$[\hat{\beta}^A]$ / Var$[\hat{\beta}^I]$; Incomplete data, truth = AR(1)

We can see that the independence estimator only starts doing badly when the correlation is quite high (say $\alpha = 0.7$). Even though there is incomplete data on half of the clusters, for fairly high correlations (say $\alpha = 0.5$ or 0.6) there is only a moderate decrease in efficiency.

Question 2:

The Madras Longitudinal Study collected monthly symptom data on schizophrenia patients after their initial hospitalisation. One scientific question of interest is whether subjects with an older age-at-onset tend to recover more/less quickly, and/or whether female subjects recover more/less quickly. Recovery is measured by a reduction in the presentation of symptoms.

For this analysis we will consider the outcome *thought disorders* which is a binary indicator of whether the patient was observed to present this "positive symptom" during the month.

The available data consist of 86 subjects each of which has a maximum of 12 measurements, month 0 (which is baseline) to month 11. Table 5 shows that 69 of the 86 subjects have complete data, while 3 subjects had fewer than 3 observations. For the purpose of this analysis the 3 subjects with either 1 or 2 observations were omitted. In addition, we assume that the missingness in the data is MCAR (although one would want to attempt to confirm this assumption).

Number of visits	1	2	3	4	5	7	8	10	11	12
Number of patients	1	2	3	2	2	1	3	2	1	69

Table 5: Number of patients for a given number of visits.

Table 6 provides a breakdown of the joint distribution of age and gender amoung the remaining 83 subjects at baseline and the 69 subjects that were followed to month 11.

		Gender	
		Female	Male
Age	< 20 years	17	12
	≥ 20 years	23	31

(a) Baseline

		Gender	
		Female	Male
Age	< 20 years	14	11
	≥ 20 years	17	23

(b) Month 11

Table 6: Joint distribution of age/gender indicators in the MADRAS data.

(a) Table 7 and Figure 3 provide the prevalance of symptoms for the four groups of interest. From Table 6(b) we see that each of the estimates in Table 7 are based on at least 11 subjects.

month	0	1	2	3	4	5	6	7	8	9	10	11
age ≥ 20 & male	62.5	71.9	61.3	55.2	35.7	39.3	28.6	28.6	17.9	14.3	11.1	14.8
age ≥ 20 & female	62.5	43.5	26.1	22.7	19.0	0.0	5.0	0.0	5.6	11.1	11.1	5.9
age < 20 & male	58.3	58.3	50.0	58.3	58.3	41.7	25.0	8.3	0.0	9.1	9.1	9.1
age < 20 & female	77.8	61.1	70.6	52.9	41.2	25.0	6.2	6.2	6.7	6.7	0.0	0.0

Table 7: Prevalance of symptoms over time for groups defined by age and gender.

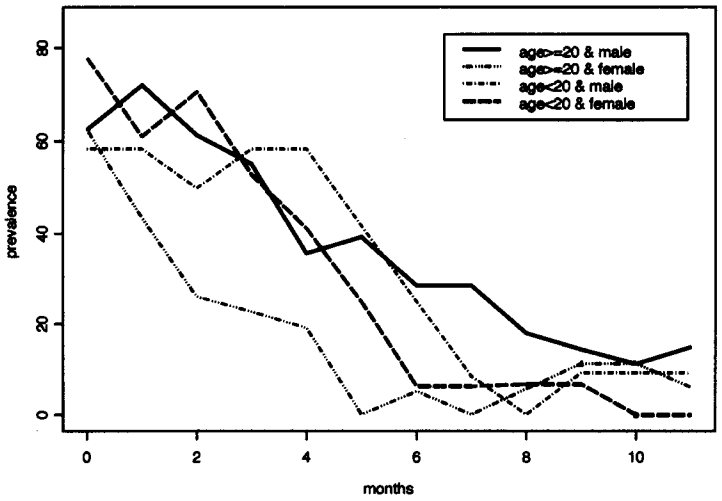

Figure 3: Prevalance of symptoms over time for groups defined by age and gender.

From both Table 7 and Figure 3 we find that there is a clear indication that for each of the four groups that the prevalance of symptoms is decreasing over time. There may be an indication that the females have a slightly steeper decline. It is not clear, however, that there are any differences between the age groups.

(b) The following provides all pairwise correlations for the outcome (binary indicator of thought disorder) across months. It seems that observations that are 1 month apart are fairly correlated, with estimates ranging from 0.48 to 0.91 but with the majority of estimates begin around 0.55-0.65. As observations move further apart the correlation goes to zero fairly quickly, suggesting that there may be a serial (autoregressive) component to the correlation structure of the data.

Correlation Matrix:

	mon 0	mon 1	mon 2	mon 3	mon 4	mon 5	mon 6	mon 7	mon 8	mon 9	mon 10	mon 11
mon 0	1	0.55	0.48	0.40	0.36	0.29	0.09	0.11	0.13	0.16	0.11	0.12
mon 1		1	0.62	0.52	0.33	0.36	0.30	0.15	0.06	0.09	0.14	0.24
mon 2			1	0.65	0.44	0.33	0.08	0.14	0.12	0.07	-0.02	0.08
mon 3				1	0.63	0.39	0.13	0.18	0.16	0.11	0.12	0.22
mon 4					1	0.53	0.16	0.11	0.14	0.19	0.20	0.20
mon 5						1	0.60	0.47	0.34	0.39	0.29	0.29
mon 6							1	0.55	0.21	0.29	0.38	0.38
mon 7								1	0.68	0.50	0.46	0.46
mon 8									1	0.48	0.58	0.45
mon 9										1	0.75	0.63
mon 10											1	0.91
mon 11												1

Pairwise odds ratios are provided below (see notes attached). In each case the value is the crude estimate of the ratio of odds of "disease" for a particular month relative to the odds of "disease" for a previous month. Each were computed by constructing a 2×2 table of the cross-classification of outcome at both times across subjects measured at both times. Some of the resulting 2×2 tables resulting in zero cells in the off-diagonals and so the odds ratios could not be computed.

We see that the odds of having a positive symptom at month 2 for an individual that had a positive outcome at month 1 are 20 times higher than the odds of a positive symptom at month 2 for an individuals that did not have a positive outcome at month 1.

```
Odds Ratio Matrix:

       mon 0 mon 1 mon 2 mon 3 mon 4 mon 5 mon 6 mon 7 mon 8 mon 9 mon 10 mon 11
 mon 0 Inf   13.14  9.25  6.70  7.10  6.27  1.79  2.19  3.29  3.93  2.80   2.88
 mon 1       Inf  20.04 11.95  4.98  9.00 10.24  2.84  1.56  1.92  3.42
 mon 2             Inf  24.18  8.18  5.33  1.55  2.41  2.42  1.57  0.88   1.82
 mon 3                   Inf  26.39  7.20  2.00  3.08  3.10  2.01  2.41   6.25
 mon 4                         Inf  14.67  2.33  1.95  2.61  3.41  4.10   4.00
 mon 5                               Inf  33.00 17.67  9.11 11.77  7.14   7.00
 mon 6                                     Inf  22.94  4.12  6.11 12.22  12.00
 mon 7                                           Inf  91.50 19.67 19.33  19.00
 mon 8                                                 Inf  20.33 40.67  20.00
 mon 9                                                       Inf 155.00  61.00
mon 10                                                             Inf
mon 11                                                                   Inf
```

It is clear from the above that at any given month the odds of a positive symptom are greatly increased for individuals with positive symptoms in the previous months. The odds ratio estimates range from around 13 to 155. There is also evidence that the lag extends further than one month. The odds ratio estimates comparing observations that are 2 months apart range from 2.33 to 61, suggesting fairly strong dependence on previous outcomes. At longer lags, although the odds ratio estimates are still (fairly) high we see that they are not as high as closer lags. This is again evidence that is suggestive of a serial component in the correlation structure.

(c) The current data are repeated measures binary data. Consequently, assuming the missingness is MCAR, we can perform regression within the GEE framework (see attached S-Plus 2000 code). Given the EDA in the previous part, we fit two models: (a) assuming an independence working correlation structure and (b) assuming a autoregressive lag 1 (AR(1)) working correlation. In both models we include main effects for time, age and gender. The questions of interest are addressed by 2 interaction terms which allow the `month` coefficient (i.e. rate over time) to depend on age and gender.

Independence:

	Estimate	Naive S.E.	Naive z	Robust S.E.	Robust z
(Intercept)	0.620	0.202	3.062	0.308	2.013
age	0.811	0.306	2.652	0.496	1.634
gender	-0.406	0.287	-1.415	0.452	-0.898
month	-0.251	0.037	-6.691	0.059	-4.236
age:month	-0.137	0.063	-2.156	0.094	-1.463
gender:month	-0.109	0.063	-1.734	0.095	-1.140

```
Estimated Scale Parameter:  1.054699
Working Correlation:        0
```

Although not shown here, a three way interaction term between time, age and gender was included in the model (for both correlation structures) but found not to be statistically significant. The inclusion of the three way interaction would allow the rate of change of prevalence to be unconstrained across the four groups of interest.

Autoregressive lag 1:

```
              Estimate Naive S.E. Naive z Robust S.E. Robust z
(Intercept)      0.518      0.296   1.747       0.294    1.761
        age      0.649      0.440   1.475       0.464    1.399
     gender     -0.200      0.412  -0.484       0.424   -0.471
      month     -0.231      0.053  -4.359       0.055   -4.210
  age:month     -0.104      0.089  -1.170       0.085   -1.220
gender:month    -0.142      0.089  -1.594       0.089   -1.599

Estimated Scale Parameter:  1.024654
Working Correlation:        0.581028

NOTE: gee() runs into problems fitting this correlation structure if you do not remove the individual with
only one observation in the dataset (see Table 5).
```

Comparing the two models we find that point estimates and robust standard error estimates are fairly close to eachother. For the AR(1) model there seems to be a fairly close correspondence between the naive and robust standard error estimates, suggesting that the AR(1) structure is fairly close to the true underlying structure. We don't see this correspondence in the independence model.

From the output for the model assuming an AR(1) structure it is clear that both being younger than 20 years old (**age** = 1) and being female (**gender** = 1) are associated with steeper declines in the rate of change of symptom prevalance. This is consistent with what we saw from Figure 3 in part (a).

To statistically assess the two interaction terms we can use a joint Wald test. The resulting test statistic is 5.42, which when compared to a χ^2 distribution with 2 degrees of freedom yields a p-value of 0.066. Thus, the data do provide marginal evidence to suggest that the decline in symptoms may depend on age and gender.

Pairwise Odds Ratios: One way of summarizing the association between two binary variables is via an odds ratio. Here the two binary variables are presence/absence of symptoms in month i (Y_i) and presence/absence of symtoms in month j (Y_j) , where without loss of generality assume that $i < j$. For these two binary indicators the pairwise odds ratio is defined as follows:

$$\begin{aligned} \text{OR} &= \frac{\text{Odds of symptoms during month } j \mid \text{symptoms at month } i}{\text{Odds of symptoms during month } j \mid \text{no symptoms at month } i} \\ &= \frac{P(Y_j = 1|Y_i = 1)}{P(Y_j = 0|Y_i = 1)} \Big/ \frac{P(Y_j = 1|Y_i = 0)}{P(Y_j = 0|Y_i = 0)}. \end{aligned}$$

This quantity can be computed as $\widehat{OR} = \frac{ad}{bc}$ from the following cross-classification of the two outcomes at months i and j:

		Month j	
		$Y_j = 1$	$Y_j = 0$
Month i	$Y_i = 1$	a	b
	$Y_i = 0$	c	d

S-Plus Code:

```
#### MADRAS schizophrenia symptom data: symptom = thought disorders
####     ID     = patient ID
####     Y      = symptom indicator
####     MONTH  = month since hospitalization
####     AGE    = age-at-onset (1= age<20 ; 0= age>=20)
####     GENDER = gender (1=female; 0=male)
####     MONTH*AGE
####     MONTH*GENDER
```

```
#
#### Splus 2000 ####
#
library( gee )
options( constrasts = c("contr.treatment", "contr.poly") )
source( "C:\\TA\\571_W03\\homework\\exercise8\\code\\stacked2wide.q" )
madras <- read.table( "C:\\TA\\571_W03\\homework\\exercise8\\code\\madras_data.txt" )[,1:5]      # Exclude interactions in baseline dataset
names( madras ) <- c( "id", "y", "month", "age", "gender" )
attach( madras )
#
#### Number of subjects with corresponding number of observations
sum(table( table ( id ) ))
#
#### Drop individuals with less than 3 observations
drop.names <- as.numeric( names( table( id )[table( id ) < 3 ] ) )
madras$y[is.element( madras$id, drop.names )] <- NA
madras <- na.omit( madras )
#
#### Number of subjects that are female/male and young/old
table( madras$age[madras$month == 0], madras$gender[madras$month == 0] )
table( madras$age[madras$month == 11], madras$gender[madras$month == 11] )
#
#### Part (a) ####
#
prev.table <- matrix( 0, 4, 12 )
for( i in 0:1 ){
  for( j in 0:1 ){
    temp.table <- table( y[age == i & gender == j], month[age == i & gender == j] )
    prev.table[((2*i)+j+1),] <- temp.table[2,] / apply( temp.table, 2, sum ) * 100
  }
}
dimnames( prev.table ) <- list( c("age>=20 & male", "age>=20 & female", "age<20 & male", "age<20 & female"), paste("mon",0:11) )
round( prev.table, 1 )
plot( 0:11, prev.table[1,], xlab = "months", ylab = "prevalence", type = "n", ylim = range( prev.table ) * 1.1 )
lines( 0:11, prev.table[1,], col = 1, lwd = 3 )
lines( 0:11, prev.table[2,], col = 4, lwd = 3, lty = 5 )
lines( 0:11, prev.table[3,], col = 6, lwd = 3, lty = 3 )
lines( 0:11, prev.table[4,], col = 8, lwd = 3, lty = 4 )
legend( 6.5, 83, col = c(1,4,6,8), lwd = c(3,3,3,3), lty = c(1,5,3,4),
     c("age>=20 & male", "age>=20 & female", "age<20 & male", "age<20 & female") )
#
#### Part (b) ####
#
stacked <- stacked2wide( id, y, month, 0:11, 0.5 )
cor.mat <- or.mat <- num.mat <- matrix( 0, 12, 12 )
for( i in 1:11 ){
  for( j in i:12 ){
    temp.stacked <- na.omit( stacked[,c(i,j)] )
    month.i <- temp.stacked[,1]
    month.j <- temp.stacked[,2]
    #
    cor.mat[i,j] <- cor( month.i, month.j )
    table.ij <- table( month.i, month.j )
    or.mat[i,j] <- (table.ij[2,2] * table.ij[1,1]) / (table.ij[1,2] * table.ij[2,1])
    num.mat[i,j] <- length( month.i )
  }
}
dimnames( cor.mat ) <- dimnames( or.mat ) <- dimnames( num.mat ) <- list( paste("mon", 0:11), paste("mon", 0:11) )
round( cor.mat, 2 )
round( or.mat, 2 )
num.mat
#
#### Part (c)
#
fit.0 <- gee( y ~ age + gender + month + age:month + gender:month,
        id = id, data = madras,
        family = binomial )
round( summary(fit.0)$coef, 3 )
#
fit.1 <- gee( y ~ age + gender + month + age:month + gender:month,
        id = id, data = madras,
        family = binomial,
        corstr = "AR-M" )
round( summary(fit.1)$coef, 3 )
#
interaction.est <- fit.1$coef[5:6]
interaction.var <- fit.1$robust.variance[5:6,5:6]
#
test.stat <- t(interaction.est) %*% solve(interaction.var) %*% interaction.est     ## 5.42192
1 - pchisq( test.stat, 2 )                                                         ## 0.06647298
```

Question 3:

Consider the GLMM for count data given by:

$$Y_{ij}|\mathbf{X}_i, b_{i,0} \sim \text{Poisson}$$

$$\log E[Y_{ij}|\mathbf{X}_i, b_{i,0}] = \beta_0 + \beta_1 X_{ij} + b_{i,0}$$

$$b_{i,0} \sim \mathcal{N}(0, \sigma^2)$$

Throughout we assume that the index i is referring to an individual subject, and that repeated measures are taken within the individuals. For example, we might think of the covariate X_{ij} as a time covariate.

(a) Using the moment generating function for the normal distribution we can compute the marginal expectation as follows:

$$\begin{aligned}
\mu_{ij} = E[Y_{ij}|\mathbf{X}_i] &= E(E[Y_{ij}|\mathbf{X}_i, b_{i,0}]) \\
&= E(e^{\beta_0+\beta_1 X_{ij}+b_{i,0}}) \\
&= e^{\beta_0+\beta_1 X_{ij}} E(e^{b_{i,0}}) \\
&= e^{\beta_0+\beta_1 X_{ij}} e^{\sigma^2/2}, \qquad M_{b_{i,0}}(1) = E(e^{b_{i,0}}) = e^{\sigma^2/2} \\
&= e^{\beta_0+\beta_1 X_{ij}+\sigma^2/2}
\end{aligned}$$

(b) β_1 is the change in the log-rate per a 1 unit change in X_{ij} holding $b_{i,0}$ constant. Thus we would expect an e^{β_1}-fold increase in the rate per a 1 unit increase in X_{ij} within an individual.

The interpretation is a *within*-subject interpretation since we are holding the subject-specific random effect $b_{0,i}$ constant. This is in contrast to the usual interpretation of (marginal) regression parameters, which refer to comparisons between populations (defined by covariate values).

(c) In the random effects model we have

$$E[Y_{ij}|\mathbf{X}_i] = e^{\beta_0+\beta_1 X_{ij}+\sigma^2/2}$$

while with marginal model we have

$$E[Y_{ij}|\mathbf{X}_i] = e^{\beta_0^*+\beta_1^* X_{ij}}$$

Thus by fitting the marginal model we reparameterize $\beta_0 + \sigma^2$ as β_0^* and β_1 as β_1^*. Consequently, the resulting estimator for β_0^*, say $\hat{\beta}_0^*$, will not be consistent for β_0 but the estimator for β_1^*, $\hat{\beta}_1^*$, will be consistent for β_1.

(d) Now consider the model that contains random intercepts and random slopes:

$$\begin{aligned} Y_{ij}|\mathbf{X}_i, b_{i,0}, b_{i,1} &\sim \text{Poisson} \\ \log E[Y_{ij}|\mathbf{X}_i, b_{i,0}, b_{i,1}] &= \beta_0 + \beta_1 X_{ij} + b_{i,0} + b_{i,1} X_{ij} \\ b_{i,0} &\sim \mathcal{N}(0, \sigma_0^2) \\ b_{i,1} &\sim \mathcal{N}(0, \sigma_1^2) \end{aligned}$$

where $b_{i,0}$ and $b_{i,1}$ are assumed independent.

Consider a unit increase in X_{ij}, holding $b_{i,0}$ and $b_{i,1}$ constant. For this contrast we would find that the resulting difference in log rates is:

$$\log E[Y_{ij}|X_{ij} = x, b_{i,0}, b_{i,1}] - \log E[Y_{ij}|X_{ij} = x, b_{i,0}, b_{i,1}] = \beta_1 + b_{i,1}.$$

To interpret β_1 we therefore need to set $b_{i,1}$ equal to zero. In the above model setting $b_{i,1} = 0$ corresponds to setting the random slope to be equal to the average of its' distribution. Since random effects are subject-specific we can think of setting the random slope equal to zero as corresponding to a subjects' average or "typical" value. Therefore, β_1 represents the **average** change in the log(rate) associated with a 1 unit increase in X_{ij} within an individual, i.e. holding $b_{i,0}$ constant.

The standard deviation of the random intercept distribution, σ_0, represents the heterogeneity among individuals at baseline (i.e when X_{ij} equals zero). That is if σ_0 is large then individuals are very heterogeneous, ie. some individuals simply start with a lower log(rate) and some start with a higher log(rate).

The standard deviation of the random slope distribution, σ_1, represents the heterogeneity in the change in the log(rate) per a 1 unit change in X between individuals. Thus if σ_1 is large, then the effect of changing X on the incident rate [log(rate)] varies greatly from individual to individual.

(e) Using the same approach as in part (a), we can derive the marginal mean for this condition model:

$$\begin{aligned} \mu_{ij} = E[Y_{ij}|\mathbf{X}_i] &= E(E[Y_{ij}|\mathbf{X}_i, b_{i,0}, b_{i,1}]) \\ &= E(e^{\beta_0 + \beta_1 X_{ij} + b_{i,0} + b_{i,1} X_{ij}}) \\ &= e^{\beta_0 + \beta_1 X_{ij}} E(e^{b_{i,0}}) E(e^{b_{i,1} X_{ij}}) \\ &= e^{\beta_0 + \beta_1 X_{ij}} e^{\sigma_0^2/2} e^{X_{ij}^2 \sigma_1^2/2} \\ &= e^{\beta_0 + \beta_1 X_{ij} + \sigma_0^2/2 + X_{ij}^2 \sigma_1^2/2} \end{aligned}$$

Question 4:

One numerical method used to obtain likelihood estimates for GLMMs is known as *Gauss-Hermite* integration. Gauss-Hermite integration uses "quadrature" points to numerically evaluate an integral. Consider the following conditional regression specification:

$$\begin{aligned} \text{logit}\left\{\text{E}[Y_{ij}|\vec{X}_i,\ b_{i,0}]\right\} &= \beta_0 + \beta_1 X_{ij} + b_{i,0} \\ b_{i,0} &\sim \mathcal{N}(0,\ \sigma^2). \end{aligned}$$

Note that the above is equivalent to the following:

$$\begin{aligned} \text{E}[Y_{ij}|\vec{X}_i,\ b_{i,0}] &= \frac{\exp[\beta_0 + \beta_1 X_{ij} + b_{i,0}]}{1 + \exp[\beta_0 + \beta_1 X_{ij} + b_{i,0}]} \\ b_{i,0} &\sim \mathcal{N}(0,\ \sigma^2). \end{aligned}$$

We would like to compute the induced marginal regression parameter. That is, we would like to compute β_0^* and β_1^* from the following marginal model that is induced by the above conditional model:

$$\text{logit}\left\{\text{E}[Y_{ij}|\vec{X}_i]\right\} = \beta_0^* + \beta_1^* X_{ij}.$$

As above, we can write down the following equivalent specification:

$$\text{E}[Y_{ij}|\vec{X}_i] = \frac{\exp[\beta_0^* + \beta_1^* X_{ij}]}{1 + \exp[\beta_0^* + \beta_1^* X_{ij}]}.$$

Notice here that, given a single covariate $X_{ij} = 0/1$, for the above marginal model β_0^* and β_1^* are given by the following:

$$\beta_0^* = \log\left[\frac{\text{E}[Y_{ij}|X_{ij}=0]}{1-\text{E}[Y_{ij}|X_{ij}=0]}\right] \quad \text{and} \quad \beta_1^* = \log\frac{\left[\frac{\text{E}[Y_{ij}|X_{ij}=1]}{1-\text{E}[Y_{ij}|X_{ij}=1]}\right]}{\left[\frac{\text{E}[Y_{ij}|X_{ij}=0]}{1-\text{E}[Y_{ij}|X_{ij}=0]}\right]}$$

To compute the marginal parameters we need to integrate out the random effects:

$$\text{E}[Y_{ij}|\vec{X}_i] = \int_{-\infty}^{\infty} \text{E}[Y_{ij}|\vec{X}_i,\ b_{i,0}]\ f_b(b_{i,0}|\sigma^2)\ \partial b_{i,0}$$

where $f_b(b_{i,0}|\sigma^2)$ is the density of the normal distribution for the random effects $b_{i,0}$. Guass-Hermite integration allows us to evaluate the above integral by choosing a number of points to be used $(z_1, ..., z_K)$, and then approximating the integral with a weighted sum:

$$\begin{aligned} \text{E}[Y_{ij}|\vec{X}_i] &\approx \sum_{i=1}^{K} \text{E}[Y_{ij}|\vec{X}_i,\ z_i] \cdot w_i \\ &= \sum_{i=1}^{K} \left\{\frac{\exp[\beta_0 + \beta_1 X_{ij} + z_i]}{1 + \exp[\beta_0 + \beta_1 X_{ij} + z_i]}\right\} \cdot w_i, \end{aligned}$$

where $w_i = f_b(z_i|\sigma = 1)$. To incoporate arbitrary variances for the random effects it is sufficient to multiply the quadrature points $\{z_i\}$ by σ before the summation.

For further details see Diggle, Liang, Heagerty, and Zeger (2002) pages 212-213.

We can therefore use these sums to numerically calculate β_0^* and β_1^* via the expressions given above. Using the functions that were provided, we can obtain $K = 20$ quadrature points and their associated weights. The table below provides the induced marginal parameter values for a range of values for σ^2:

		Intercept	Slope
Conditional		-2	1
Marginal	$\sigma = 0.5$	-1.910	0.962
	$\sigma = 1.0$	-1.692	0.861
	$\sigma = 1.5$	-1.450	0.738

We can see that as the heterogeneity increases (i.e. σ increases) then the disparity between the conditional and marginal parameter also increases.

S-Plus Code:

```
#
#### Gauss-Hermite integration ####
#
source( "gauss.hermite.q" )
#
### Returns the value of the expit function, given a random effect:
g.b <- function( x, b ){
    value <- exp(-2 + (1 * x) + b) / (1 + exp(-2 + (1 * x) + b))
    return( value )
}
#
sigma <- 1.5
#
### Get the quadrature points and weights
quad.points <- get.GH.z( 20 ) * sigma
quad.weights <- get.GH.w( 20 )
#
### Compute the sums for the two cases; when X = 0 and when X = 1
sum.0 <- g.b( x = 0, b = quad.points ) %*% quad.weights
sum.1 <- g.b( x = 1, b = quad.points ) %*% quad.weights
#
### Compute odds for the two cases
odds.0 <- sum.0 / (1 - sum.0)
odds.1 <- sum.1 / (1 - sum.1)
#
### Compute parameters
beta.0 <- log( odds.0 )
beta.1 <- log( odds.1 / odds.0 )
#
### Output
round( c( beta.0, beta.1 ), 3 )
```