Practice 11 Solutions

You need `R` packages `gee` and `MASS` for doing this practice. The question focuses on analysing the `OME` data stored in `MASS`. Type `help(OME)` and `help(gee)` to get details of the data and GEE solver.

The following commands are used to get the analysis results.

```
head(OME)
fm <- gee(cbind(Correct, Trials - Correct) ~ Loud + Age + OME + Noise, id = ID,
          data = OME, family = binomial, corstr = "exchangeable")
sfm <- summary(fm)
attributes(fm); attributes(sfm)
```

1. How many children were included in the data?

   - 77 children were included in the study.

   ```
   length(table(OME$ID))    #[1] 77
   head(OME)

     ID Age OME Loud      Noise Correct Trials
   1  1  30 low   35   coherent       1      4
   2  1  30 low   35 incoherent       4      5
   3  1  30 low   40   coherent       0      3
   4  1  30 low   40 incoherent       1      1
   5  1  30 low   45   coherent       2      4
   6  1  30 low   45 incoherent       2      2

   tail(OME)

         ID Age OME Loud      Noise Correct Trials
   1092 100  18 N/A   45   coherent       1      2
   1093 100  18 N/A   45 incoherent       3      4
   1094 100  18 N/A   50   coherent       7      7
   1095 100  18 N/A   50 incoherent       3      3
   1096 100  18 N/A   55   coherent       4      5
   1097 100  18 N/A   55 incoherent       5      5
   ```

2. How many tests did the child with `ID=1` attend to in the data? Namely, what is the cluster size associated with child 1?

   - 20 tests were done for the child with `ID=1`.

   ```
   sum(OME$ID==1)    #[1] 20
   ```

3. Let y_{it} and n_{it} be the number of correct responses and the number of trials, respectively, for the child with `ID=i` at test t. Define $y^*_{it} = y_{it}/n_{it}$. Write down the marginal model fitted by the GEE approach in this analysis. The model should include the following components: marginal means of y^*_{it}, marginal variances of y^*_{it}, and correlation between y^*_{ij} and y^*_{ik}.

- $E(y^*_{ij}|\cdot) = \dfrac{\exp(\eta_{ij})}{1+\exp(\eta_{ij})}$ with

$$\begin{aligned}\hat{\eta}_{ij} &= -6.954 + 0.167\cdot\text{Loud} + 0.02\cdot\text{Age} - 0.097\cdot\text{OME.high}\\ &\quad -0.320\cdot\text{OME.low} + 1.312\cdot\text{Noise.incoherent}.\end{aligned}$$

- $\text{Var}(y^*_{ij}|\cdot) = \dfrac{\phi}{n_{it}}\cdot\dfrac{\exp(\eta_{ij})}{[1+\exp(\eta_{ij})]^2}$ with $\hat{\phi} = 1.116$ and $\hat{\eta}_{ij}$ as above.
- $\text{Corr}(y^*_{ij}, y^*_{ik}) = \alpha$ is assumed in this analysis, i.e. exchangeable correlation structure. $\hat{\alpha} = -0.01001777$.

```
summary(fm)

 Coefficients:
                   Estimate  Naive S.E.     Naive z Robust S.E.    Robust z
(Intercept)     -6.95403269 0.344815565 -20.1673978 0.244535253 -28.4377512
Loud             0.16684213 0.007509126  22.2185817 0.005460094  30.5566395
Age              0.02003283 0.003699210   5.4154354 0.003425720   5.8477724
OMEhigh         -0.09666362 0.168193526  -0.5747167 0.156324984  -0.6183504
OMElow          -0.32041010 0.133828827  -2.3941785 0.122289621  -2.6200923
Noiseincoherent  1.31206035 0.095810530  13.6943230 0.098468293  13.3246988

Estimated Scale Parameter:  1.116139       Number of Iterations:  3
Working Correlation
             [,1]        [,2]        [,3]        [,4]        [,5]        [,6]
 [1,]  1.00000000 -0.01001777 -0.01001777 -0.01001777 -0.01001777 -0.01001777
...

fm$robust.variance

                  (Intercept)          Loud           Age       OMEhigh
(Intercept)      5.979749e-02 -1.251861e-03 -8.942811e-05 -9.138445e-04
Loud            -1.251861e-03  2.981263e-05 -1.964369e-06  4.984232e-05
Age             -8.942811e-05 -1.964369e-06  1.173556e-05 -3.160966e-04
OMEhigh         -9.138445e-04  4.984232e-05 -3.160966e-04  2.443750e-02
OMElow          -4.739901e-04  2.106300e-05 -2.950897e-04  1.191210e-02
Noiseincoherent -5.811760e-03  8.403785e-05 -1.016171e-06 -1.114490e-03
                       OMElow Noiseincoherent
(Intercept)     -0.0004739901   -5.811760e-03
Loud             0.0000210630    8.403785e-05
Age             -0.0002950897   -1.016171e-06
OMEhigh          0.0119121013   -1.114490e-03
OMElow           0.0149547515   -1.079979e-03
Noiseincoherent -0.0010799793    9.696005e-03
```

4. Give the commands for finding the estimates and their robust variance matrix for the regression parameters in the marginal model.

- `fm$robust.variance` gives

	Intercept	Loud	Age	OME.high	OME.low	Noise.incoherent
Intercept	0.0598	-0.00125	-0.0000894	-0.000914	-0.000474	-0.00581
Loud		0.0000298	-0.0000196	-0.0000498	0.000021	0.000084
Age			0.0000117	-0.000316	-0.000295	-0.00000102
OME.high				0.0244	0.0119	-0.00111
OME.low					0.01495	-0.00108
Noise.incoherent						0.00967

5. Use the R output to estimate the mean and variance of y_{11}, and the covariance between y_{11} and y_{12}.

- $\hat{\eta}_{11} = (-6.954, 0.167, 0.02, -0.097, -0.32, 1.312) \cdot (1, 35, 30, 0, 1, 0)^T = -0.834$.
- $\hat{E}(y_{11}^*) = \frac{e^{-0.834}}{1+e^{-0.834}} = 0.3028$. $n_{11} = 4$. So $\hat{E}(y_{11}) = 4 \times 0.3028 = 1.211214$.
- $\widehat{\text{Var}}(y_{11}) = n_{11}\hat{\phi}\hat{E}(y_{11}^*)[1 - \hat{E}(y_{11}^*)] = 4 \times 1.116 \times 0.3028 \times (1 - 0.3028)$
  $= 0.8445 \times 1.1161 = 0.9425282$.
- $$\begin{aligned} \widehat{\text{Cov}}(y_{11}, y_{12}) &= \sqrt{\widehat{\text{Var}}(y_{11})} \cdot \sqrt{\widehat{\text{Var}}(y_{12})} \cdot \widehat{\text{Corr}}(y_{11}, y_{12}) \\ &= \sqrt{0.9425282} \cdot \sqrt{1.318396} \cdot (-0.01001777) = -0.01116712 \end{aligned}$$

  where $\widehat{\text{Var}}(y_{12}) = n_{12}\hat{\phi}\hat{E}(y_{12}^*)[1 - \hat{E}(y_{12}^*)] = 5 \times 1.116 \times 0.6173 \times (1 - 0.6173) = 1.318396$.