Homework 3

Mairead Dillon

2023-02-04

# (a) Which variables are between-cluster variables, and which are within-cluster variables?

The between-cluster variables are id, sex, and baseage. The within-cluster variables are resp, age, xp, time, and season.

# (b) Fit a logistic regression model with resp as the outcome variable and predictors age, xp and sex. Interpret the age effect.

For every one month increase in age, there is a 2.7% decrease in the odds that the patient has a respiratory infection when holding all other variables constant.

# Logistic regression  
mod1 <- glmer(resp ~ age + xp + sex + (1 | id), data=resp, family="binomial")  
summary(mod1)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: resp ~ age + xp + sex + (1 | id)  
## Data: resp  
##   
## AIC BIC logLik deviance df.resid   
## 696.6 722.1 -343.3 686.6 1193   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -0.8228 -0.3022 -0.2313 -0.1782 4.9331   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## id (Intercept) 0.7915 0.8897   
## Number of obs: 1198, groups: id, 275  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.504964 0.199341 -12.566 < 2e-16 \*\*\*  
## age -0.027179 0.006648 -4.088 4.35e-05 \*\*\*  
## xp 0.466198 0.488367 0.955 0.340   
## sex -0.426960 0.261595 -1.632 0.103   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) age xp   
## age 0.155   
## xp -0.105 -0.104   
## sex -0.424 0.017 0.089

# (c) Expand the previous model to include between-cluster and within-cluster age effects. Please specify your model and indicate how you constructed your predictors for age. You do not need to fit the new model in part (c).

There are two age predictors - one for the between-cluster effect and one for the within-cluster effect. The between cluster predictor is multiplied by the mean of age, while the within-cluster predictor is multiplied by age minus the mean of age.

# (d) Now fit a logistic regression model with baseline age, change in age from baseline, xp and sex. Give careful interpretations of the two age effects. Contrast them to those in part (b).

For every one month increase in baseline age, the odds of a patient having a respiratory infection decrease by 3.1% when holding all other variables constant. For every one month increase in change in age from baseline, the odds of a patient having a respiratory infection increase by 1.2%. An one month increase in change in baseline age in this model causes a slightly greater decrease in the odds of a patient having a respiratory infection than the same change in the age variable in the previous model. An one month increase in change in age from baseline in this model causes an increase in the odds of a patient having a respiratory infection, which contrasts to the decrease caused by an one month increase in age in the previous model.

# Create change in age from baseline variable  
resp <- resp %>%  
 mutate(change\_age = age - baseage)  
  
# Logistic regression  
mod2 <- glmer(resp ~ baseage + change\_age + xp + sex + (1 | id), data=resp, family="binomial")  
summary(mod2)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: resp ~ baseage + change\_age + xp + sex + (1 | id)  
## Data: resp  
##   
## AIC BIC logLik deviance df.resid   
## 694.5 725.1 -341.3 682.5 1192   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -0.7519 -0.3101 -0.2212 -0.1700 5.3335   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## id (Intercept) 0.8622 0.9286   
## Number of obs: 1198, groups: id, 275  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -2.854623 0.275755 -10.352 < 2e-16 \*\*\*  
## baseage -0.031766 0.007303 -4.350 1.36e-05 \*\*\*  
## change\_age 0.012037 0.020508 0.587 0.557   
## xp 0.524484 0.493716 1.062 0.288   
## sex -0.422844 0.266618 -1.586 0.113   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) baseag chng\_g xp   
## baseage 0.365   
## change\_age -0.596 -0.005   
## xp -0.116 -0.102 0.026   
## sex -0.311 0.018 0.007 0.087

# (e) Fit the models in (b) and (d), but this time including season effects (treat it as a categorical predictor). Give explanations for any large changes in the age effects you observe when season is added to the models compared to your previous results. Use some descriptive statistics or graphics to support your explanations.

When season is added to the model in (b), the change in odds of a patient having respiratory infection is about the same for an one month increase in age when all other variables are held constant. When season is added to the model in (c), the change in odds of a patient having a respiratory infection is about the same for an one month increase in baseline age when all other variables are held constant. In this same model, the odds of a patient having a respiratory infection now increase by 3.8% when there is an one month increase in change in age from baseline and all other variables are held constant, which is greater than the odds increase of 1.2% seen in (c). This change in the effect of change in age from baseline on the odds of a patient having respiratory disease is unsurprising because the change in age from baseline greatly varies based on season, as seen in the table below.

# Refit model from part b  
mod3 <- glmer(resp ~ age + xp + sex + as.factor(season) + (1 | id), data=resp, family="binomial")

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 0.00691893 (tol = 0.002, component 1)

summary(mod3)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: resp ~ age + xp + sex + as.factor(season) + (1 | id)  
## Data: resp  
##   
## AIC BIC logLik deviance df.resid   
## 683.6 724.3 -333.8 667.6 1190   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -1.0084 -0.2959 -0.2215 -0.1579 5.6054   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## id (Intercept) 0.8005 0.8947   
## Number of obs: 1198, groups: id, 275  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -3.429310 0.437357 -7.841 4.47e-15 \*\*\*  
## age -0.025866 0.006649 -3.890 0.000100 \*\*\*  
## xp 0.642884 0.491296 1.309 0.190688   
## sex -0.428125 0.263484 -1.625 0.104192   
## as.factor(season)2 1.406314 0.425566 3.305 0.000951 \*\*\*  
## as.factor(season)3 0.658900 0.444676 1.482 0.138406   
## as.factor(season)4 0.626898 0.495987 1.264 0.206251   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) age xp sex as.()2 as.()3  
## age 0.046   
## xp -0.117 -0.127   
## sex -0.193 0.016 0.086   
## as.fctr(s)2 -0.867 0.017 0.076 -0.007   
## as.fctr(s)3 -0.815 -0.005 0.066 0.007 0.833   
## as.fctr(s)4 -0.720 0.036 -0.007 -0.006 0.738 0.706  
## optimizer (Nelder\_Mead) convergence code: 0 (OK)  
## Model failed to converge with max|grad| = 0.00691893 (tol = 0.002, component 1)

# Refit model from part c  
mod4 <- glmer(resp ~ baseage + change\_age + xp + sex + as.factor(season) + (1 | id), data=resp, family="binomial")

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :  
## Model failed to converge with max|grad| = 0.075193 (tol = 0.002, component 1)

summary(mod4)

## Generalized linear mixed model fit by maximum likelihood (Laplace  
## Approximation) [glmerMod]  
## Family: binomial ( logit )  
## Formula: resp ~ baseage + change\_age + xp + sex + as.factor(season) +   
## (1 | id)  
## Data: resp  
##   
## AIC BIC logLik deviance df.resid   
## 675.0 720.8 -328.5 657.0 1189   
##   
## Scaled residuals:   
## Min 1Q Median 3Q Max   
## -0.9231 -0.2919 -0.2057 -0.1419 6.2087   
##   
## Random effects:  
## Groups Name Variance Std.Dev.  
## id (Intercept) 0.9346 0.9668   
## Number of obs: 1198, groups: id, 275  
##   
## Fixed effects:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) -4.139306 0.502549 -8.237 < 2e-16 \*\*\*  
## baseage -0.033532 0.007522 -4.458 8.27e-06 \*\*\*  
## change\_age 0.037019 0.020805 1.779 0.075188 .   
## xp 0.772116 0.502029 1.538 0.124051   
## sex -0.420907 0.272937 -1.542 0.123040   
## as.factor(season)2 1.649912 0.434643 3.796 0.000147 \*\*\*  
## as.factor(season)3 0.678999 0.449212 1.512 0.130653   
## as.factor(season)4 0.813870 0.503895 1.615 0.106276   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation of Fixed Effects:  
## (Intr) baseag chng\_g xp sex as.()2 as.()3  
## baseage 0.249   
## change\_age -0.416 -0.029   
## xp -0.149 -0.133 0.033   
## sex -0.170 0.020 0.000 0.083   
## as.fctr(s)2 -0.825 -0.061 0.140 0.099 -0.007   
## as.fctr(s)3 -0.707 -0.019 -0.033 0.080 0.006 0.819   
## as.fctr(s)4 -0.680 -0.010 0.122 0.006 -0.006 0.740 0.695  
## optimizer (Nelder\_Mead) convergence code: 0 (OK)  
## Model failed to converge with max|grad| = 0.075193 (tol = 0.002, component 1)

|  | Season 1 (N=183) | Season 2 (N=424) | Season 3 (N=414) | Season 4 (N=177) | Overall (N=1198) |
| --- | --- | --- | --- | --- | --- |
| **Change in Age from Baseline (Months)** |  |  |  |  |  |
| 0 | 0 (0%) | 229 (54.0%) | 0 (0%) | 0 (0%) | 229 (19.1%) |
| 3 | 0 (0%) | 0 (0%) | 213 (51.4%) | 0 (0%) | 213 (17.8%) |
| 6 | 0 (0%) | 0 (0%) | 0 (0%) | 176 (99.4%) | 176 (14.7%) |
| 9 | 183 (100%) | 0 (0%) | 0 (0%) | 1 (0.6%) | 184 (15.4%) |
| 12 | 0 (0%) | 195 (46.0%) | 0 (0%) | 0 (0%) | 195 (16.3%) |
| 15 | 0 (0%) | 0 (0%) | 201 (48.6%) | 0 (0%) | 201 (16.8%) |