STATISTICAL MODELLING   Part IIC, Michaelmas 2019
Practical 6: Binomial regression   By courtesy of Dr. S. A. Bacallado and Dr. R. Shah

Solutions

In the log-likelihood ratio test comparing the model `MyopiaLogReg1` to the intercept only model, the p-value is 2.6×10^{-6} which is extremely small, so we should reject the null model with only an intercept.

Exercise 1

```
anova(MyopiaLogReg1,MyopiaLogReg2,test="LR")
```

```
## Analysis of Deviance Table
##
## Model 1: myopic ~ gender + sportHR + readHR + compHR + studyHR + TVHR +
##     mumMyopic + dadMyopic
## Model 2: myopic ~ gender + sportHR + readHR + compHR + studyHR + TVHR +
##     mumMyopic + dadMyopic + mumMyopic:dadMyopic
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       609     439.60
## 2       608     438.01  1   1.5918   0.2071
```

This test cannot reject the simpler model without interactions between `mumMyopic` and `dadMyopic`.

Exercise 2

```
MyopiaLogReg3 <- glm(myopic ~ . -compHR - TVHR,
                     data = Myopia, family = binomial)
anova(MyopiaLogReg3,MyopiaLogReg1,test="LR")
```

```
## Analysis of Deviance Table
##
## Model 1: myopic ~ (gender + sportHR + readHR + compHR + studyHR + TVHR +
##     mumMyopic + dadMyopic) - compHR - TVHR
## Model 2: myopic ~ gender + sportHR + readHR + compHR + studyHR + TVHR +
##     mumMyopic + dadMyopic
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1       611     440.46
## 2       609     439.60  2  0.86043   0.6504
```

The hours of computer use and TV watching don't seem to be collectively significant.

Exercise 3

If we include the variable `mumPlusdadMyopic` in the model `ModLogReg3`, the column space of the design matrix does not change, since this variable is the sum of two variables already in the model. Therefore, the fitted values shouldn't change. To make the design of full rank, we must impose a corner point constraint, and we shall require that the coefficient for `dadMyopic` is 0. Therefore, we can interpret the coefficient for `mumMyopic` as the difference in the effects of myopia in the mother and father. To test the hypothesis that the effects are equal, we can use the z-test for this coefficient.

```
mumPlusdadMyopic <- (dadMyopic == "Yes") + (mumMyopic == "Yes")
MyopiaLogReg4 <- glm(myopic ~ . -compHR - TVHR + mumPlusdadMyopic - dadMyopic,
                     data = Myopia, family = binomial)
summary(MyopiaLogReg4)
```

```
## 
## Call:
## glm(formula = myopic ~ . - compHR - TVHR + mumPlusdadMyopic -
##     dadMyopic, family = binomial, data = Myopia)
## 
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -1.0382  -0.5746  -0.4311  -0.2904   2.5691
## 
## Coefficients:
##                  Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -2.57268    0.36941  -6.964  3.3e-12 ***
## gendermale       -0.30898    0.24975  -1.237 0.216031
## sportHR          -0.04186    0.01785  -2.344 0.019066 *
## readHR            0.09620    0.03827   2.514 0.011952 *
## studyHR          -0.06065    0.06490  -0.935 0.350023
## mumMyopicYes     -0.12016    0.35822  -0.335 0.737300
## mumPlusdadMyopic  0.98788    0.26204   3.770 0.000163 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## (Dispersion parameter for binomial family taken to be 1)
## 
##     Null deviance: 480.08  on 617  degrees of freedom
## Residual deviance: 440.46  on 611  degrees of freedom
## AIC: 454.46
## 
## Number of Fisher Scoring iterations: 5
```

We cannot reject the hypothesis that the effects are equal.

Exercise 4

In order to represent this effect, we can include `mumPlusdadMyopic` in addition to an indicator for the event that both mother and father are myopic.

```
mumAnddadMyopic <- mumPlusdadMyopic==2
MyopiaLogReg5 <- glm(myopic ~ . -compHR - TVHR + mumPlusdadMyopic + mumAnddadMyopic
                     - dadMyopic - mumMyopic ,
                     data = Myopia, family = binomial)
```

Smoking exercise

```
summary(SmokingLogReg1)
```

```
## 
## Call:
## glm(formula = propDied ~ Age.group + Smoker, family = binomial,
##     weights = total)
## 
## Deviance Residuals:
##     Min      1Q   Median      3Q      Max
## -2.3947  -0.3841   0.6008   1.3799   3.4824
## 
## Coefficients:
```

```
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -7.687751   0.447646 -17.174   <2e-16 ***
## Age.group    0.124957   0.007274  17.178   <2e-16 ***
## SmokerYes    0.266053   0.168702   1.577    0.115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 641.496  on 13  degrees of freedom
## Residual deviance:  32.572  on 11  degrees of freedom
## AIC: 85.568
##
## Number of Fisher Scoring iterations: 5
```

The odds of dying (ratio of the probabilities of dying and not dying) get multiplied by $\exp(0.12497) \approx 1.13$ for every year of age (and for the rest of covariates fixed), since the age is represented in a scale of years.