STATISTICAL MODELLING Part IIC, Lent 2018
Practical 7: Binomial regression and Poisson regression By courtesy of Dr. R. Shah

Solutions

The model for football scores stored in the object `LogLinModel` is the following. If $Y_{i,j,\text{away}}$, $Y_{i,j,\text{home}}$ are the numbers of goals scored by team i against team j away and at home, respectively, then

$$Y_{i,j,\text{away}} \sim^{ind.} \text{Poisson}(\exp(\mu + \alpha_{\text{away}} + \beta_i + \gamma_j)),$$

$$Y_{i,j,\text{home}} \sim^{ind.} \text{Poisson}(\exp(\mu + \alpha_{\text{home}} + \beta_i + \gamma_j))$$

and the $Y_{i,j,\text{away}}$ and $Y_{i,j,\text{home}}$ are mutually independent, where $\alpha_{\text{away}} = \beta_{\text{Leicester}} = \gamma_{\text{Leicester}} = 0$ (corner point constraints).

```
summary(LogLinMod)
```

```
##
## Call:
## glm(formula = GoalsScored ~ HomeAway + By + Against, family = poisson)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.1477  -1.0860  -0.0816   0.5266   3.5319
##
## Coefficients:
##                       Estimate Std. Error z value Pr(>|z|)
## (Intercept)            0.03261    0.26353   0.124 0.901530
## HomeAwayHome           0.23510    0.07711   3.049 0.002298 **
## ByArsenal              0.68790    0.24865   2.767 0.005666 **
## ByAston Villa         -0.66531    0.34529  -1.927 0.054003 .
## ByBurnley             -0.02381    0.28686  -0.083 0.933853
## ByChelsea              0.78644    0.24515   3.208 0.001337 **
## ByCrystal Palace       0.17565    0.27318   0.643 0.520228
## ByEverton              0.27522    0.26955   1.021 0.307246
## ByHull                -0.03521    0.28670  -0.123 0.902265
## ByLiverpool            0.44911    0.25962   1.730 0.083654 .
## ByMan City             0.80421    0.24415   3.294 0.000988 ***
## ByMan United           0.55581    0.25348   2.193 0.028327 *
## ByNewcastle            0.24968    0.27092   0.922 0.356742
## ByQPR                  0.11468    0.28157   0.407 0.683803
## BySouthampton          0.37611    0.26177   1.437 0.150784
## ByStoke                0.18238    0.27329   0.667 0.504550
## BySunderland          -0.15786    0.29614  -0.533 0.593994
## BySwansea              0.19171    0.27316   0.702 0.482781
## ByTottenham            0.51812    0.25794   2.009 0.044566 *
## ByWest Brom           -0.02984    0.28689  -0.104 0.917171
## ByWest Ham             0.43364    0.26042   1.665 0.095883 .
## AgainstArsenal        -0.36999    0.24285  -1.524 0.127628
## AgainstAston Villa    -0.17836    0.22635  -0.788 0.430717
## AgainstBurnley         0.04692    0.21569   0.218 0.827786
## AgainstChelsea        -0.59486    0.26437  -2.250 0.024445 *
## AgainstCrystal Palace -0.10960    0.22525  -0.487 0.626576
## AgainstEverton        -0.13354    0.22366  -0.597 0.550451
## AgainstHull           -0.17248    0.22813  -0.756 0.449605
## AgainstLiverpool      -0.33343    0.24077  -1.385 0.166106
```

```
## AgainstMan City       -0.44830    0.24742  -1.812 0.070004 .
## AgainstMan United     -0.43996    0.25120  -1.751 0.079870 .
## AgainstNewcastle      -0.02623    0.21937  -0.120 0.904832
## AgainstQPR             0.10587    0.21560   0.491 0.623403
## AgainstSouthampton    -0.76085    0.27248  -2.792 0.005233 **
## AgainstStoke          -0.23629    0.23188  -1.019 0.308198
## AgainstSunderland     -0.13642    0.22523  -0.606 0.544711
## AgainstSwansea        -0.23130    0.23195  -0.997 0.318682
## AgainstTottenham      -0.13673    0.22819  -0.599 0.549032
## AgainstWest Brom      -0.20940    0.23208  -0.902 0.366918
## AgainstWest Ham       -0.24599    0.23362  -1.053 0.292365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 634.32  on 535  degrees of freedom
## Residual deviance: 521.89  on 496  degrees of freedom
## AIC: 1527
##
## Number of Fisher Scoring iterations: 5
```

The coefficient α_{home} is estimated to be 0.235105, and the effect is statistically significant at 5% significance level (if we ignore the fact that the responses are correlated, since there are pairs of responses derived from the same match). So we interpret the "home advantage" as multiplying the baseline expectation for the number of goals by

$$\exp(0.235) = 1.26,$$

that is, increasing the expected number of goals by 26%.