HW 4

Ted Henson

2/7/2020

Problem 4

a)

```
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.1548078779 -0.0005799429  0.0035405940  0.0088255719 -0.0057094223
##      Adjacent
## -0.0006630311
```

```
## [1] 716.8458
```

The values of the coefficients and the deviance is shown in the output above.

b)

The form of the adjusted dependent variable is:

```
z = eta + (y-mu)*(d.eta.d.mu)
```

The eta is 4.060443, 3.4339872, 1.0986123, 3.2188758, 0.6931472, 2.8903718, 3.1780538, 2.3025851, 2.0794415, 0.6931472, 4.574711, 4.5325995, 4.060443, 1.6094379, 3.6888795, 5.8493248, 3.9318256, 0.6931472, 4.6443909, 4.6821312, 2.4849066, 4.2484952, 5.6347896, 5.4680601, 6.0958246, 4.1271344, 5.6524892, 3.7841896, 2.7725887, 3.0445224, the d.eta.d.mu is 0.0172414, 0.0322581, 0.3333333, 0.04, 0.5, 0.0555556, 0.0416667, 0.1, 0.125, 0.5, 0.0103093, 0.0107527, 0.0172414, 0.2, 0.025, 0.0028818, 0.0196078, 0.5, 0.0096154, 0.0092593, 0.0833333, 0.0142857, 0.0035714, 0.0042194, 0.0022523, 0.016129, 0.0035088, 0.0227273, 0.0625, 0.047619, the variance of mu is 58, 31, 3, 25, 2, 18, 24, 10, 8, 2, 97, 93, 58, 5, 40, 347, 51, 2, 104, 108, 12, 70, 280, 237, 444, 62, 285, 44, 16, 21, and the weights are 58, 31, 3, 25, 2, 18, 24, 10, 8, 2, 97, 93, 58, 5, 40, 347, 51, 2, 104, 108, 12, 70, 280, 237, 444, 62, 285, 44, 16, 21.

c)

```
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.5191545412 -0.0005298484  0.0031643557  0.0025188990 -0.0037899780
##      Adjacent
## -0.0006623523
```

```
## [1] 576.8075
```

The intercept is fairly close to the values computed by the glm function. The Area and Adjacent variables are very close to the final values. The other values appear to be on the right track, but are not close to their final values yet.

d)

```
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.2102594447 -0.0005651969  0.0034606226  0.0077171134 -0.0052400871
##      Adjacent
## -0.0006604828
```

```
## [1] 570.9648
```

This deviance is fairly close to the glm deviance, but still can probably converge closer.

e)

```
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.1562582546 -0.0005793855  0.0035379237  0.0087861184 -0.0056868875
##      Adjacent
## -0.0006630167
## [1] "Iteration 1 Deviance: 725.51031962005"
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.1548090631 -0.0005799422  0.0035405910  0.0088255087 -0.0057093801
##      Adjacent
## -0.0006630313
## [1] "Iteration 2 Deviance: 760.699799799015"
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.1548078779 -0.0005799429  0.0035405940  0.0088255719 -0.0057094223
##      Adjacent
## -0.0006630311
## [1] "Iteration 3 Deviance: 761.977437469427"
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.1548078779 -0.0005799429  0.0035405940  0.0088255719 -0.0057094223
##      Adjacent
## -0.0006630311
## [1] "Iteration 4 Deviance: 761.979247760923"
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.1548078779 -0.0005799429  0.0035405940  0.0088255719 -0.0057094223
##      Adjacent
## -0.0006630311
## [1] "Iteration 5 Deviance: 761.979247765408"
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.1548078779 -0.0005799429  0.0035405940  0.0088255719 -0.0057094223
##      Adjacent
## -0.0006630311
## [1] "Iteration 6 Deviance: 761.979247765408"
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.1548078779 -0.0005799429  0.0035405940  0.0088255719 -0.0057094223
##      Adjacent
## -0.0006630311
## [1] "Iteration 7 Deviance: 761.979247765408"
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.1548078779 -0.0005799429  0.0035405940  0.0088255719 -0.0057094223
##      Adjacent
## -0.0006630311
## [1] "Iteration 8 Deviance: 761.979247765408"
##   (Intercept)          Area     Elevation       Nearest         Scruz
```

```
##  3.1548078779 -0.0005799429  0.0035405940  0.0088255719 -0.0057094223
##      Adjacent
## -0.0006630311
## [1] "Iteration 9 Deviance: 761.979247765408"
##   (Intercept)          Area     Elevation       Nearest         Scruz
##  3.1548078779 -0.0005799429  0.0035405940  0.0088255719 -0.0057094223
##      Adjacent
## -0.0006630311
## [1] "Iteration 10 Deviance: 761.979247765408"
```

The deviance is very close to the target after 10 iterations and appears to have converged. All of the coefficients are very close to the glm's as well.

Problem 6

a)

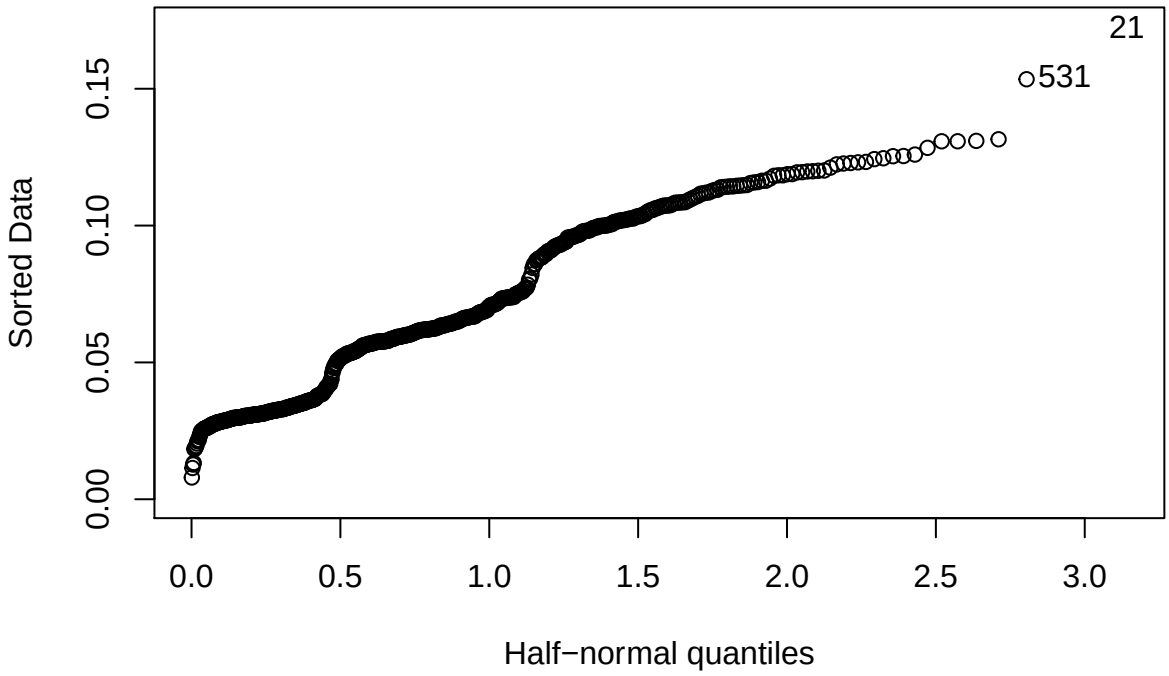
```
##
## Call:
## glm(formula = Shots ~ Team + Position + Tackles.per.90.min +
##     Passes.per.90.min, family = "poisson", data = worldcup)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.9451  -1.4487  -0.3227   0.5258   5.4445
##
## Coefficients:
##                     Estimate Std. Error z value Pr(>|z|)
## (Intercept)         0.164837   0.200122   0.824 0.410119
## TeamArgentina       0.778391   0.221968   3.507 0.000454 ***
## TeamAustralia      -0.157955   0.258458  -0.611 0.541105
## TeamBrazil          0.832805   0.216579   3.845 0.000120 ***
## TeamCameroon        0.069575   0.247975   0.281 0.779038
## TeamChile           0.282551   0.235277   1.201 0.229780
## TeamDenmark        -0.094604   0.252068  -0.375 0.707430
## TeamEngland         0.460226   0.229875   2.002 0.045277 *
## TeamFrance         -0.167531   0.259100  -0.647 0.517899
## TeamGermany         0.688142   0.219869   3.130 0.001749 **
## TeamGhana           0.840148   0.214732   3.913 9.13e-05 ***
## TeamGreece         -0.034924   0.253378  -0.138 0.890372
## TeamHonduras       -1.035233   0.330693  -3.130 0.001745 **
## TeamItaly           0.096256   0.247014   0.390 0.696776
## TeamIvory Coast     0.293131   0.241622   1.213 0.225061
## TeamJapan           0.142848   0.242102   0.590 0.555169
## TeamMexico          0.364535   0.237609   1.534 0.124986
## TeamNetherlands     0.788152   0.219050   3.598 0.000321 ***
## TeamNew Zealand    -0.633606   0.314859  -2.012 0.044183 *
## TeamNigeria        -0.169823   0.259024  -0.656 0.512065
## TeamNorth Korea     0.174491   0.252334   0.692 0.489247
## TeamParaguay        0.348329   0.230826   1.509 0.131285
## TeamPortugal        0.457938   0.230425   1.987 0.046883 *
## TeamSerbia          0.165891   0.242190   0.685 0.493369
## TeamSlovakia       -0.018986   0.247267  -0.077 0.938796
## TeamSlovenia       -0.517359   0.302901  -1.708 0.087634 .
```

```
## TeamSouth Africa    0.198270   0.247507   0.801 0.423093
## TeamSouth Korea     0.418820   0.234687   1.785 0.074327 .
## TeamSpain           1.008485   0.214527   4.701 2.59e-06 ***
## TeamSwitzerland    -0.441761   0.282767  -1.562 0.118223
## TeamUSA             0.425214   0.224804   1.891 0.058560 .
## TeamUruguay         0.606591   0.216224   2.805 0.005026 **
## PositionForward     1.215280   0.082578  14.717  < 2e-16 ***
## PositionGoalkeeper -3.922002   1.003696  -3.908 9.32e-05 ***
## PositionMidfielder  0.807259   0.082679   9.764  < 2e-16 ***
## Tackles.per.90.min -0.031253   0.018116  -1.725 0.084490 .
## Passes.per.90.min  -0.006863   0.001797  -3.819 0.000134 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 2182.6  on 594  degrees of freedom
## Residual deviance: 1487.0  on 558  degrees of freedom
## AIC: 2661.3
##
## Number of Fisher Scoring iterations: 6
```

Tackles and passes per 90 minutes have a negative relationship with shots per 90 minutes. Passes per 90 minutes is significant by the .0005 threshold and Tackles per 90 minutes may or may not be significant.

b)

```
##     Amoah
## 0.1728024
```



Amoah has high leverage since he is a forward, plays for Gahana, and has a lot of tackles per 90 minutes. These three combinations are unusual predictor values and have the potential to affect the fit of the model.

c)

```
##      Amoah
## 0.02235704
```

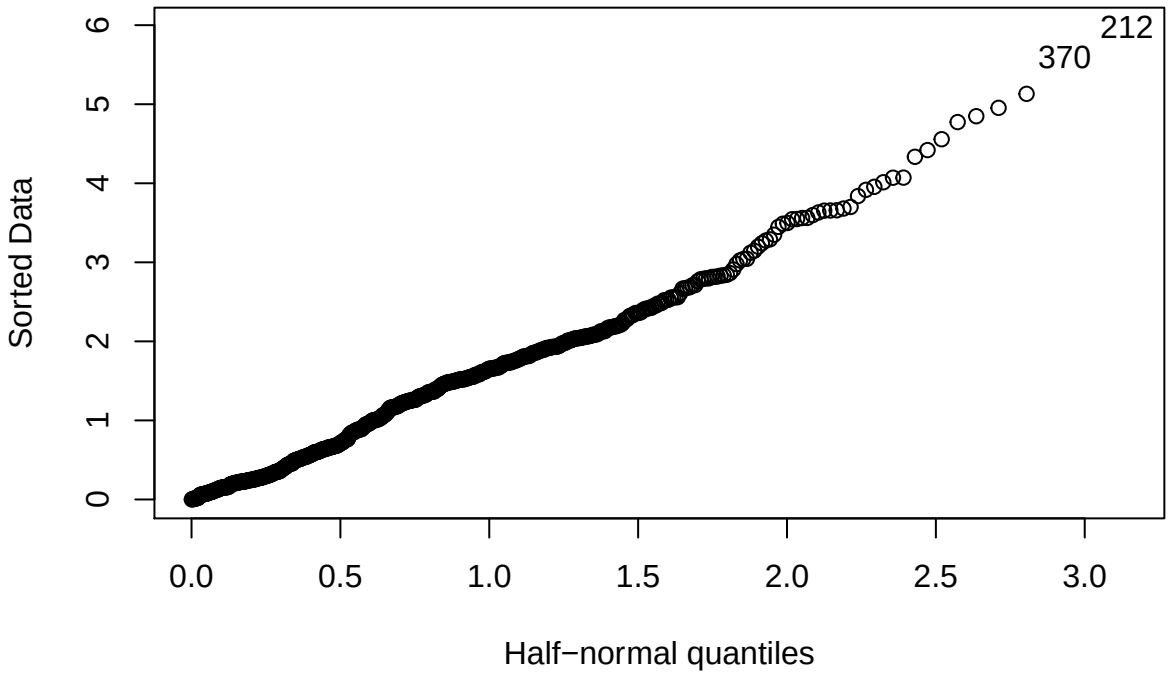
Amoah has high influence since he has the unusual predictor values mentioned above, but he also has zero shots, influencing the model when the weights would predict he would have a lot of shots.

d)

```
##      Gyan
## 0.1867534
```

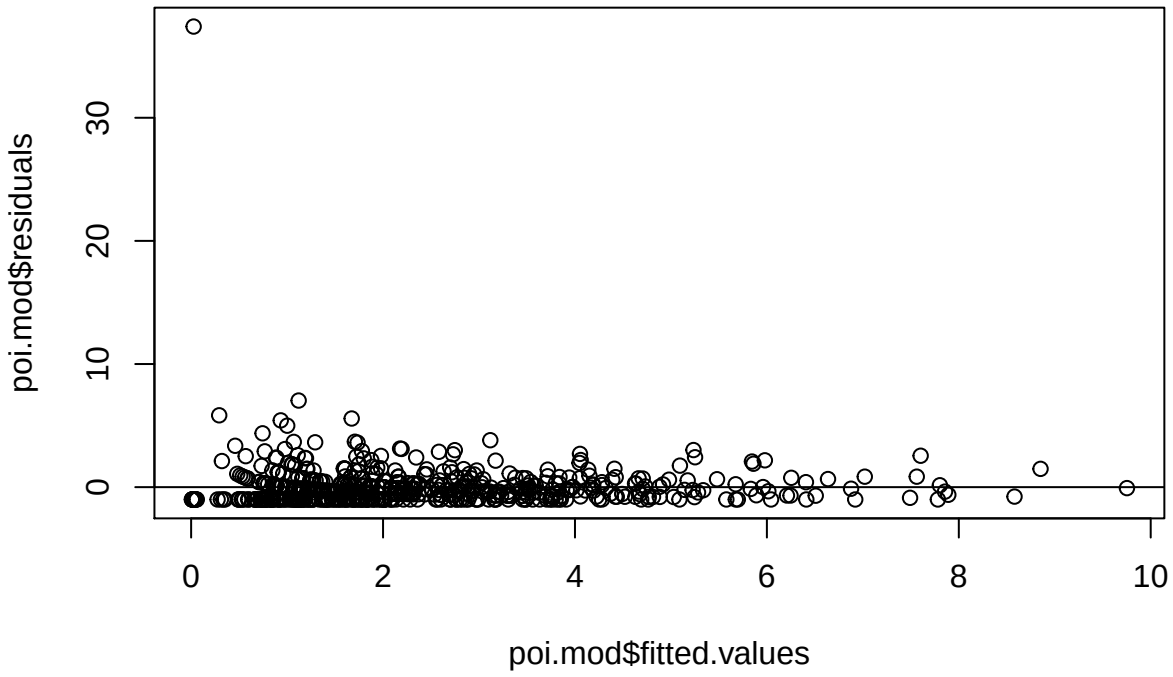
Gyan has the highest number of shots cross the entire dataset and a low number of passes per 90 minutes compared to his team.

e)



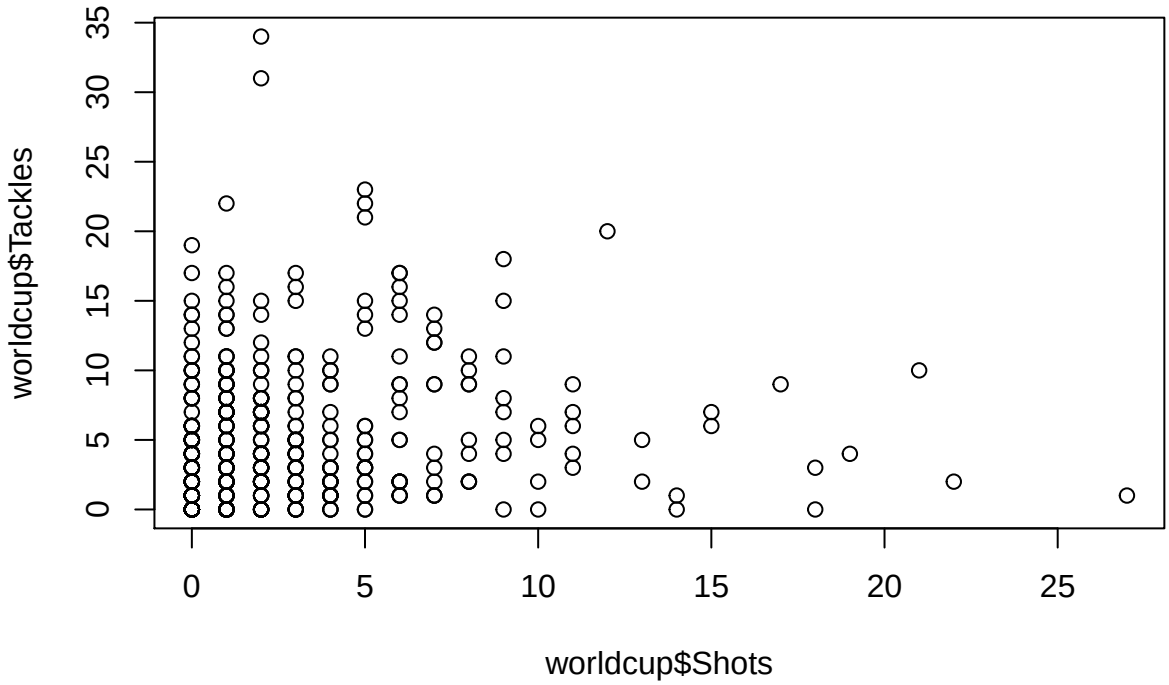
The player with the largest jacknife residual is 11, 2, 501, 27, 151, 1, 0, 0.179640718562874, 27.125748502994. As with the Cook's distance, Gyan has the highest number of shots cross the entire dataset and a low number of passes per 90 minutes compared to his team.

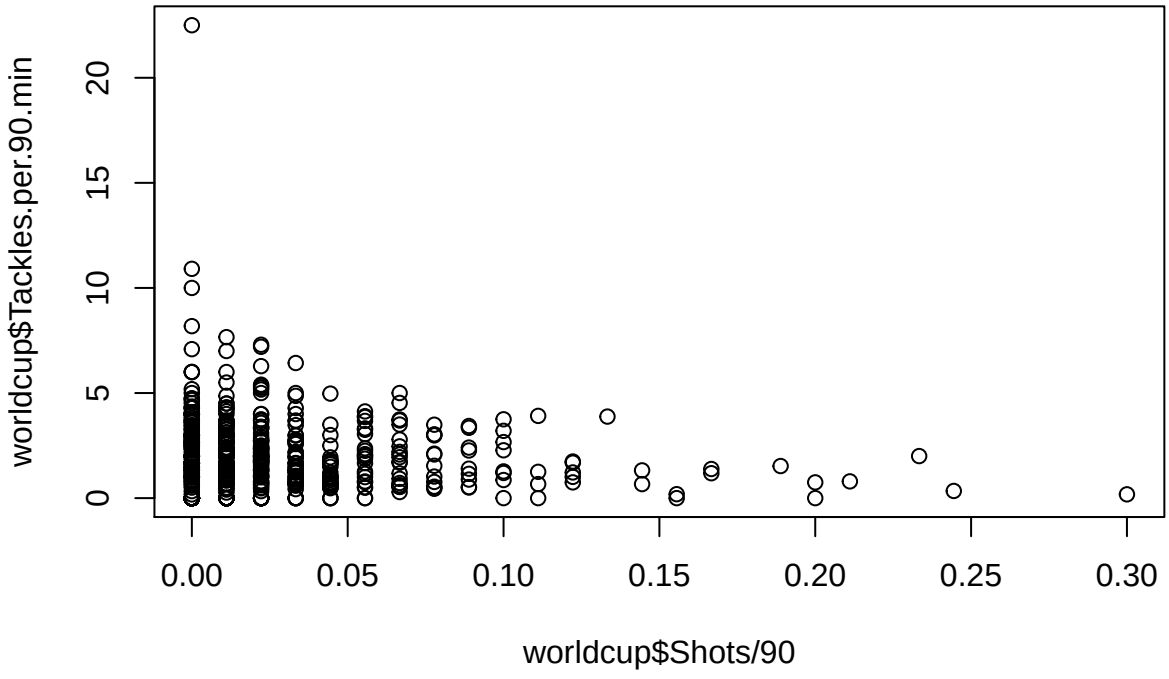
f)

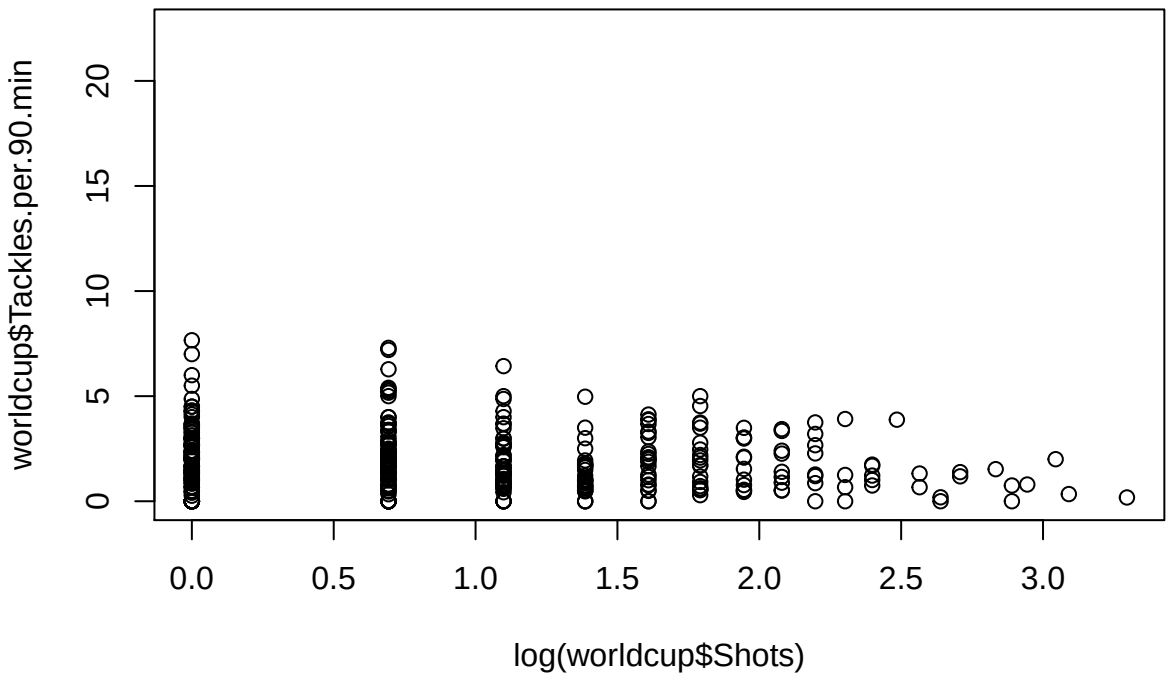


The horizontal line is drawn at residuals of 0. Ideally the points should be normally distributed throughout the line. This plot indicates the residuals are larger on the lower quantiles, and there is an especially large residual for one point fitted at almost 0 shots.

g)

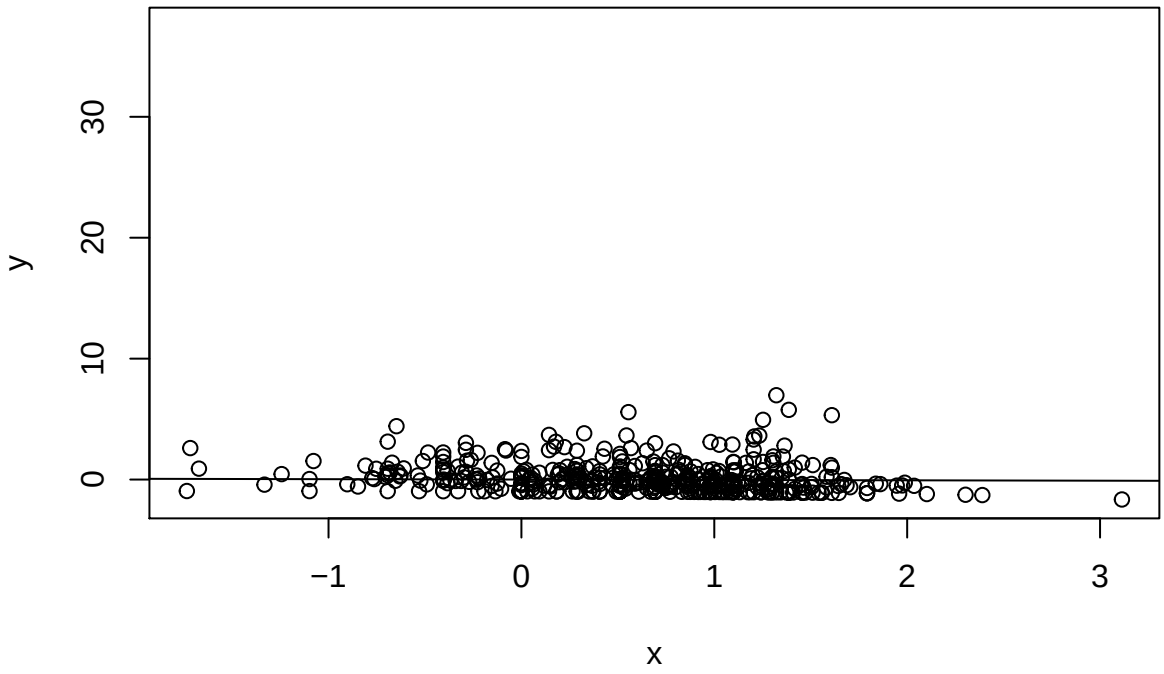

worldcup$Tackles
0
5
10
15
20
25
30
35
0
5
10
15
20
25
worldcup$Shots


worldcup$Tackles.per.90.min
0
5
10
15
20
0.00
0.05
0.10
0.15
0.20
0.25
0.30
worldcup$Shots/90

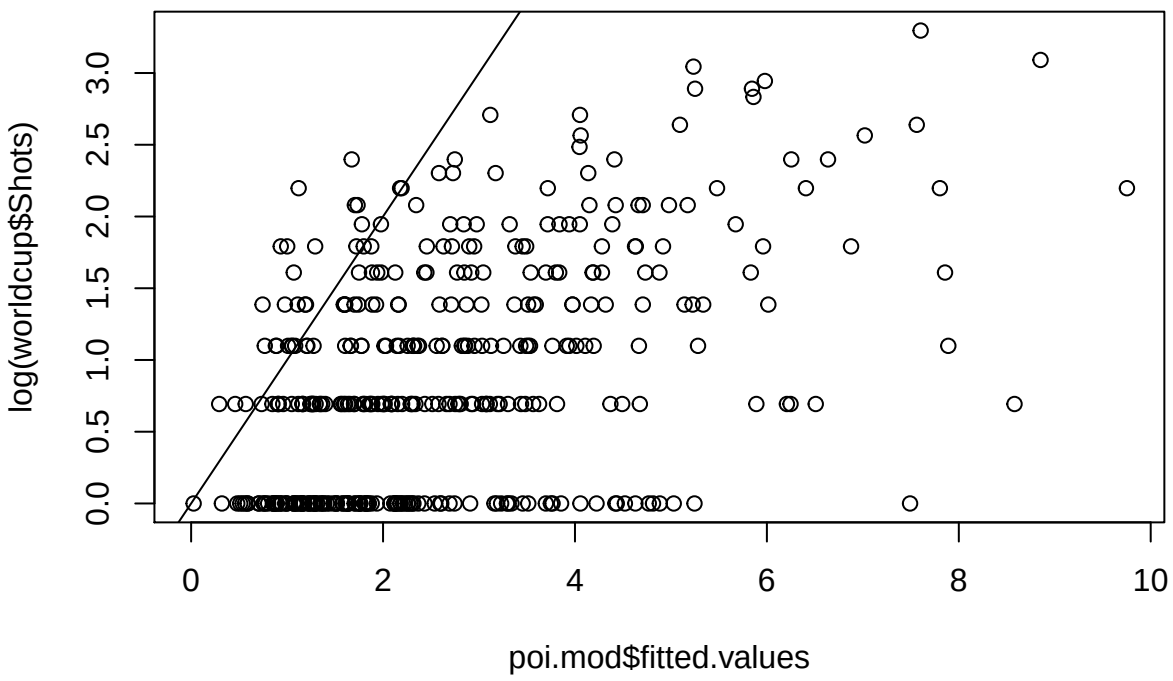


The plots appear to show that there is a slightly negative relationship between the linearized shots per game and the tackles per 90 minutes.

h)



f)



Our link function may not be the best for our data as the plot of fitted values versus the linearized response shows that the fitted values are much higher than the linearized response.