

CATEGORICAL DATA ANALYSIS

FINAL PROJECT

Categorical Data Analysis in NBA

Author:
Minzhao LIU

April 15, 2010

Contents

1	Introduction	2
1.1	Teams	2
1.2	Players	2
1.3	Data	3
2	Model Selection and Fitting	3
2.1	Selecting Covariates	3
2.1.1	Responses v.s. Each Covariates	3
2.1.2	Multi-collinearity	4
2.2	Model Fitting and Diagnostics	4
2.2.1	Forward Selection	4
2.2.2	Diagnostics	6
3	Some Inference from Model	6
3.1	Conditional Association	6
3.1.1	Using Logit Models to Test Conditional Independence	6
3.1.2	Cochran-Mantel-Haenszel Test of Conditional Independence	7
3.2	Parameters Interpretation and Conclusion	7
4	Discussion	8
4.1	Longitudinal Data Perspective	8
4.2	Link Function Test	8
4.3	Overdispersion	9
4.4	Other Predictors	9

Abstract

This article mainly uses binary generalized linear models to analyze the data from NBA players. By logistic model, the data from Yao Ming are used to predict odds to win. Results show certain covariates need to be selected and they are quite significantly related to team victory. This model also presents how personal behaviors change game result and how odds ratio varies between various divisions. Those conclusions may help coach to design strategy for some particular player and particular opponents. Longitudinal perspective, link function test and overdispersion situation are also discussed in this article.

Keywords: binary glm, model selection, diagnostic, multicollinearity, conditional association, longitudinal, link function test.

1 Introduction

1.1 Teams

NBA is short for national basketball association, and there are 30 teams in the league, consisting 2 conferences, east and west, as well as 6 divisions. Each division contains 5 teams. In regular season, each team has to play 82 games, in which, there are 4 games (2 home, 2 away) for each opponent in the same conference, and 2 or 3 games for every opponent in the other conference. This kind of schedule suffices us a sufficient large dataset, as long as the player is healthy enough to play.

1.2 Players

Basically, there are 5 starters and several bench players in one team, and for each game, about 10 players will show up on the court for various minutes. Although basketball is said to be a team sports, only 5 players could be on the court at the same time, thus, individual contributes to the game results much more than one in football or soccer games.

Although each game has only two results, win or lose, which we will take as binary categorical responses, every player has lots of statistics during the game. Our goal is to find out how those individual figures change the game result.

In addition, there is a statistics evaluating the whole performance of one player in one game, called *Efficiency*, which could be calculated as

$$Efficiency = PTS + TOT + AST + STL + BLK - (FGA - FGM) - (FGA - FTM) - TO \quad (1)$$

1.3 Data

This data is from <http://www.databasebasketball.com/>, and Yao Ming is picked up as our objective, who is a center serving *Houston Rockets*. He played 481 games during 2002 Oct-2009 May, 6 seasons, of which 283 games are won and 198 are lose. For details, please see appendix.

2 Model Selection and Fitting

Model selection [Alan \[2002\]](#) for logistic regression faces several issues. Here two of them will be mainly discussed, which exploratory covariates should be added into the model, and whether to include interaction or which are not significant such that they could be removed. Goal is that using simplest model to smoothly fit the complicated data well. Other issues like link function selection, longitudinal perspective and overdispersion will be discussed in section 4

2.1 Selecting Covariates

Before choosing various kinds of models, some crucial exploratory terms need to be selected instead of those useless ones. Moreover, those covariates with strong collinearity should be eliminated.

2.1.1 Responses v.s. Each Covariates

In order to find out which predictors have effects on the response Y , it is helpful and useful by inspecting their marginal relations from pictures. Suppose taking win or lose as response variable Y , other personal statistics as covariates. Graphs are drawn for continuous predictors and contingency tables are made for categorical ones.

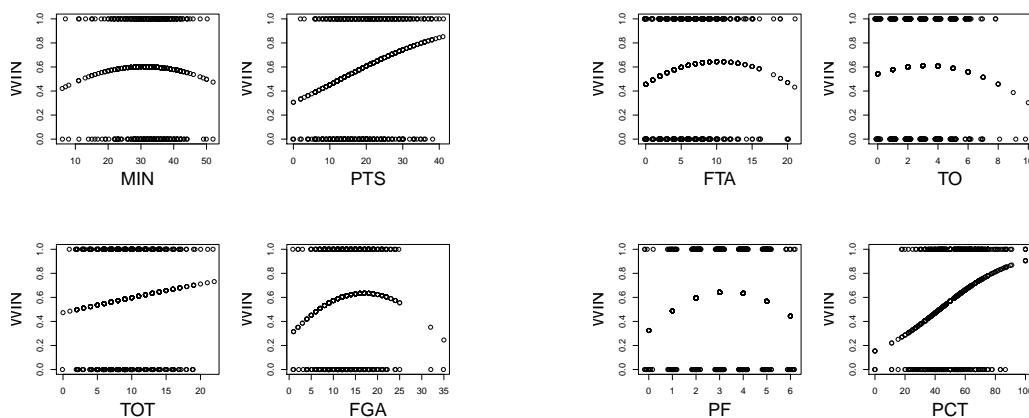


Figure 1: Responses v.s Each Covariates

DIV	Lose	Win
1	20	41
2	24	37
3	15	38
4	49	38
5	38	67
6	52	62

Table 1: Division v.s Win

Home	Lose	Win
0	123	120
1	75	163

Table 2: Home v.s. Win

From the graphics and tables above, intuitively response Y has association to some extent with PTS , PCT , and $Home$. As model shows later, those predictors fit the model well.

2.1.2 Multi-collinearity

As multicollinearity draws the caution in ordinary regression model, it also needs to be taken into consideration for generalized linear models. If model is fitted with multi-collinearity, one variable would seem to be non-significant because of its overlapping with other predictors and it could be predicted by other predictors. One good thing about eliminating redundant covariates is that it reduces standard errors of other estimated effects.

As to detect predictors multi-collinearity, one way is to make scatter plot. Here is the plot for main covariates.

Seen from the graph 2, EFF , MIN , PTS , TOT , FGA have strong correlation and could be regarded as one group. Only one of the them would be selected into the model. Meanwhile, this group, as well as FTA , TO , PF , PCT are mutually independent. The same result could be drew from their correlation matrix, which is not shown here.

2.2 Model Fitting and Diagnostics

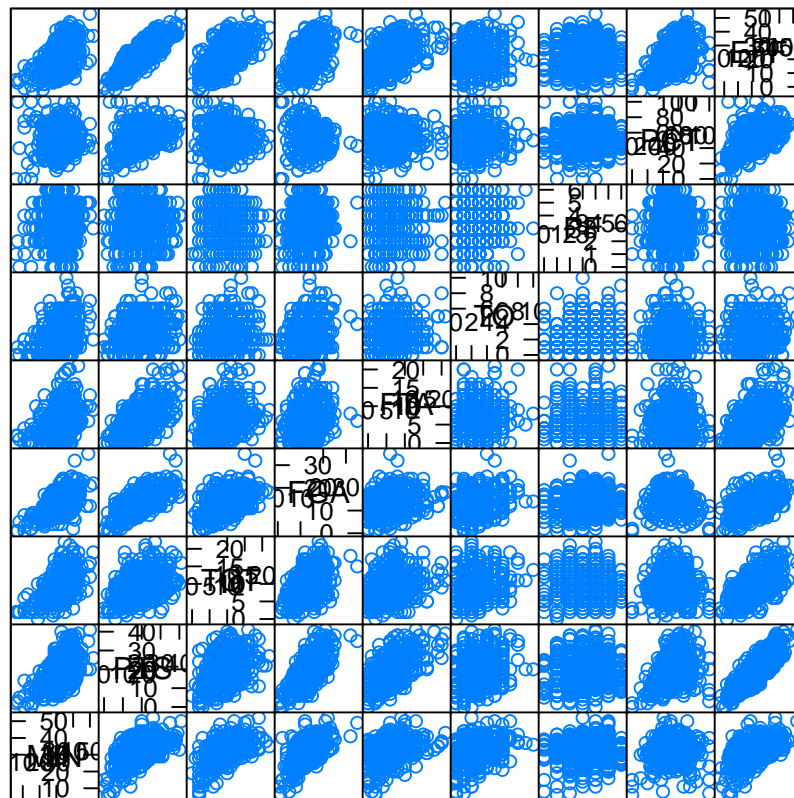
2.2.1 Forward Selection

Goodman (1971) Goodman [1971] proposed methods analogous to forward selection and backward elimination in ordinary regression. Here forward procedure is also used in generalized linear model selection.

First fit the null model, and then add each main effect into the model to see if there is significant reduction in the deviance.

	HOME	DIV	MIN	PTS	TOT	FGA	FTA	TO	PF	PCT
p-value	0.00	0.80	0.89	0.00	0.08	0.16	0.18	0.92	0.55	0.00

Table 3: P-value for main terms



Scatter Plot Matrix

Figure 2: Predictors Correlation

By output in appendix and table 3, fortunately we found p-value for *home*, *points*, *percentage* are highly significant less than 0.001, and *total rebounds* is barely significant ($P = 0.08$). Since the critical value 0.05 is arbitrary, one could also take *FGA*, *FTA* as significant if he needs ($P \simeq 0.17$).

Next step, consider those interaction terms. Given the model includes *home*, *points*, *percentage*, *total rebounds* main term, and group them into factor.

	HOME:PTS	HOME:TOT	HOME:PCT	PTS:TOT	PTS:PCT	TOT:PCT
p-value	1.00	0.99	0.70	1.00	1.00	1.00

Table 4: P-value for interactions

Voila, from the output in appendix and table 4, the procedure could end with no interaction needed. The 'final' model would be

$$\text{WIN} = \text{HOME} + \text{PTS.f} + \text{TOT.f} + \text{PCT.f} \quad (2)$$

2.2.2 Diagnostics

Don't be scared about the huge deviance residual in output for 'final' model above, as they are from the ungrouped data. The following diagnostics from grouped data would show how goodness-of-fit the model is.

Since a single residual is usually uninformative, lack of fit is not fully supported by detecting large deviance. Plots of residuals against linear predictors also have limited use, consisting simply of two parallel lines of dots.

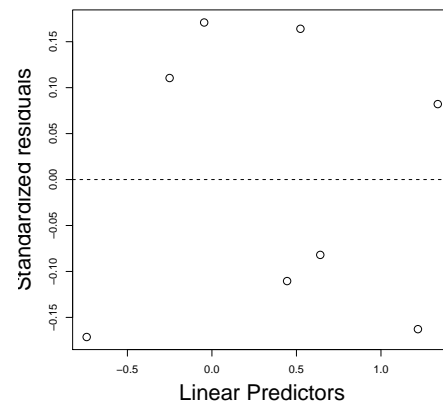
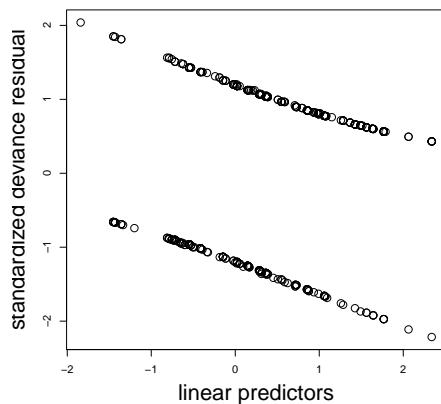


Figure 3: Individual standardized residuals v.s linear predictors Figure 4: Grouped residuals vs linear predictors

For convenience, just group *points* predictor into $[0, 10]$, $(10, 20]$, $(20, 30]$, $(30, 50]$.

This time, the plot 4 shows pretty good. There's no lack of fit for this model with $G^2 = 0.06$ with $df = 3$. The fit is satisfactory. See appendix for detail.

3 Some Inference from Model

3.1 Conditional Association

One interesting issue is to test conditional independence as a logit model for $HOME \times WIN \times DIV$ contingency table. Here are two methods.

3.1.1 Using Logit Models to Test Conditional Independence

Suppose Y stand for win or lose, X is home or away, Z represents divisions. The logit model is

$$\text{logit}(P(Y = 1)) = \alpha + \beta X_i + \beta_k^Z, \quad i = 1, 2 \quad k = 1, 2, \dots, 6$$

where $X_1 = 1, X_2 = 0$.

From the output in appendix, the null hypothesis is rejected and estimated common odds ratio is $\exp(\beta) = 2.3$, which means given division, the odds for win at home is 2.3 times the odds away.

3.1.2 Cochran-Mantel-Haenszel Test of Conditional Independence

Mantel and Haenszel [Mantel and Haenszel \[1959\]](#) proposed a non-model based test for conditional independence.

```
> dat <- table(HOME, WIN, DIV)
> dat
, , DIV = 3
, , DIV = 1
      WIN
HOME 0 1
0 10 22
1 10 19
, , DIV = 2
      WIN
HOME 0 1
0 16 12
1 8 25
, , DIV = 4
      WIN
HOME 0 1
0 32 13
1 17 25
, , DIV = 5
      WIN
HOME 0 1
0 26 29
1 12 38
, , DIV = 6
      WIN
HOME 0 1
0 29 27
1 23 35
```

Note although the order of *HOME* and *WIN* is reversed from usual, the odds ratio stay the same by its invariance after exchanging rows and columns, thus it's legitimate to deal with this contingency table.

```
> mantelhaen.test(dat)
```

Mantel-Haenszel chi-squared test with continuity correction

```
data: dat
Mantel-Haenszel X-squared = 17.8742, df = 1, p-value = 2.36e-05
alternative hypothesis: true common odds ratio is not equal to 1
95 percent confidence interval:
 1.56085 3.31546
sample estimates:
common odds ratio
      2.27485
```

where the result coincides the one from logit model.

3.2 Parameters Interpretation and Conclusion

Referring to the final model equation 2, the odds at home is $\exp(0.7) \simeq 2$ times the odds at away, which is pretty high. When other predictors fixed, the odds is

highest when Yao gets points between 20 and 30. It is very weird that the odds for win is not higher when Yao gets more rebound than 15. Field throw percentage is also a crucial factor. When field throw percentage arises to 60% up, the odds is at least $\exp(1.72 - 0.94) = 2.18$ times the one at usual.

Some suggestion for coach on Yao Ming may be 1. to put emphasis on offense rather than defense 2. to grab 20+, 10+ 3. to control the time on court in order to have stable field throw percentage, which may depend on strength.

4 Discussion

4.1 Longitudinal Data Perspective

This study can also be regarded as repeated measurement, since Yao met those opponents almost every season. Those game results are supposed to be highly correlated. In addition, because of health problem, Yao sometimes would have a rest and miss some games with certain opponents, those missing data would cause bias in estimates by usual maximum likelihood method. One way to deal with this longitudinal data and missing values is *generalized estimation equations* (GEE) [Liang and Zeger \[1986\]](#). An advantage of the GEE method is that different objects could have different numbers of observations.

From the output in appendix, those estimates are similar to the ML estimates, and the corresponding conclusions are also analogue.

4.2 Link Function Test

By substituting logit link with probit, complementary log-log, cauchit link, those estimates do not change dramatically, as well as the goodness of fit for those model.

Another way to testing the link is by Stukel [Stukel \[1988\]](#), who proposes an extended link family with two additional parameters, α_1 and α_2 . If $\alpha_1 = \alpha_2 = 0$, then it is the usual logit link. While $\alpha_1 = \alpha_2 > 0$ ($\alpha_1 = \alpha_2 < 0$), it goes to a link that rises more quickly (slowly) than the logit. If $\alpha_1 \neq \alpha_2$, the extended link is asymmetric.

Stukel suggests score tests for several possible alternative hypothesis to the logit link. (a) $H_a : \alpha_1 \neq 0, \alpha_2 = 0$, (b) $H_a : \alpha_1 = 0, \alpha_2 \neq 0$, (c) $H_a : \alpha_i \neq 0$. Here's the illustration on our logistic model.

	$\alpha_i \neq 0$	$\alpha_1 \neq 0$	$\alpha_2 \neq 0$
p-value	0.6128	0.437	0.3242

Table 5: Stukel Score Link Test for Three Alternative Hypothesis

From the output in appendix and table 5, either of them shows significance, such that we could trust and use logit link safely.

4.3 Overdispersion

If the model does not fit well, overdispersion might be the reason, though it does not happen here. Several methods could be adopted. 1) quasi binomial 2) betabinomial 3) generalized linear mixed models to handle overdispersion problem.

There are corresponding families in `glm` or `vglm` function for quasi binomial and betabinomial when using R. For SAS, *PROC GLIMMIX*, *PROC NLMIXED* could deal with this situation. *glmer*, *glmmPQL*, *glmmADMB*, *MCMCglmm*, *glmm-BUGS* in R are also used and they have separate advantages and limitations.

4.4 Other Predictors

As mentioned earlier, efficiency could be a score to evaluate how one player performs during the game. Since some positions may have less balls to handle, or some positions have to do dirty work, whose contribution could not shown up by those data. Recently another score \pm stat has been put up to evaluate players. \pm measures how many points the team win or lose when that player is on the court. That data could also be negative obviously. \pm stat is an efficient way to find out which combination has the magic chemistry in the team. Further study might take those combination including two, three or four players as predictors.

References

- A. Alan. Categorical data analysis. *Editorial John Willey & Sons*, 2002.
- L.A. Goodman. The analysis of multidimensional contingency tables: Stepwise procedures and direct estimation methods for building models for multiple classifications. *Technometrics*, 13(1):33–61, 1971.
- K.Y. Liang and S.L. Zeger. Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13, 1986.
- N. Mantel and W. Haenszel. Statistical aspects of the analysis of data from retrospective studies. *J Natl Cancer Inst*, 22(4):719–48, 1959.
- T.A. Stukel. Generalized logistic models. *Journal of the American Statistical Association*, 83(402):426–431, 1988.

```

# data
YEAR DATE DAY OPP MIN FGM FGA FTM FTA HOME WIN OREB DREB TOT AST STL
1 2002 Oct. 30 IND 11 0 1 0 0 0 0 0 0 2 2 0 0
2 2002 Nov. 1 DEN 13 1 5 0 0 0 0 0 1 2 5 7 1 0
3 2002 Nov. 2 TOR 24 4 7 0 2 1 1 1 1 1 3 4 0 0
4 2002 Nov. 5 SEA 15 0 2 0 0 1 0 1 1 1 2 3 1 0
5 2002 Nov. 9 GSW 7 1 1 1 2 1 1 1 1 1 1 2 0 0
6 2002 Nov. 12 POR 15 2 2 3 3 1 1 1 1 3 4 2 0
TO BLK PF PTS PCT EFF DIV CON
1 2 0 3 0 0.0000 -1 2 EAST
2 0 0 2 2 20.0000 6 5 WEST
3 2 1 3 8 57.1429 6 1 EAST
4 1 0 4 0 0.0000 1 5 WEST
5 1 0 2 3 100.0000 3 6 WEST
6 2 1 0 7 100.0000 12 5 WEST

```

where

$$\begin{cases} Home = 1 & Home \\ Home = 0 & Away \end{cases}, \begin{cases} Win = 1 & Win \\ Win = 0 & Lose \end{cases}$$

```

# first forward procedure
> yao.logit1 <- glm(WIN~1,binomial)
> add1(yao.logit1, ~HOME+factor(DIV)+MIN+PTS+TOT+FGA+FTA+TO+PF+PCT,
test='F')
Single term additions

```

Model:

WIN ~ 1

	Df	Deviance	AIC	F value	Pr(F)	
<none>		651.7	653.7			
HOME	1	633.4	637.4	13.808	0.000226	***
factor(DIV)	5	635.9	647.9	0.471	0.797584	
MIN	1	651.7	655.7	0.018	0.893000	
PTS	1	626.8	630.8	19.033	1.57e-05	***
TOT	1	647.5	651.5	3.086	0.079596	.
FGA	1	649.0	653.0	1.973	0.160758	
FTA	1	649.3	653.3	1.767	0.184341	
TO	1	651.7	655.7	0.010	0.920758	
PF	1	651.2	655.2	0.358	0.550109	
PCT	1	611.5	615.5	31.528	3.34e-08	***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Warning message:

```

In add1.glm(yao.logit1, ~HOME + factor(DIV) + MIN + PTS + TOT + :
F test assumes quasibinomial family

```

```
# second forward procedure
> yao.logit2 <- update(yao.logit1, .~.+HOME+PTS.f+TOT.f+PCT.f)
> add1(yao.logit2, ~HOME+PTS.f+TOT.f+PCT.f+HOME:PTS.f+HOME:TOT.f
+HOME:PCT.f+PTS.f:TOT.f+PTS.f:PCT.f+ TOT.f:PCT.f, test='F')
Single term additions

Model:
WIN ~ HOME + PTS.f + TOT.f + PCT.f
              Df Deviance   AIC F value Pr(F)
<none>                578.4 602.4
HOME:PTS.f    3      578.3 608.3    0.004 1.000
HOME:TOT.f    3      578.1 608.1    0.029 0.993
HOME:PCT.f    4      567.6 599.6    0.549 0.700
PTS.f:TOT.f   8      571.7 611.7    0.084 1.000
PTS.f:PCT.f   9      570.9 612.9    0.075 1.000
TOT.f:PCT.f  12      566.2 614.2    0.068 1.000
Warning message:
In add1.glm(yao.logit2, ~HOME + PTS.f + TOT.f + PCT.f + HOME:PTS.f + :
  F test assumes quasibinomial family

# final model
> summary(yao.logit2)

Call:
glm(formula = WIN ~ HOME + PTS.f + TOT.f + PCT.f, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.205  -1.008   0.598   0.955   1.994

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.4297    0.4612   -3.10  0.00194 **
HOME            0.6959    0.2052    3.39  0.00069 ***
PTS.f(10,20]    0.0940    0.3517    0.27  0.78917
PTS.f(20,30]    0.8017    0.4086    1.96  0.04975 *
PTS.f(30,50]    0.5257    0.5541    0.95  0.34275
TOT.f(5,10]    -0.0216    0.3129   -0.07  0.94486
TOT.f(10,15]    0.5493    0.3589    1.53  0.12581
TOT.f(15,23]   -0.5052    0.5063   -1.00  0.31838
PCT.f(30,40]    0.6402    0.4889    1.31  0.19037
PCT.f(40,50]    0.8081    0.4729    1.71  0.08750 .
PCT.f(50,60]    0.9446    0.4974    1.90  0.05755 .
PCT.f(60,100]   1.7216    0.4849    3.55  0.00038 ***
---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 651.71 on 480 degrees of freedom
Residual deviance: 578.37 on 469 degrees of freedom
AIC: 602.4

Number of Fisher Scoring iterations: 4

```
# group data
> home.win <- as.data.frame(table(HOME,PTS.f,WIN))
> home.win <- reshape(home.win, timevar='WIN',idvar=c('HOME','PTS.f'),
direction='wide')
> colnames(home.win) <- c('home','pts.f','lose','win')
> home.fit
```

```
Call: glm(formula = cbind(win, lose) ~ home + pts.f,
family = binomial(), data = home.win)
```

```
Coefficients:
(Intercept)      home1 pts.f(10,20] pts.f(20,30] pts.f(30,50]
      -0.741       0.695       0.491       1.382       1.264
```

Degrees of Freedom: 7 Total (i.e. Null); 3 Residual

Null Deviance: 45.4

Residual Deviance: 0.0611 AIC: 43.4

```
> home.fit <- glm(cbind(win,lose)~ home+pts.f, binomial(), home.win)
> summary(home.fit)
```

Call:

```
glm(formula = cbind(win, lose) ~ home + pts.f, family = binomial(),
data = home.win)
```

Deviance Residuals:

1	2	3	4	5	6	7	8
-0.1080	0.1169	0.0509	-0.0596	-0.0487	0.0482	0.1155	-0.1078

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	-0.741	0.281	-2.64	0.00829	**
home1	0.695	0.196	3.55	0.00038	***
pts.f(10,20]	0.491	0.299	1.64	0.10114	
pts.f(20,30]	1.382	0.320	4.32	1.6e-05	***
pts.f(30,50]	1.264	0.463	2.73	0.00631	**

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 45.402787 on 7 degrees of freedom
Residual deviance: 0.061126 on 3 degrees of freedom
AIC: 43.35

Number of Fisher Scoring iterations: 3

```
# logit link for conditional independence
> yao.logit2 <- glm(WIN~HOME+factor(DIV),binomial)
> summary(yao.logit2)
```

Call:

```
glm(formula = WIN ~ HOME + factor(DIV), family = binomial)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.797	-1.249	0.782	1.032	1.471

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.351	0.288	1.22	0.2231
HOME	0.835	0.194	4.31	1.6e-05 ***
factor(DIV)2	-0.352	0.386	-0.91	0.3614
factor(DIV)3	0.207	0.416	0.50	0.6198
factor(DIV)4	-1.020	0.356	-2.87	0.0041 **
factor(DIV)5	-0.157	0.347	-0.45	0.6500
factor(DIV)6	-0.593	0.338	-1.75	0.0797 .

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 651.71 on 480 degrees of freedom
Residual deviance: 616.85 on 474 degrees of freedom
AIC: 630.8

Number of Fisher Scoring iterations: 4

```
> exp(coef(yao.logit2)['HOME'])
HOME
2.30464
```

```
# GEE
> library(gee)
> yao.gee <- gee(WIN~HOME+factor(DIV)+PTS.f, id=YEAR,
  family=binomial, corstr='exchangeable')
summary(yao.gee)
summary(yao.gee)
>
GEE:  GENERALIZED LINEAR MODELS FOR DEPENDENT DATA
gee S-function, version 4.13 modified 98/01/27 (1998)

Model:
  Link:                               Logit
  Variance to Mean Relation: Binomial
  Correlation Structure:      Exchangeable

Call:
gee(formula = WIN ~ HOME + factor(DIV) + PTS.f, id = YEAR,
    family = binomial, corstr = "exchangeable")

Summary of Residuals:
      Min       1Q   Median       3Q      Max
-0.852127 -0.460014  0.190669  0.363263  0.815421

Coefficients:
              Estimate Naive S.E.   Naive z Robust S.E.  Robust z
(Intercept)  -0.4208126   0.387513 -1.085932   0.355066 -1.185166
HOME          0.7215157   0.201642  3.578204   0.255438  2.824617
factor(DIV)2 -0.3057166   0.401892 -0.760694   0.425898 -0.717816
factor(DIV)3  0.0828629   0.429408  0.192970   0.306413  0.270429
factor(DIV)4 -1.0648175   0.369623 -2.880818   0.328896 -3.237552
factor(DIV)5 -0.1026079   0.357922 -0.286677   0.419724 -0.244465
factor(DIV)6 -0.5820502   0.350262 -1.661759   0.246651 -2.359817
PTS.f(10,20]  0.5662433   0.310847  1.821614   0.197153  2.872106
PTS.f(20,30]  1.4506816   0.333275  4.352810   0.459362  3.158038
PTS.f(30,50]  1.2452296   0.478216  2.603906   0.427261  2.914444

Estimated Scale Parameter: 1.01931
Number of Iterations: 3

# stukel function
stukel <- function(object, alternative = c("both", "alpha1",
  "alpha2")) { DNAME <- deparse(substitute(object))
  METHOD <- "Stukel's test of the logistic link"
  alternative <- match.arg(alternative)
```

```

eta <- predict(object, type = "link")
etasq <- 0.5 * eta * eta
etapos <- eta > 0
dv <- matrix(0, nrow = length(eta), ncol = 2)
dv[etapos,1] <- etasq[etapos]
dv[!etapos,2] <- - etasq[!etapos]
colnames(dv) <- c("z1", "z2")
oinfo <- vcov(object)
oX <- qr.X(object$qr)
ImH <- - oX %*% oinfo %*% t(oX)
diag(ImH) <- 1 + diag(ImH)
wdv <- sqrt(object$weights) * dv
qmat <- t(wdv) %*% ImH %*% wdv
sc <- apply(dv * (object$weights * residuals(object,
"working")), 2, sum)
allstat <- c(sc * sc / diag(qmat), sc %*% solve(qmat) %*% sc)
names(allstat) <- c("alpha1", "alpha2", "both")
allpar <- c(1,1,2)
names(allpar) <- names(allstat)
allpval <- pchisq(allstat, allpar, lower.tail=FALSE)
STATISTIC <- allstat[alternative]
PARAMETER <- allpar[alternative]
names(PARAMETER) <- "df"
PVAL <- allpval[alternative]
names(allpar) <- rep("df", 3)
structure(list(statistic = STATISTIC,
               parameter = PARAMETER,
               p.value = PVAL,
               alternative = alternative,
               method = METHOD, data.name = DNAME,
               allstat = allstat, allpar = allpar, allpval = allpval
            ),
          class = "htest")
}

# stukel score link test
> stukel(yao.logit2)

Stukel's test of the logistic link

data:  yao.logit2
both = 0.9793, df = 2, p-value = 0.6128
alternative hypothesis: both

```



```
> stukel(yao.logit2,alternative='alpha1')
```

Stukel's test of the logistic link

data: yao.logit2

alpha1 = 0.6042, df = 1, p-value = 0.437

alternative hypothesis: alpha1

```
> stukel(yao.logit2,alternative='alpha2')
```

Stukel's test of the logistic link

data: yao.logit2

alpha2 = 0.972, df = 1, p-value = 0.3242

alternative hypothesis: alpha2