GLM: France Soccer Championship, 2007-2008

**Goals:**

Analyze a real data set on soccer championship (and compare the results with that of Lab 5)
Learn how to construct CI for model parameters and predicted proportions

**Assignments:**

Use the data on France Soccer Championship 2007-2008.

a) Construct a GLM to predict the probability to win a game (response) using the current standing of both teams (explanatory). (Decide whether you need the quadratic terms in this problem; justify your choice).
b) Discuss the model results. What is the predicted dependence of the game result on the team standing? Does this supports/contradicts the possibility of bribing? Why or why not?
c) Find 95% CIs for all model parameters.
d) Find a 95% CI for the probability to win when the team score is 50 and the opponent’s score is 10.

**Reports:** Assignments require printed report, which will consist of R-results (do not print the entire session, only the necessary results!) and plots. Describe briefly the theoretical background for the methods you use, including necessary formulas, and make short statements about result interpretation. Consult instructor if you have any questions about the level of detail or formatting of your report.

Reports are due on December 1st
Essential R commands:

Session management:
- help()
- ls()
- getwd()
- setwd()
- library()
- data()
- save()
- load()
- read.table()
- class()
- names()
- rm()

Vectors:
- c()
- seq()
- rep()
- factor()
- cbind()
- rbind()

Data summaries:
- mean()
- sd()
- median()
- quantile()
- summary()

Graphs:
- par()
- plot()
- points()
- lines()
- mosaicplot()
- text()

GLMs:
- glm()
- family()
- summary.glm()
- predict.glm()
- confint()
- summary()
- predict()
# read data
T<-read.table(file='France_2007_2008.txt',header=TRUE)

# make R recognize the names of variables from "T"
attach(T)

# Tie indicator
T<-factor(R1==1)  # Tie indicator

# yellow background
par(bg='yellow')

# plot tie indicator vs. player 1 points
plot(P1,jitter(as.numeric(T)),pch=19,col='blue')

# plot tie indicator vs. P1-P2
plot(P1-P2,jitter(as.numeric(T),amount=.1),pch=19,col='blue',
ylab='Jittered tie indicator (FALSE=1)')

# Histograms of point difference for tie/no tie
hist(P1[T==FALSE]-P2[T==FALSE],col='blue',
main='No tie',xlab='Point difference',xlim=c(-40,40))

hist(P1[T==TRUE]-P2[T==TRUE],col='blue',
main='Tie',xlab='Point difference',xlim=c(-40,40))

# GLM
#===============================================
g<-glm(T~I(P1-P2)+I((P1-P2)^2),family=binomial)
g<-glm(T~I((P1-P2)^2),family=binomial)
summary(g)

# CI for model parameters
#===============================================
confint(g,level=.99)

# CI for predicted proportions
#===============================================
p<-predict(g,data.frame(P1=10,P2=5),type='response',se.fit=TRUE)
L<-p$fit - qnorm(1-.025)*p$se.fit
R<-p$fit + qnorm(1-.025)*p$se.fit

# Plot predictions
#===============================================
plot(P1-P2,predict(g,data.frame(P1,P2),type='response'),
pch=19,col='blue',ylab='Probability of a tie')
grid()