**TASK I**

**I.I**

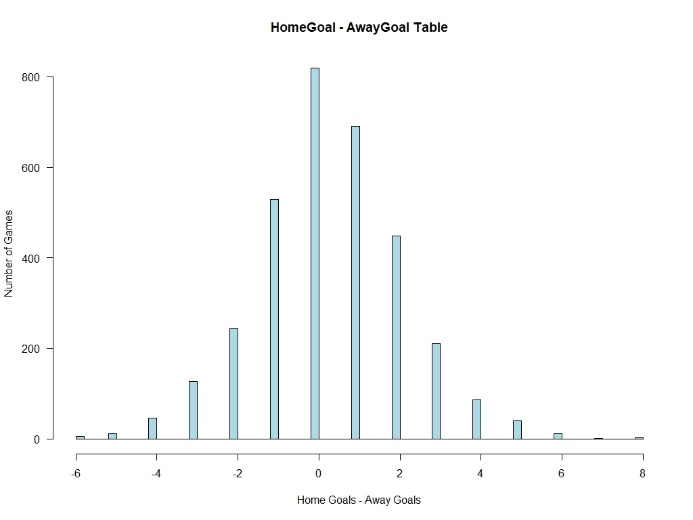
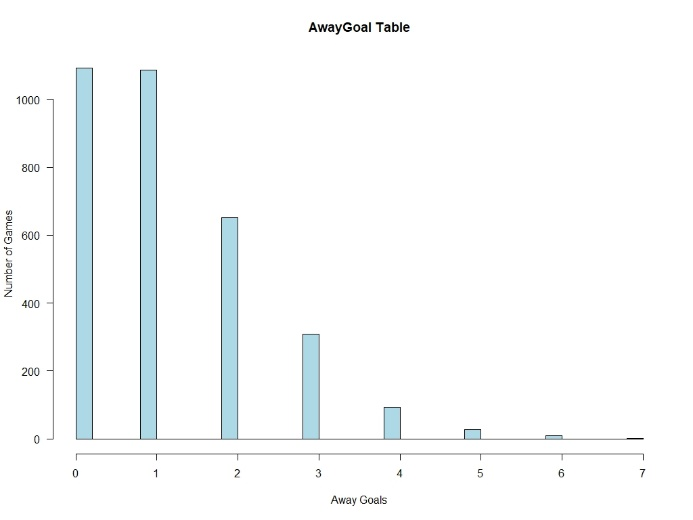
In this part of the Task I, in order to create theHomeGoal Table and AwayGoal Table, the HomeGoal and AwayGoal columns are added to the matches table and the number of goals are stated. The histograms below belong to HomeGoal, AwayGoal and HomeGoal-AwayGoal respectively.

**Figure 1: HomeGoal table with number of games**

(goals)

(goals)

(goals)



**Figure 3: AwayGoal table with number of games**

**Figure 2: Difference of HomeGoal and AwayGoal table with number of games**

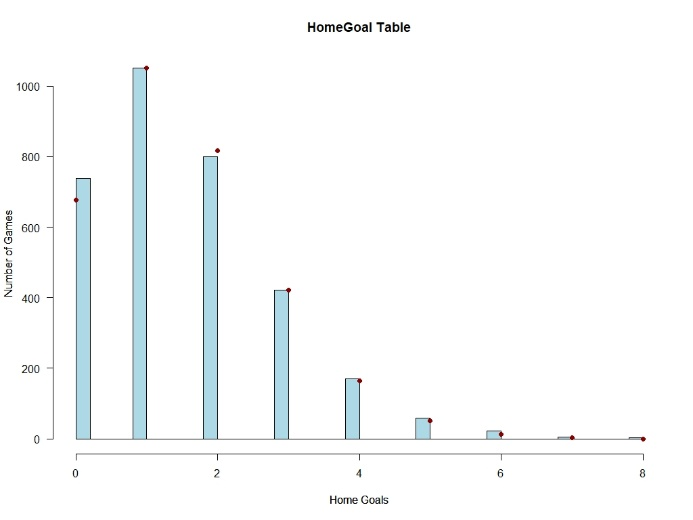
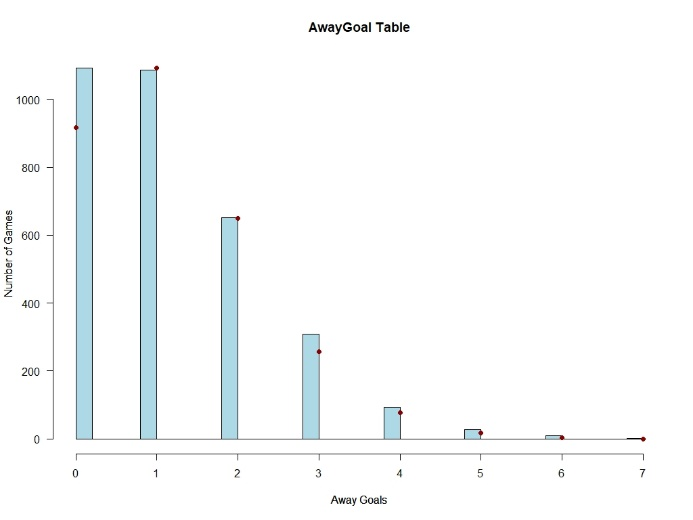
**of games**

(goals)

(goals)

**I.II**

In this part of the Task 1, observing the distribution type, we claim that HomeGoal and AwayGoal are Poisson distributed. In order to verify our statement, calculating the mean of the HomeGoal, AwayGoal and HomeGoal-AwayGoal we found parameters for Poisson distribution to compare the sample distribution and theoretic Poisson distribution. The Poisson distribution with lambda=1.553776 in range 0:8 is plotted over to each histogram.

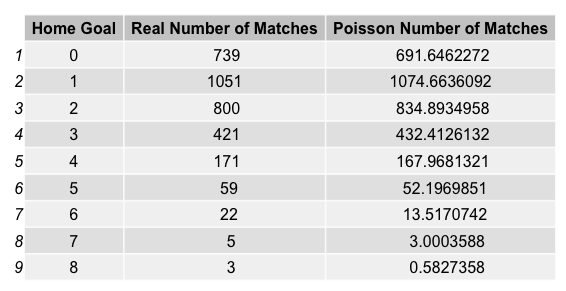
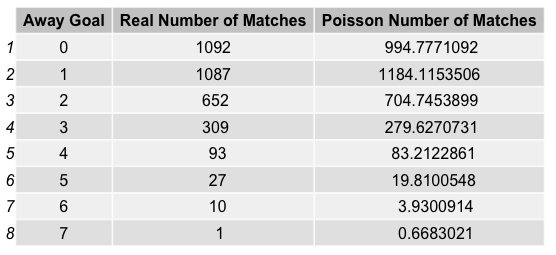


**Figure 5 : Away goals with poisson distribution**

(goals)

**Figure 4: Home goals with poisson distribution**

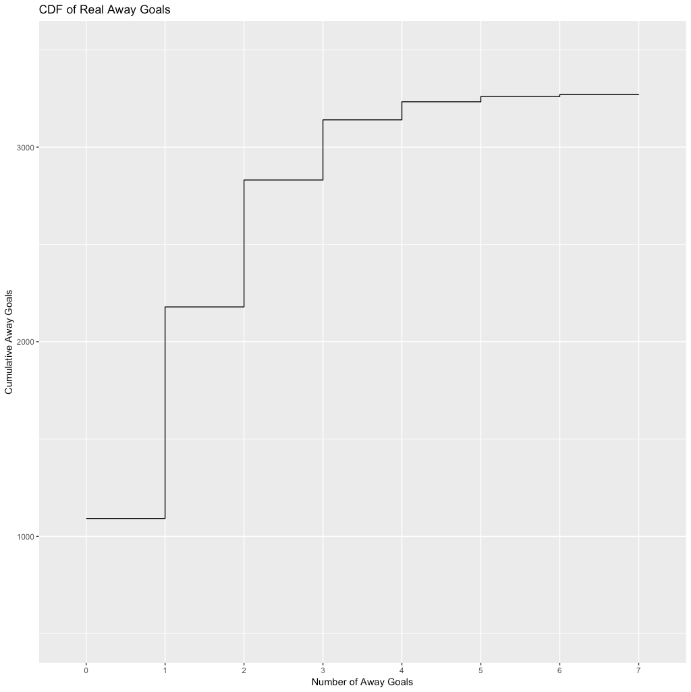
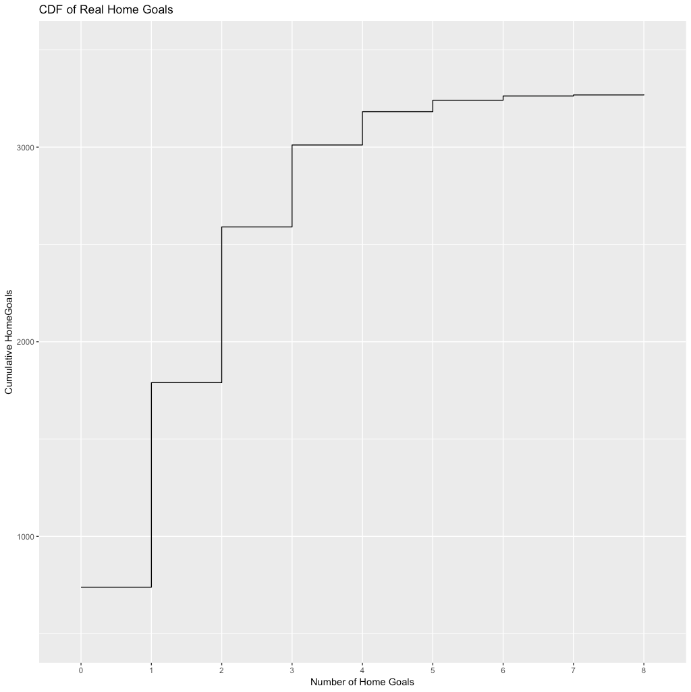
(goals)



**Table 2: Number of Matches from Data and Number of Matches from Theoretical Poisson Distribution According to Away Goal**

**Table 1: Number of Matches from Data and Number of Matches from Theoretical Poisson Distribution According to Home Goal**

We observed the similarity between the data and theoretical Poisson distribution from the Figure 4 and Figure 5. However, in order to support our claim, we formed Table 1 and Table 2 and obtained the numerical values for each Home Goal number and Away Goal number for data and Poisson distribution. Evaluating the result from Figure 4, Figure 5 and Table 1, Table 2, we concluded that our claim is true, and data is consistent with the Poisson distribution.

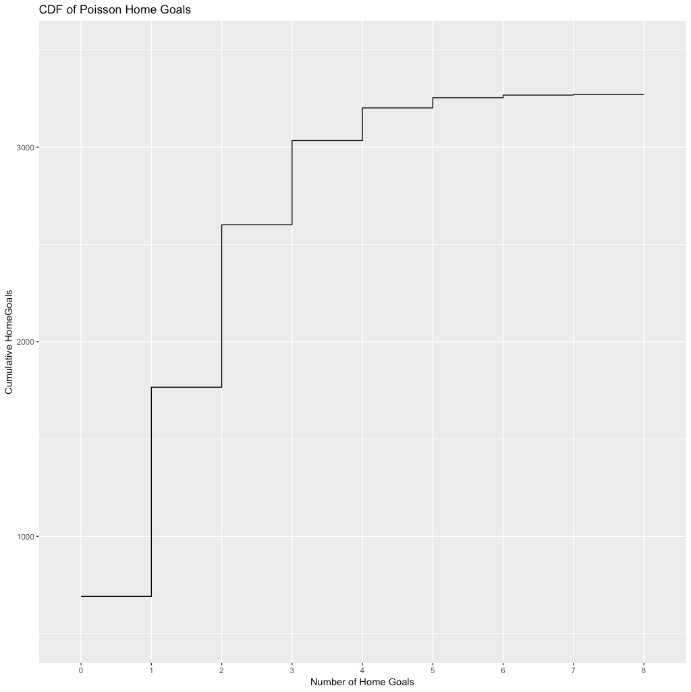
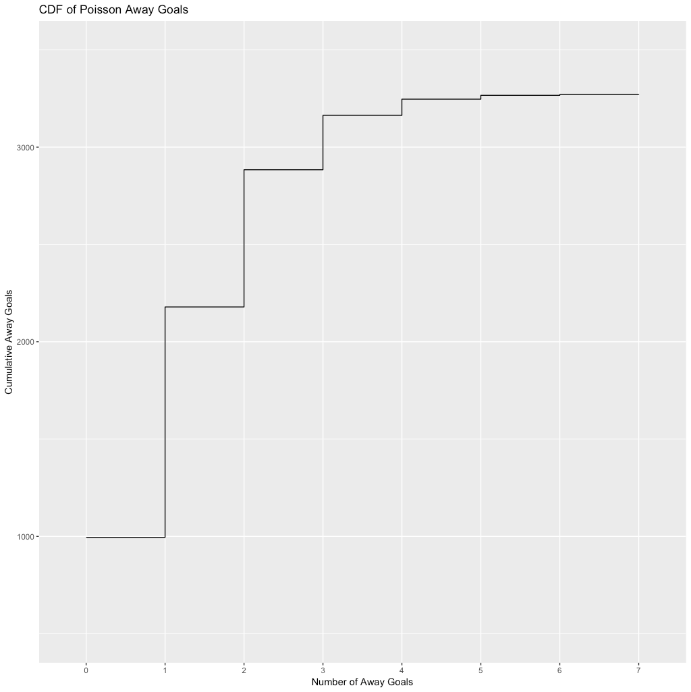


**Figure 7: CDF of Real Away Goals with Number of Home Goals and Cumulative Away Goals**

**Figure 6: CDF of Real Home Goals with Number of Home Goals and Cumulative Home Goals**

In order to calculate the expected number of games corresponding to each quantile (number of goals) with Poisson distribution, we plotted cumulative distribution functions for Real Home Goals, Real Away Goals and Poisson Home Goals and Poisson Away Goals.

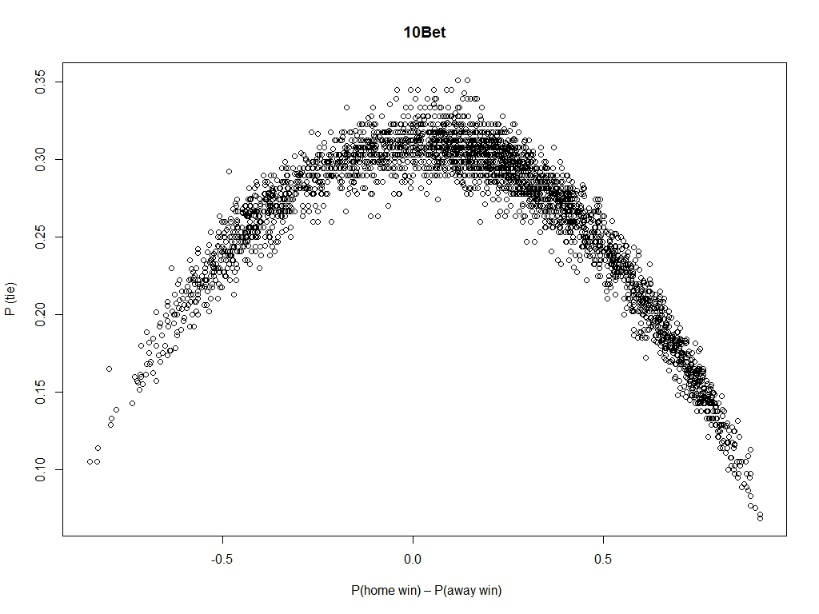
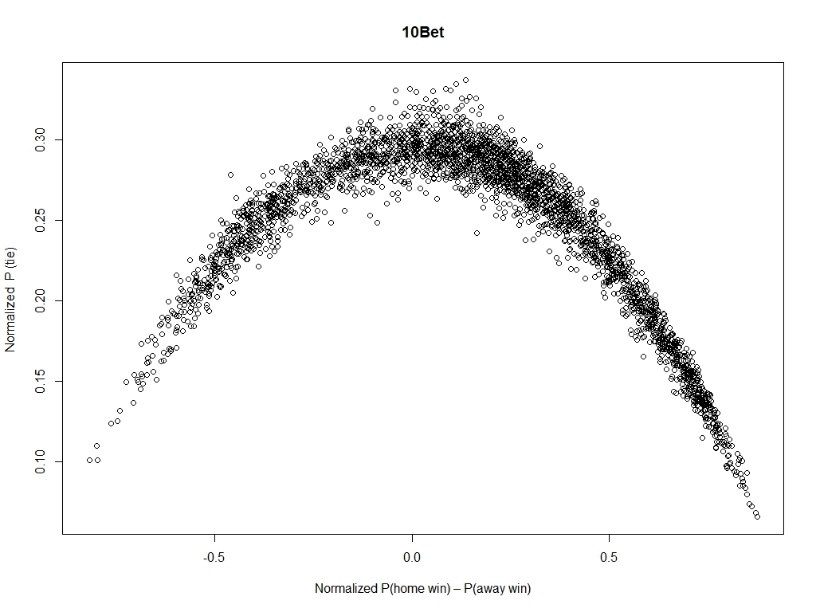
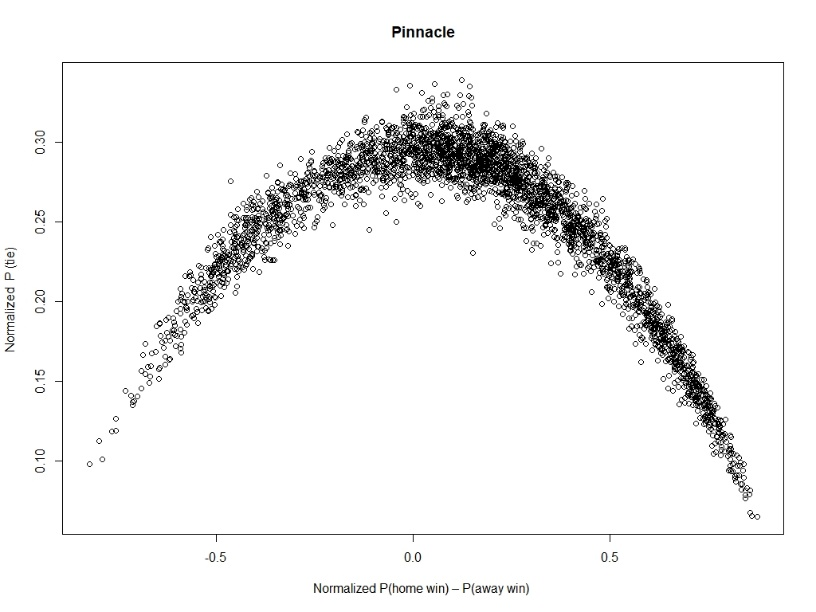
The similarity between Real and Poisson plots are observed to be high again. To conclude, our claim is turned to be true.



**Figure 9: CDF of Poisson Away Goals with Number of Away Goals and Cumulative Away Goals**

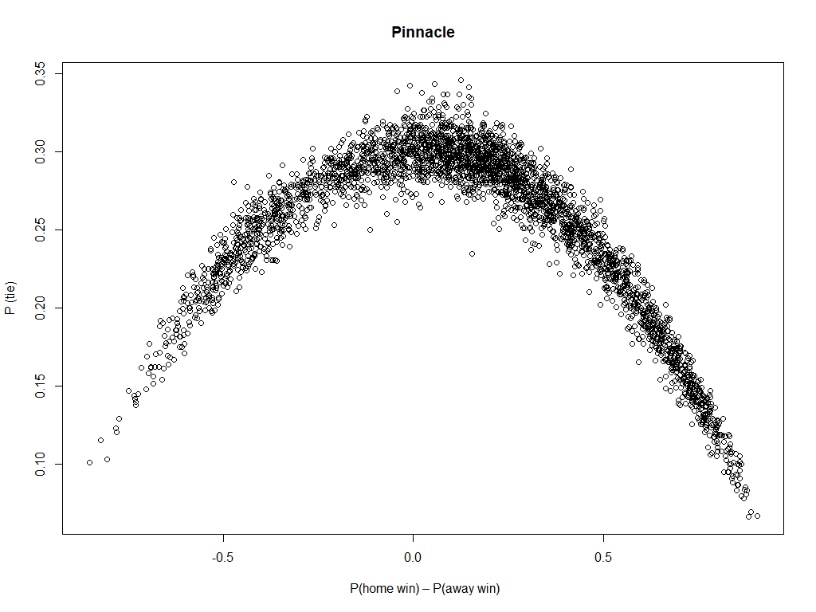
**Figure 8: CDF of Poisson Home Goals with Number of Home Goals and Cumulative Home Goals**

**TASK II**

 In this part, we first calculated the P(home win),P(away win) and P(tie) by dividing 1 by odds for each result that are given by bookmarkers. Since the total probability given by bookmarkers sum up to a value bigger than 1, we normalized the probabilities. Then, we created two plots for each bookmarker that is chosen, the first plots for each bookmarker shows the non-normalized probabilities and the second ones represent the normalized probabilities. We discretized P(home win)–P(away win) values into bins like [-1,-0.95),[-0.95,-0.90).. to (0.95,1] and calculate the number of games ended as raw in the corresponding bin. In order to observe the differences between them, the two plots are put one under the other.

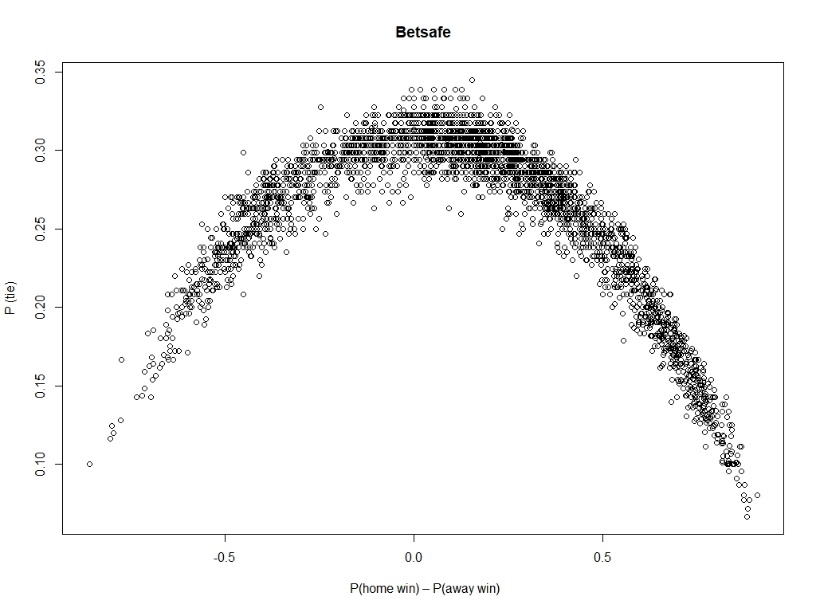
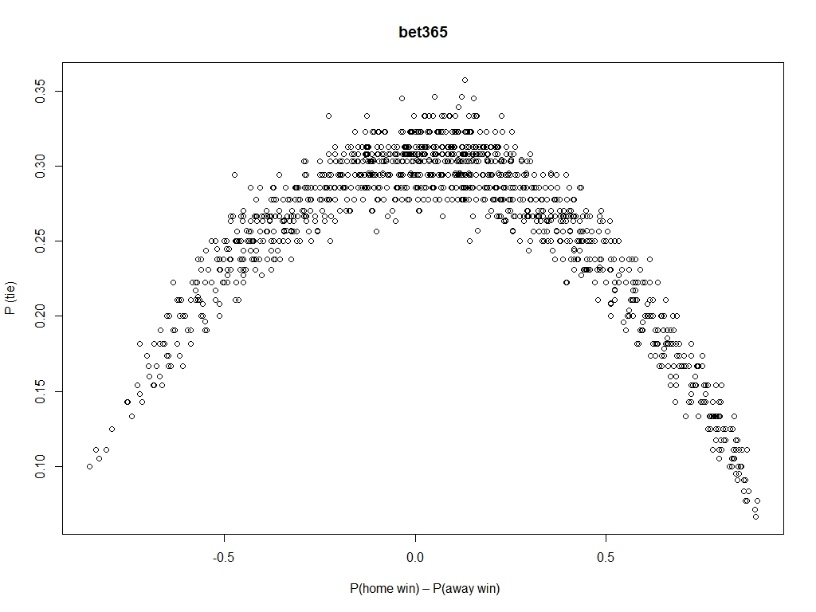
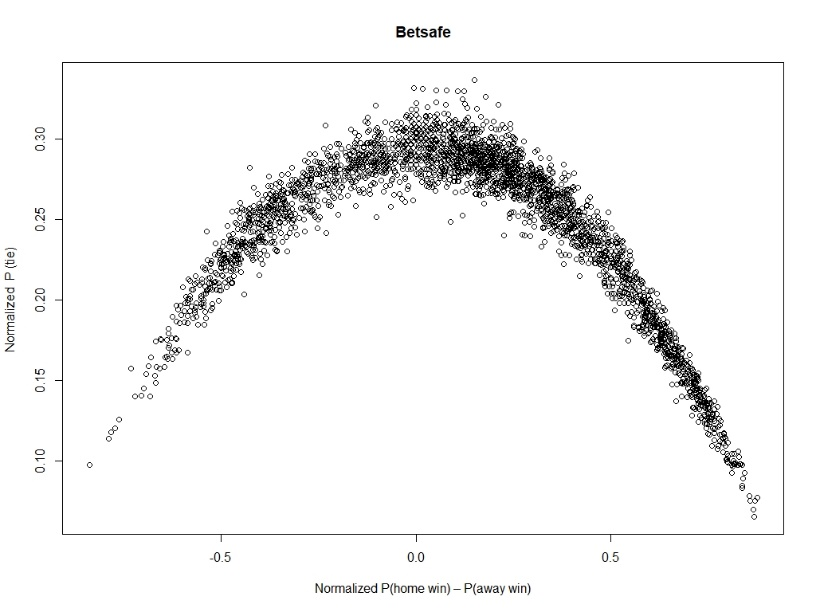
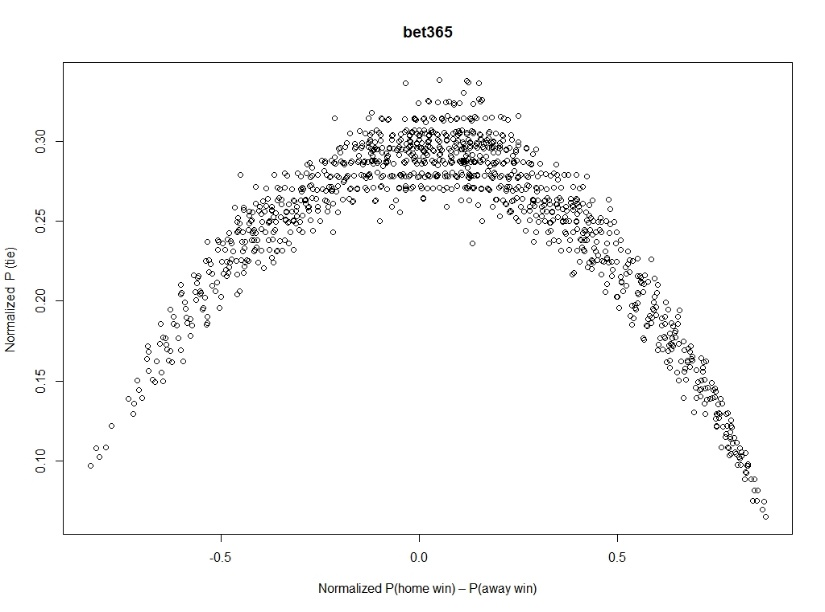
**Figure 11: 10Bet’s P(home win)-P(away win) vs P(tie)**

**Figure 10: Pinnacle’s P(home win)-P(away win) vs P(tie)**



**Figure 13: 10Bet’s Normalized P(home win)-P(away win) vs Normalized P(tie)**

**Figure 12: Pinnacle’s Normalized P(home win)-P(away win) vs Normalized P(tie)**



**Figure 15: Bet365’s P(home win)-P(away win) vs P(tie)**

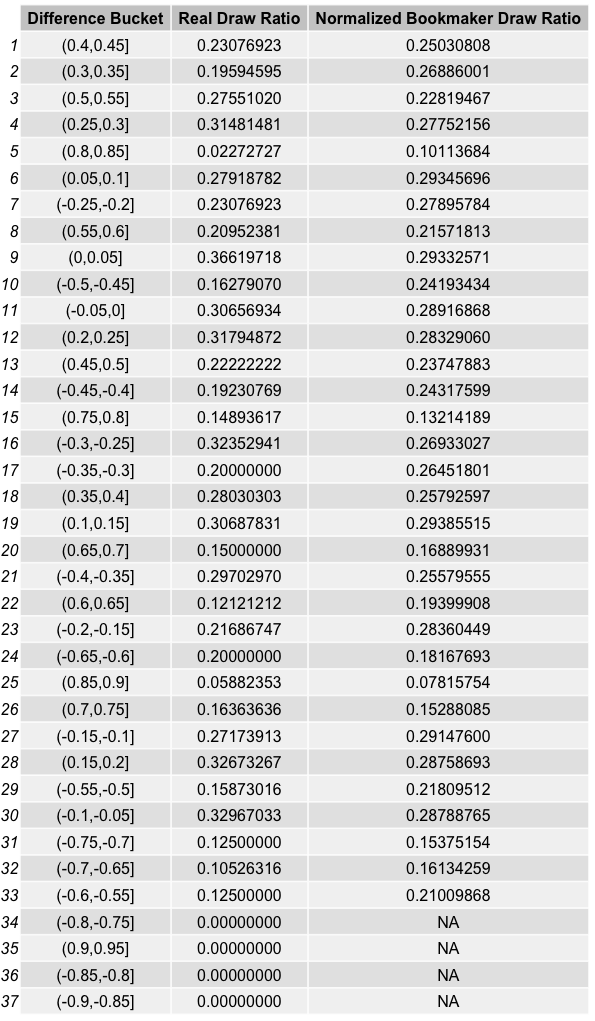
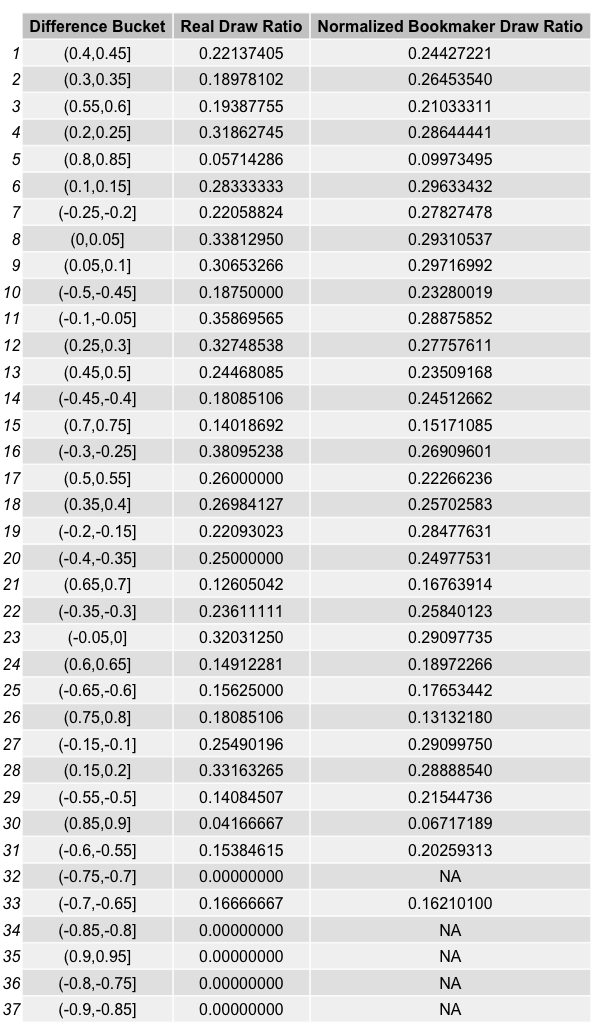
**Figure 16: Betsafe’s Normalized P(home win)-P(away win) vs Normalized P(tie)**

**Figure 17: Bet365’s Normalized P(home win)-P(away win) vs Normalized P(tie)**

**Figure 14: Betsafe’s P(home win)-P(away win) vs P(tie)**

**II.IV**

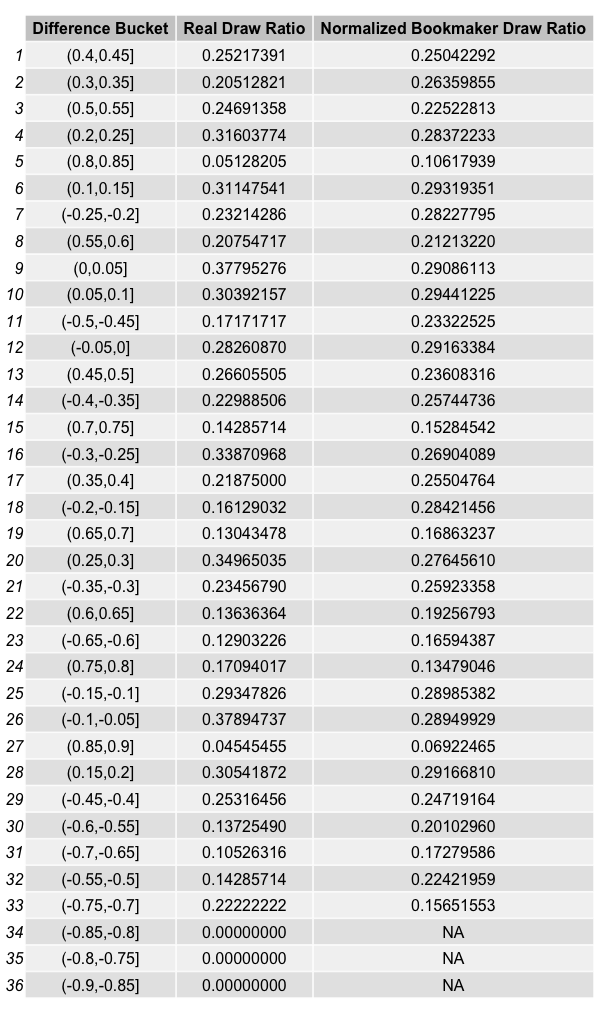
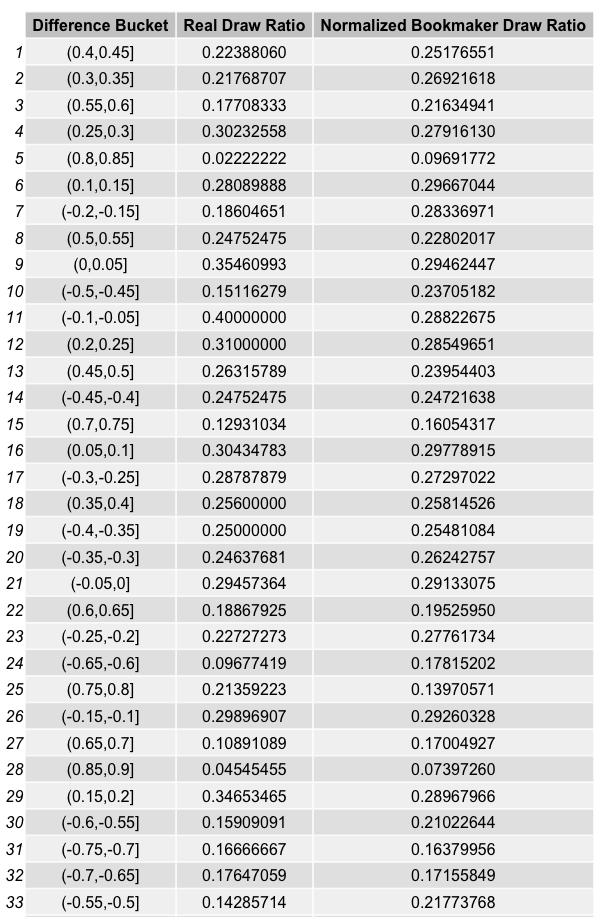
When the difference between real draw ratio and bookmaker draw ratio is observed, it makes sense to bet on draw if real draw ratio is bigger than bookmaker draw ratio. Since the real probability of a match ending draw has larger probability than the probability that bookmaker estimate, there is a chance to gain money due to bias.



**Table4:Result Summary Ratios of Betsafe**

**Table3:Result Summary Ratios of Pinnacle**

**Table5:Result Summary Ratios of bet365**

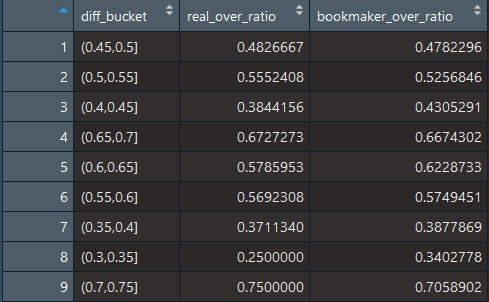


**Table6:Result Summary Ratios of Bet10**

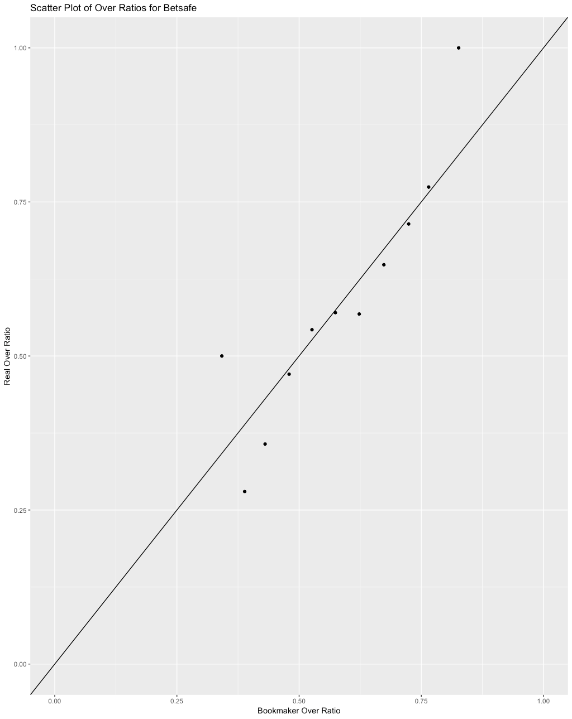
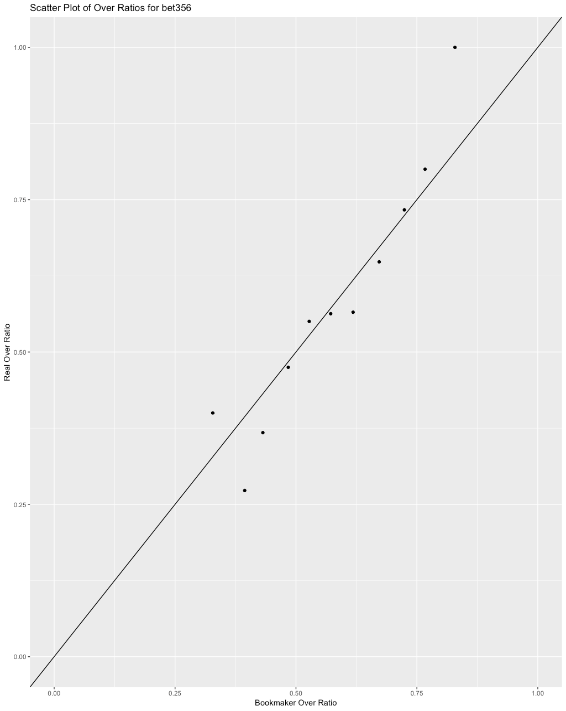
**III.I**

In this part, first we filtered the odds table by taking the ‘ou’ betType and 2.5 totalhandicap for the bookmarker Pinnacle. We observed that there are more than 1 over odds for some matches given by the Pinnacle so in order to obtain the latest odd , we created a new table called latest\_odds. Then, similar to the previous task, we calculated the probabilities of over and under using the odds given by Pinnacle and then normalized the probabilities.

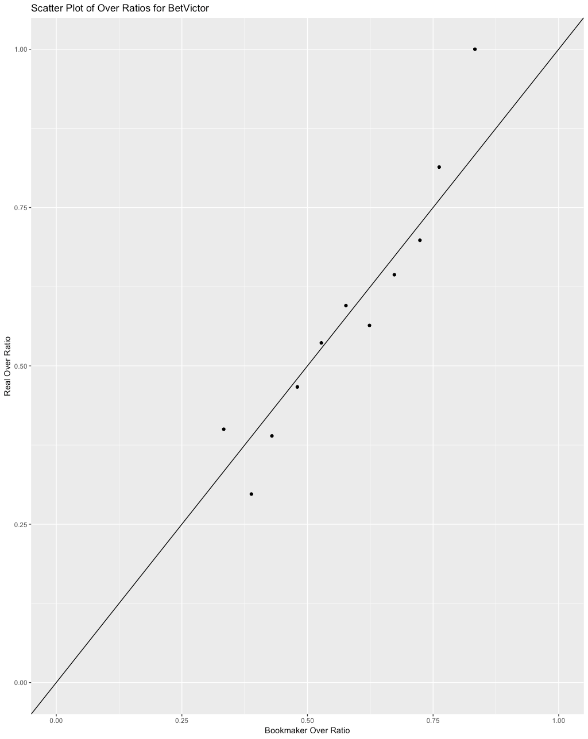
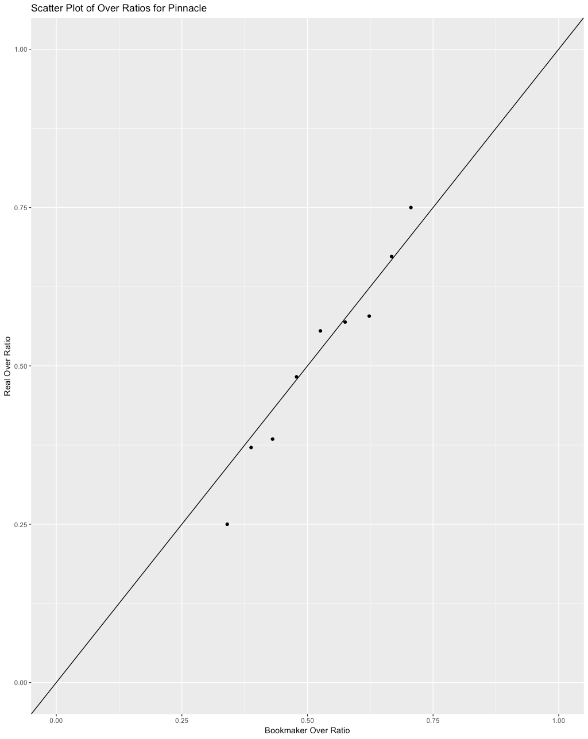
In order to detect the over ending matches, we wanted to search for total goals bigger than 2.5, thus more than or equal to 3. To do so, we calculated TotalGoal and merged it to the latest\_odds table. For the over ending matches, we calculated real\_over\_ratio and bookmaker\_over\_ratio. While doing so, there occurred NA cells in TotalGoal column due to matches not played yet, we solved this problem by using na.rm=TRUE.

 **Result Summary Table**

**Table 7:Result Summary Table for Pinnacle**



**Figure 18: Scatter Plot of bet365 Figure 19: Scatter Plot of Betsafe**

In order to visualize our work, we used ggplot2 library and created a scatter plot with reference line x=y. We did this for 4 different bookmakers and observed how data is distributed along the reference line.

**Figure 20: Scatter Plot of BetVictor Figure 21: Scatter Plot of Pinnacle**

**III.II**

In this part, we tried to observe the reliability of a bookmaker in years. We converted the epoch time units to Turkey’s local time and date. Next, we determined a certain bucket range in which we compared the mean of over probabilities given by bookmaker and mean of real over probabilities in each year.

metin, harita içeren bir resim

Açıklama otomatik olarak oluşturuldu

:2011 :2015 :2013

:2012 :2017

:2016 :2018

**Ratio**

**Ratio**

**Figure 22: Representation of change of odds**

**R codes for Plotting Figures**i. Code for Figure 1:

summary\_by\_homegoal=matches[,list(count=.N),by=list(matchId,HomeGoal)]

factor(summary\_by\_homegoal$HomeGoal)

table\_for\_homegoal=table(summary\_by\_homegoal$HomeGoal)

table\_for\_homegoal

hist(summary\_by\_homegoal$HomeGoal,main = "HomeGoal Table", xlab = "Home Goals", ylab = "Number of Games", las =1, breaks = 30,col='light blue')

ii. Code for Figure 2:

summary\_by\_homegoal\_and\_awaygoal=matches[,list(count=.N),by=list(matchId,HomeGoal,AwayGoal)]

summary\_by\_homegoal\_and\_awaygoal

homegoal\_minus\_awaygoal=summary\_by\_homegoal\_and\_awaygoal[,list(count=.N),by=list(matchId,HomeGoal-AwayGoal)]

homegoal\_minus\_awaygoal[,c("HomeGoal-AwayGoal"):=HomeGoal]

homegoal\_minus\_awaygoal$HomeGoal=NULL

homegoal\_minus\_awaygoal$count=NULL

homegoal\_minus\_awaygoal

hist(homegoal\_minus\_awaygoal$`HomeGoal-AwayGoal`,main = "HomeGoal - AwayGoal Table", xlab = "Home Goals - Away Goals", ylab = "Number of Games", las =1, breaks = 60,col='light blue')

iii. Code for Figure 3:

summary\_by\_awaygoal=matches[,list(count=.N),by=list(matchId,AwayGoal)]

factor(summary\_by\_awaygoal$AwayGoal)

table\_for\_awaygoal=table(summary\_by\_awaygoal$AwayGoal)

table\_for\_awaygoal

hist(summary\_by\_awaygoal$AwayGoal,main = "AwayGoal Table", xlab = "Away Goals", ylab = "Number of Games",las=1, breaks = 30,col='light blue')

iv. Code for Figure 6:

summary\_by\_homegoal=matches[,list(count=.N),by=list(matchId,HomeGoal)]

factor(summary\_by\_homegoal$HomeGoal)

table\_for\_homegoal=table(summary\_by\_homegoal$HomeGoal)

table\_for\_homegoal

hist(summary\_by\_homegoal$HomeGoal,main = "HomeGoal Table", xlab = "Home Goals", ylab = "Number of Games", las =1, breaks = 30,col='light blue')

mean\_homegoal=mean(matches$HomeGoal,na.rm = T)

mean\_homegoal

par(new=TRUE)

plot(dpois(x=0:8,lambda=mean\_homegoal), xlab = "Home Goals",ylab="Number of Games",axes=F,col='dark red',pch=19)

HomeGoal\_pois=c(dpois(0,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(1,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(2,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(3,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(4,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(5,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(6,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(7,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(8,mean\_homegoal)\*sum(table\_for\_homegoal))

real\_vs\_poison\_homegoal=data.table(Real\_HomeGoal=table\_for\_homegoal,Poison\_HomeGoal=HomeGoal\_pois)

ggplot(real\_vs\_poison\_homegoal, aes(real\_vs\_poison\_homegoal$Real\_HomeGoal.V1, cumsum(real\_vs\_poison\_homegoal$Real\_HomeGoal.N))) +

geom\_step(aes(group=1))+

ggtitle("CDF of Real Home Goals")+

xlab("Number of Home Goals")+

ylab("Cumulative HomeGoals")+

ylim(500, 3500)

v. Code for Figure 7:

summary\_by\_awaygoal=matches[,list(count=.N),by=list(matchId,AwayGoal)]

factor(summary\_by\_awaygoal$AwayGoal)

table\_for\_awaygoal=table(summary\_by\_awaygoal$AwayGoal)

table\_for\_awaygoal

hist(summary\_by\_awaygoal$AwayGoal,main = "AwayGoal Table", xlab = "Away Goals", ylab = "Number of Games",las=1, breaks = 30,col='light blue')

mean\_awaygoal=mean(matches$AwayGoal,na.rm = T)

par(new=TRUE)

plot(dpois(x=0:7,lambda=mean\_awaygoal), xlab = "Away Goals",ylab="Number of Games",axes=F,col='dark red',pch=19)

AwayGoal\_pois=c(dpois(0,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(1,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(2,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(3,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(4,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(5,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(6,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(7,mean\_awaygoal)\*sum(table\_for\_awaygoal))

real\_vs\_poison\_awaygoal=data.table(Real\_AwayGoal=table\_for\_awaygoal,Poison\_AwayGoal=AwayGoal\_pois)

ggplot(real\_vs\_poison\_awaygoal, aes(real\_vs\_poison\_awaygoal$Real\_AwayGoal.V1, cumsum(real\_vs\_poison\_awaygoal$Real\_AwayGoal.N))) +

geom\_step(aes(group=1))+

ggtitle("CDF of Real Away Goals")+

xlab("Number of Away Goals")+

ylab("Cumulative Away Goals")+

ylim(500, 3500)

vi. Code for Figure 8:

summary\_by\_homegoal=matches[,list(count=.N),by=list(matchId,HomeGoal)]

factor(summary\_by\_homegoal$HomeGoal)

table\_for\_homegoal=table(summary\_by\_homegoal$HomeGoal)

table\_for\_homegoal

hist(summary\_by\_homegoal$HomeGoal,main = "HomeGoal Table", xlab = "Home Goals", ylab = "Number of Games", las =1, breaks = 30,col='light blue')

mean\_homegoal=mean(matches$HomeGoal,na.rm = T)

mean\_homegoal

par(new=TRUE)

plot(dpois(x=0:8,lambda=mean\_homegoal), xlab = "Home Goals",ylab="Number of Games",axes=F,col='dark red',pch=19)

HomeGoal\_pois=c(dpois(0,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(1,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(2,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(3,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(4,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(5,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(6,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(7,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(8,mean\_homegoal)\*sum(table\_for\_homegoal))

real\_vs\_poison\_homegoal=data.table(Real\_HomeGoal=table\_for\_homegoal,Poison\_HomeGoal=HomeGoal\_pois)

ggplot(real\_vs\_poison\_homegoal, aes(real\_vs\_poison\_homegoal$Real\_HomeGoal.V1, cumsum(real\_vs\_poison\_homegoal$Poison\_HomeGoal))) +

geom\_step(aes(group=1))+

ggtitle("CDF of Poisson Home Goals")+

xlab("Number of Home Goals")+

ylab("Cumulative HomeGoals")+

ylim(500, 3500)

vii. Code for Figure 9:

summary\_by\_awaygoal=matches[,list(count=.N),by=list(matchId,AwayGoal)]

factor(summary\_by\_awaygoal$AwayGoal)

table\_for\_awaygoal=table(summary\_by\_awaygoal$AwayGoal)

table\_for\_awaygoal

hist(summary\_by\_awaygoal$AwayGoal,main = "AwayGoal Table", xlab = "Away Goals", ylab = "Number of Games",las=1, breaks = 30,col='light blue')

mean\_awaygoal=mean(matches$AwayGoal,na.rm = T)

par(new=TRUE)

plot(dpois(x=0:7,lambda=mean\_awaygoal), xlab = "Away Goals",ylab="Number of Games",axes=F,col='dark red',pch=19)

AwayGoal\_pois=c(dpois(0,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(1,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(2,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(3,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(4,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(5,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(6,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(7,mean\_awaygoal)\*sum(table\_for\_awaygoal))

real\_vs\_poison\_awaygoal=data.table(Real\_AwayGoal=table\_for\_awaygoal,Poison\_AwayGoal=AwayGoal\_pois)

ggplot(real\_vs\_poison\_awaygoal, aes(real\_vs\_poison\_awaygoal$Real\_AwayGoal.V1, cumsum(real\_vs\_poison\_awaygoal$Poison\_AwayGoal))) +

geom\_step(aes(group=1))+

ggtitle("CDF of Poisson Away Goals")+

xlab("Number of Away Goals")+

ylab("Cumulative Away Goals")+

ylim(500, 3500)

viii. Code for Figure 10:

filtered\_odds=odds[betType=='1x2' & bookmaker=='Pinnacle']

filtered\_odds[,c('betType','bookmaker','totalhandicap'):=NULL]

filtered\_odds=filtered\_odds[order(matchId, oddtype,date)]

latest\_odds=filtered\_odds[,list(final\_odd=odd[.N]),by=list(matchId,oddtype)]

help(dcast)

latest\_odds=dcast(latest\_odds,matchId~oddtype,value.var='final\_odd')

temp=matches[,list(matchId,date\_of\_match,home,away,MatchResult)]

matches\_with\_odds=merge(temp,latest\_odds,by='matchId')

summary\_odds\_by\_result=matches\_with\_odds[,list(mean\_home=mean(odd1),

mean\_draw=mean(oddX),mean\_away=mean(odd2),.N),by=list(MatchResult)]

matches\_with\_odds[,prob\_home:=1/odd1]

matches\_with\_odds[,prob\_draw:=1/oddX]

matches\_with\_odds[,prob\_away:=1/odd2]

matches\_with\_odds[,total\_prob:=prob\_home+prob\_draw+prob\_away]

matches\_with\_odds[,home\_away\_diff:=prob\_home-prob\_away]

plot(matches\_with\_odds[,list(home\_away\_diff,prob\_draw)])

cut\_levels=c(-20:20)/20

matches\_with\_odds[,diff\_bucket:=cut(home\_away\_diff,cut\_levels)]

result\_summary=matches\_with\_odds[,list(real\_draw\_ratio=sum(MatchResult=='draw', na.rm = T)/.N,draw\_prob\_bookmaker=mean(prob\_draw[MatchResult=='draw'], na.rm = T)),by=list(diff\_bucket)]

ix. Code for Figure 11:

filtered\_odds2=odds[betType=='1x2' & bookmaker=='10Bet']

filtered\_odds2[,c('betType','bookmaker','totalhandicap'):=NULL]

filtered\_odds2=filtered\_odds2[order(matchId, oddtype,date)]

latest\_odds2=filtered\_odds2[,list(final\_odd=odd[.N]),by=list(matchId,oddtype)]

latest\_odds2=dcast(latest\_odds2,matchId~oddtype,value.var='final\_odd')

temp2=matches[,list(matchId,date\_of\_match,home,away,MatchResult)]

matches\_with\_odds2=merge(temp2,latest\_odds2,by='matchId')

summary\_odds\_by\_result2=matches\_with\_odds2[,list(mean\_home=mean(odd1),

mean\_draw=mean(oddX),mean\_away=mean(odd2),.N),by=list(MatchResult)]

matches\_with\_odds2[,prob\_home:=1/odd1]

matches\_with\_odds2[,prob\_draw:=1/oddX]

matches\_with\_odds2[,prob\_away:=1/odd2]

matches\_with\_odds2[,total\_prob:=prob\_home+prob\_draw+prob\_away]

matches\_with\_odds2[,home\_away\_diff:=prob\_home-prob\_away]

plot(matches\_with\_odds2[,list(home\_away\_diff,prob\_draw)])

matches\_with\_odds2[,diff\_bucket:=cut(home\_away\_diff,cut\_levels)]

result\_summary2=matches\_with\_odds2[,list(real\_draw\_ratio=sum(MatchResult=='draw', na.rm = T)/.N,draw\_prob\_bookmaker=mean(prob\_draw[MatchResult=='draw'], na.rm = T)),by=list(diff\_bucket)]

xi. Code for Figure 12:

matches\_with\_odds[,P\_home:=prob\_home/total\_prob]

matches\_with\_odds[,P\_away:=prob\_away/total\_prob]

matches\_with\_odds[,P\_draw:=prob\_draw/total\_prob]

matches\_with\_odds[,P\_home\_away\_diff:=P\_home-P\_away]

P\_summary\_odds\_by\_result=matches\_with\_odds[,list(mean\_home=mean(P\_home),

mean\_draw=mean(P\_draw),mean\_away=mean(P\_away),.N),by=list(MatchResult)]

plot(matches\_with\_odds[,list(P\_home\_away\_diff,P\_draw)])

matches\_with\_odds[,P\_diff\_bucket:=cut(P\_home\_away\_diff,cut\_levels)]

P\_result\_summary=matches\_with\_odds[,list(real\_draw\_ratio=sum(MatchResult=='draw', na.rm = T)/.N,P\_draw\_prob\_bookmaker=mean(P\_draw[MatchResult=='draw'], na.rm = T)),by=list(diff\_bucket)]

xii. Code for Figure 13:

matches\_with\_odds2[,P\_home:=prob\_home/total\_prob]

matches\_with\_odds2[,P\_away:=prob\_away/total\_prob]

matches\_with\_odds2[,P\_draw:=prob\_draw/total\_prob]

matches\_with\_odds2[,P\_home\_away\_diff:=P\_home-P\_away]

P\_summary\_odds\_by\_result2=matches\_with\_odds2[,list(mean\_home=mean(P\_home),

mean\_draw=mean(P\_draw),mean\_away=mean(P\_away),.N),by=list(MatchResult)]

plot(matches\_with\_odds2[,list(P\_home\_away\_diff,P\_draw)])

matches\_with\_odds2[,P\_diff\_bucket:=cut(P\_home\_away\_diff,cut\_levels)]

P\_result\_summary2=matches\_with\_odds2[,list(real\_draw\_ratio=sum(MatchResult=='draw', na.rm = T)/.N,P\_draw\_prob\_bookmaker=mean(P\_draw[MatchResult=='draw'], na.rm = T)),by=list(diff\_bucket)]

xiii. Code for Figure 14:

filtered\_odds3=odds[betType=='1x2' & bookmaker=='Betsafe']

filtered\_odds3[,c('betType','bookmaker','totalhandicap'):=NULL]

filtered\_odds3=filtered\_odds3[order(matchId, oddtype,date)]

latest\_odds3=filtered\_odds3[,list(final\_odd=odd[.N]),by=list(matchId,oddtype)]

latest\_odds3=dcast(latest\_odds3,matchId~oddtype,value.var='final\_odd')

temp3=matches[,list(matchId,date\_of\_match,home,away,MatchResult)]

matches\_with\_odds3=merge(temp3,latest\_odds3,by='matchId')

summary\_odds\_by\_result3=matches\_with\_odds3[,list(mean\_home=mean(odd1),

mean\_draw=mean(oddX),mean\_away=mean(odd2),.N),by=list(MatchResult)]

matches\_with\_odds3[,prob\_home:=1/odd1]

matches\_with\_odds3[,prob\_draw:=1/oddX]

matches\_with\_odds3[,prob\_away:=1/odd2]

matches\_with\_odds3[,total\_prob:=prob\_home+prob\_draw+prob\_away]

matches\_with\_odds3[,home\_away\_diff:=prob\_home-prob\_away]

plot(matches\_with\_odds3[,list(home\_away\_diff,prob\_draw)])

matches\_with\_odds3[,diff\_bucket:=cut(home\_away\_diff,cut\_levels)]

result\_summary3=matches\_with\_odds3[,list(real\_draw\_ratio=sum(MatchResult=='draw', na.rm = T)/.N,draw\_prob\_bookmaker=mean(prob\_draw[MatchResult=='draw'], na.rm = T)),by=list(diff\_bucket)]

xiv. Code for Figure 15:

filtered\_odds4=odds[betType=='1x2' & bookmaker=='bet365']

filtered\_odds4[,c('betType','bookmaker','totalhandicap'):=NULL]

filtered\_odds4=filtered\_odds4[order(matchId, oddtype,date)]

latest\_odds4=filtered\_odds4[,list(final\_odd=odd[.N]),by=list(matchId,oddtype)]

latest\_odds4=dcast(latest\_odds4,matchId~oddtype,value.var='final\_odd')

temp4=matches[,list(matchId,date\_of\_match,home,away,MatchResult)]

matches\_with\_odds4=merge(temp4,latest\_odds4,by='matchId')

summary\_odds\_by\_result4=matches\_with\_odds4[,list(mean\_home=mean(odd1),

mean\_draw=mean(oddX),mean\_away=mean(odd2),.N),by=list(MatchResult)]

matches\_with\_odds4[,prob\_home:=1/odd1]

matches\_with\_odds4[,prob\_draw:=1/oddX]

matches\_with\_odds4[,prob\_away:=1/odd2]

matches\_with\_odds4[,total\_prob:=prob\_home+prob\_draw+prob\_away]

matches\_with\_odds4[,home\_away\_diff:=prob\_home-prob\_away]

plot(matches\_with\_odds4[,list(home\_away\_diff,prob\_draw)])

matches\_with\_odds4[,diff\_bucket:=cut(home\_away\_diff,cut\_levels)]

result\_summary4=matches\_with\_odds4[,list(real\_draw\_ratio=sum(MatchResult=='draw', na.rm = T)/.N,draw\_prob\_bookmaker=mean(prob\_draw[MatchResult=='draw'], na.rm = T)),by=list(diff\_bucket)]

xv. Code for Figure 16:

matches\_with\_odds3[,P\_home:=prob\_home/total\_prob]

matches\_with\_odds3[,P\_away:=prob\_away/total\_prob]

matches\_with\_odds3[,P\_draw:=prob\_draw/total\_prob]

matches\_with\_odds3[,P\_home\_away\_diff:=P\_home-P\_away]

P\_summary\_odds\_by\_result3=matches\_with\_odds3[,list(mean\_home=mean(P\_home),

mean\_draw=mean(P\_draw),mean\_away=mean(P\_away),.N),by=list(MatchResult)]

plot(matches\_with\_odds3[,list(P\_home\_away\_diff,P\_draw)])

matches\_with\_odds3[,P\_diff\_bucket:=cut(P\_home\_away\_diff,cut\_levels)]

P\_result\_summary3=matches\_with\_odds3[,list(real\_draw\_ratio=sum(MatchResult=='draw', na.rm = T)/.N,P\_draw\_prob\_bookmaker=mean(P\_draw[MatchResult=='draw'], na.rm = T)),by=list(diff\_bucket)]

xvi. Code for Figure 17:

matches\_with\_odds4[,P\_home:=prob\_home/total\_prob]

matches\_with\_odds4[,P\_away:=prob\_away/total\_prob]

matches\_with\_odds4[,P\_draw:=prob\_draw/total\_prob]

matches\_with\_odds4[,P\_home\_away\_diff:=P\_home-P\_away]

P\_summary\_odds\_by\_result4=matches\_with\_odds4[,list(mean\_home=mean(P\_home),

mean\_draw=mean(P\_draw),mean\_away=mean(P\_away),.N),by=list(MatchResult)]

plot(matches\_with\_odds4[,list(P\_home\_away\_diff,P\_draw)])

matches\_with\_odds4[,P\_diff\_bucket:=cut(P\_home\_away\_diff,cut\_levels)]

P\_result\_summary4=matches\_with\_odds4[,list(real\_draw\_ratio=sum(MatchResult=='draw', na.rm = T)/.N,P\_draw\_prob\_bookmaker=mean(P\_draw[MatchResult=='draw'], na.rm = T)),by=list(diff\_bucket)]

xvii. Code for Figure 18:

filtered\_odds2=odds[betType=='ou'& bookmaker=='bet365'& totalhandicap==2.5]

filtered\_odds2=filtered\_odds2[order(matchId,date)]

latest\_odds2=filtered\_odds2[,list(final\_odd=odd[.N]),by=list(matchId,oddtype)]

latest\_odds2=dcast(latest\_odds2,matchId~oddtype,value.var='final\_odd')

latest\_odds2[,prob\_over:=1/over]

latest\_odds2[,prob\_under:=1/under]

Total\_odds2=latest\_odds2$prob\_over+latest\_odds2$prob\_under

latest\_odds2[,Total\_odds2:=latest\_odds2$prob\_over+latest\_odds2$prob\_under]

latest\_odds2[,P\_over:=prob\_over/Total\_odds2]

latest\_odds2[,P\_under:=prob\_under/Total\_odds2]

cut\_levels=c(0:20)/20

latest\_odds2[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches[,TotalGoal:=HomeGoal+AwayGoal]

temp2=matches[,list(matchId,date,TotalGoal)]

latest\_odds2=merge(temp2,latest\_odds2,by='matchId')

result\_summary2=latest\_odds2[, list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N, bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

real\_over\_r2=result\_summary2$real\_over\_ratio

b2=result\_summary2$bookmaker\_over\_ratio

ggplot(result\_summary2,aes(x=b2, y=real\_over\_r2))+

geom\_point()+

geom\_abline(slope = 1, intercept = 0)+

ggtitle("Scatter Plot of Over Ratios for bet356")+

xlab("Bookmaker Over Ratio")+

ylab("Real Over Ratio")+

xlim(0,1)+

ylim(0,1)

xviii. Code for Figure 19:

filtered\_odds3=odds[betType=='ou'& bookmaker=='Betsafe'& totalhandicap==2.5]

filtered\_odds3=filtered\_odds3[order(matchId,date)]

latest\_odds3=filtered\_odds3[,list(final\_odd=odd[.N]),by=list(matchId,oddtype)]

latest\_odds3=dcast(latest\_odds3,matchId~oddtype,value.var='final\_odd')

latest\_odds3[,prob\_over:=1/over]

latest\_odds3[,prob\_under:=1/under]

Total\_odds3=latest\_odds3$prob\_over+latest\_odds3$prob\_under

latest\_odds3[,Total\_odds3:=latest\_odds3$prob\_over+latest\_odds3$prob\_under]

latest\_odds3[,P\_over:=prob\_over/Total\_odds3]

latest\_odds3[,P\_under:=prob\_under/Total\_odds3]

cut\_levels=c(0:20)/20

latest\_odds3[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches[,TotalGoal:=HomeGoal+AwayGoal]

temp3=matches[,list(matchId,date,TotalGoal)]

latest\_odds3=merge(temp3,latest\_odds3,by='matchId')

result\_summary3=latest\_odds3[,

list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

real\_over\_r3=result\_summary3$real\_over\_ratio

b3=result\_summary3$bookmaker\_over\_ratio

ggplot(result\_summary3,aes(x=b3, y=real\_over\_r3))+

geom\_point()+

geom\_abline(slope = 1, intercept = 0)+

ggtitle("Scatter Plot of Over Ratios for Betsafe")+

xlab("Bookmaker Over Ratio")+

ylab("Real Over Ratio")+

xlim(0,1)+

ylim(0,1)

xix. Code for Figure 20:

filtered\_odds4=odds[betType=='ou'& bookmaker=='BetVictor'& totalhandicap==2.5]

filtered\_odds4=filtered\_odds4[order(matchId,date)]

latest\_odds4=filtered\_odds4[,list(final\_odd=odd[.N]),by=list(matchId,oddtype)]

latest\_odds4=dcast(latest\_odds4,matchId~oddtype,value.var='final\_odd')

latest\_odds4[,prob\_over:=1/over]

latest\_odds4[,prob\_under:=1/under]

Total\_odds4=latest\_odds4$prob\_over+latest\_odds4$prob\_under

latest\_odds4[,Total\_odds4:=latest\_odds4$prob\_over+latest\_odds4$prob\_under]

latest\_odds4[,P\_over:=prob\_over/Total\_odds4]

latest\_odds4[,P\_under:=prob\_under/Total\_odds4]

cut\_levels=c(0:20)/20

latest\_odds4[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches[,TotalGoal:=HomeGoal+AwayGoal]

temp4=matches[,list(matchId,date,TotalGoal)]

latest\_odds4=merge(temp4,latest\_odds4,by='matchId')

result\_summary4=latest\_odds4[,

list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

real\_over\_r4=result\_summary4$real\_over\_ratio

b4=result\_summary4$bookmaker\_over\_ratio

ggplot(result\_summary4,aes(x=b4, y=real\_over\_r4))+

geom\_point()+

geom\_abline(slope = 1, intercept = 0)+

ggtitle("Scatter Plot of Over Ratios for BetVictor")+

xlab("Bookmaker Over Ratio")+

ylab("Real Over Ratio")+

xlim(0,1)+

ylim(0,1)

xx. Code for Figure 21:

filtered\_odds=odds[betType=='ou'& bookmaker=='Pinnacle'& totalhandicap==2.5]

filtered\_odds=filtered\_odds[order(matchId,date)]

latest\_odds=filtered\_odds[,list(final\_odd=odd[.N]),by=list(matchId,oddtype)]

latest\_odds=dcast(latest\_odds,matchId~oddtype,value.var='final\_odd')

latest\_odds[,prob\_over:=1/over]

latest\_odds[,prob\_under:=1/under]

Total\_odds=latest\_odds$prob\_over+latest\_odds$prob\_under

latest\_odds[,Total\_odds:=latest\_odds$prob\_over+latest\_odds$prob\_under]

latest\_odds[,P\_over:=prob\_over/Total\_odds]

latest\_odds[,P\_under:=prob\_under/Total\_odds]

cut\_levels=c(0:20)/20

latest\_odds[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches[,TotalGoal:=HomeGoal+AwayGoal]

temp=matches[,list(matchId,date,TotalGoal)]

latest\_odds=merge(temp,latest\_odds,by='matchId')

result\_summary=latest\_odds[,list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N, bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

real\_over\_r=result\_summary$real\_over\_ratio

b=result\_summary$bookmaker\_over\_ratio

ggplot(result\_summary,aes(x=b, y=real\_over\_r))+

geom\_point()+

geom\_abline(slope = 1, intercept = 0)+

ggtitle("Scatter Plot of Over Ratios for Pinnacle")+

xlab("Bookmaker Over Ratio")+

ylab("Real Over Ratio")+

xlim(0,1)+

ylim(0,1)

xxi. Code for Figure 22:

require(lubridate)

matches[,timestamp:=as\_datetime(date,tz='Turkey')]

matches[,date\_of\_match:=date(timestamp)]

latest\_odds[,date\_of\_match:=date(timestamp)]

latest\_odds[,timestamp:=as\_datetime(date,tz='Turkey')]

filtered\_odds[,timestamp:=as\_datetime(date,tz='Turkey')]

odds[,timestamp:=as\_datetime(date,tz='Turkey')]

temp=matches[,list(matchId,date\_of\_match)]

latest\_odds=merge(latest\_odds,temp,by='matchId')

matches\_of\_2011=latest\_odds[date\_of\_match.x>'2011-01-01' & date\_of\_match.x<'2012-01-01']

cut\_levels=c(25:30)/50

matches\_of\_2011[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches\_of\_2011=matches\_of\_2011[complete.cases(matches\_of\_2011)]

result\_summary\_2011=matches\_of\_2011[,

list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

matches\_of\_2012=latest\_odds[date\_of\_match.x>'2012-01-01' & date\_of\_match.x<'2013-01-01']

matches\_of\_2012[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches\_of\_2012=matches\_of\_2012[complete.cases(matches\_of\_2012)]

result\_summary\_2012=matches\_of\_2012[,

list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

matches\_of\_2013=latest\_odds[date\_of\_match.x>'2013-01-01' & date\_of\_match.x<'2014-01-01']

matches\_of\_2013[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches\_of\_2013=matches\_of\_2013[complete.cases(matches\_of\_2013)]

result\_summary\_2013=matches\_of\_2013[,list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

matches\_of\_2014=latest\_odds[date\_of\_match.x>'2014-01-01' & date\_of\_match.x<'2015-01-01']

matches\_of\_2014[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches\_of\_2014=matches\_of\_2014[complete.cases(matches\_of\_2014)]

result\_summary\_2014=matches\_of\_2014[,list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

matches\_of\_2015=latest\_odds[date\_of\_match.x>'2015-01-01' & date\_of\_match.x<'2016-01-01']

matches\_of\_2015[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches\_of\_2015=matches\_of\_2015[complete.cases(matches\_of\_2015)]

result\_summary\_2015=matches\_of\_2015[,list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

matches\_of\_2016=latest\_odds[date\_of\_match.x>'2016-01-01' & date\_of\_match.x<'2017-01-01']

matches\_of\_2016[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches\_of\_2016=matches\_of\_2016[complete.cases(matches\_of\_2016)]

result\_summary\_2016=matches\_of\_2016[,list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

matches\_of\_2017=latest\_odds[date\_of\_match.x>'2017-01-01' & date\_of\_match.x<'2018-01-01']

matches\_of\_2017[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches\_of\_2017=matches\_of\_2017[complete.cases(matches\_of\_2017)]

result\_summary\_2017=matches\_of\_2017[,list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

matches\_of\_2018=latest\_odds[date\_of\_match.x>'2018-01-01' & date\_of\_match.x<'2019-01-01']

matches\_of\_2018[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches\_of\_2018=matches\_of\_2018[complete.cases(matches\_of\_2018)]

result\_summary\_2018=matches\_of\_2018[,list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,

bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

matches\_of\_2019=latest\_odds[date\_of\_match.x>'2019-01-01' & date\_of\_match.x<'2020-01-01']

matches\_of\_2019[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches\_of\_2019=matches\_of\_2019[complete.cases(matches\_of\_2019)]

result\_summary\_2019=matches\_of\_2019[,list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N, bookmaker\_over\_ratio=mean(prob\_over[TotalGoal>=3],na.rm=TRUE)),by=list(diff\_bucket)]

order1<-order(result\_summary\_2011$diff\_bucket)

result\_summary\_2011=result\_summary\_2011[order1,]

order2<-order(result\_summary\_2012$diff\_bucket)

result\_summary\_2012=result\_summary\_2012[order2,]

order3<-order(result\_summary\_2013$diff\_bucket)

result\_summary\_2013=result\_summary\_2013[order3,]

order4<-order(result\_summary\_2014$diff\_bucket)

result\_summary\_2014=result\_summary\_2014[order4,]

order5<-order(result\_summary\_2015$diff\_bucket)

result\_summary\_2015=result\_summary\_2015[order5,]

order6<-order(result\_summary\_2016$diff\_bucket)

result\_summary\_2016=result\_summary\_2016[order6,]

order7<-order(result\_summary\_2017$diff\_bucket)

result\_summary\_2017=result\_summary\_2017[order7,]

order8<-order(result\_summary\_2018$diff\_bucket)

result\_summary\_2018=result\_summary\_2018[order8,]

order9<-order(result\_summary\_2019$diff\_bucket)

result\_summary\_2019=result\_summary\_2019[order9,]

plot(result\_summary\_2011$bookmaker\_over\_ratio,result\_summary\_2011$real\_over\_ratio,axes=T,col='dark red', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")

lines(result\_summary\_2011$bookmaker\_over\_ratio,result\_summary\_2011$real\_over\_ratio,col='dark red')

par(new=TRUE)

plot(result\_summary\_2012$bookmaker\_over\_ratio,result\_summary\_2012$real\_over\_ratio,axes=F,col='yellow', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")

lines(result\_summary\_2012$bookmaker\_over\_ratio,result\_summary\_2012$real\_over\_ratio,col='yellow')

par(new=TRUE)

plot(result\_summary\_2013$bookmaker\_over\_ratio,result\_summary\_2013$real\_over\_ratio,axes=F,col='dark blue',xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")

lines(result\_summary\_2013$bookmaker\_over\_ratio,result\_summary\_2013$real\_over\_ratio,col='dark blue')

par(new=TRUE)

plot(result\_summary\_2014$bookmaker\_over\_ratio,result\_summary\_2014$real\_over\_ratio,axes=F,col='black', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")

lines(result\_summary\_2014$bookmaker\_over\_ratio,result\_summary\_2014$real\_over\_ratio)

par(new=TRUE)

plot(result\_summary\_2015$bookmaker\_over\_ratio,result\_summary\_2015$real\_over\_ratio,axes=F,col='gray', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")

lines(result\_summary\_2015$bookmaker\_over\_ratio,result\_summary\_2015$real\_over\_ratio,col='gray')

par(new=TRUE)

plot(result\_summary\_2016$bookmaker\_over\_ratio,result\_summary\_2016$real\_over\_ratio,axes=F,col='green', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")

lines(result\_summary\_2016$bookmaker\_over\_ratio,result\_summary\_2016$real\_over\_ratio,col='green')

par(new=TRUE)

plot(result\_summary\_2017$bookmaker\_over\_ratio,result\_summary\_2017$real\_over\_ratio,axes=F,col='orange', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")

lines(result\_summary\_2017$bookmaker\_over\_ratio,result\_summary\_2017$real\_over\_ratio,col='orange')

par(new=TRUE)

plot(result\_summary\_2018$bookmaker\_over\_ratio,result\_summary\_2018$real\_over\_ratio,axis=F,col='brown', xlim = c(5:6)/10, ylim = c(0:1),xlab="Real", ylab = "Bookmaker")

lines(result\_summary\_2018$bookmaker\_over\_ratio,result\_summary\_2018$real\_over\_ratio,col='brown')

par(new=TRUE)

abline(h=0.55)

**R codes for Tables**

i. Code for Table 1:

summary\_by\_homegoal=matches[,list(count=.N),by=list(matchId,HomeGoal)]

factor(summary\_by\_homegoal$HomeGoal)

table\_for\_homegoal=table(summary\_by\_homegoal$HomeGoal)

table\_for\_homegoal

hist(summary\_by\_homegoal$HomeGoal,main = "HomeGoal Table", xlab = "Home Goals", ylab = "Number of Games", las =1, breaks = 30,col='light blue')

mean\_homegoal=mean(matches$HomeGoal,na.rm = T)

mean\_homegoal

par(new=TRUE)

plot(dpois(x=0:8,lambda=mean\_homegoal), xlab = "Home Goals",ylab="Number of Games",axes=F,col='dark red',pch=19)

HomeGoal\_pois=c(dpois(0,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(1,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(2,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(3,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(4,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(5,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(6,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(7,mean\_homegoal)\*sum(table\_for\_homegoal),

dpois(8,mean\_homegoal)\*sum(table\_for\_homegoal))

real\_vs\_poison\_homegoal=data.table(Real\_HomeGoal=table\_for\_homegoal,Poison\_HomeGoal=HomeGoal\_pois)

ii. Code for Table 2:

summary\_by\_awaygoal=matches[,list(count=.N),by=list(matchId,AwayGoal)]factor(summary\_by\_awaygoal$AwayGoal)

table\_for\_awaygoal=table(summary\_by\_awaygoal$AwayGoal)

table\_for\_awaygoal

hist(summary\_by\_awaygoal$AwayGoal,main = "AwayGoal Table", xlab = "Away Goals", ylab = "Number of Games",las=1, breaks = 30,col='light blue')

mean\_awaygoal=mean(matches$AwayGoal,na.rm = T)

par(new=TRUE)

plot(dpois(x=0:7,lambda=mean\_awaygoal), xlab = "Away Goals",ylab="Number of Games",axes=F,col='dark red',pch=19)

AwayGoal\_pois=c(dpois(0,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(1,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(2,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(3,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(4,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(5,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(6,mean\_awaygoal)\*sum(table\_for\_awaygoal),

dpois(7,mean\_awaygoal)\*sum(table\_for\_awaygoal))

real\_vs\_poison\_awaygoal=data.table(Real\_AwayGoal=table\_for\_awaygoal,Poison\_AwayGoal=AwayGoal\_pois)

iii. Code for Table 3,4,5,6: By proceeding Pinnacle’s, Betsafe’s, bet365’s, Bet10’s Normalized P(home win)-P(away win) vs Normalized P(tie) codes in figure 12,13,16,17

names(P\_result\_summary)[1]<-("Difference Bucket")

names(P\_result\_summary)[2]<-("Real Draw Ratio")

names(P\_result\_summary)[3]<-("Normalized Bookmaker Draw Ratio")

grid.table(P\_result\_summary)

names(P\_result\_summary2)[1]<-("Difference Bucket")

names(P\_result\_summary2)[2]<-("Real Draw Ratio")

names(P\_result\_summary2)[3]<-("Normalized Bookmaker Draw Ratio")

grid.table(P\_result\_summary2)

names(P\_result\_summary3)[1]<-("Difference Bucket")

names(P\_result\_summary3)[2]<-("Real Draw Ratio")

names(P\_result\_summary3)[3]<-("Normalized Bookmaker Draw Ratio")

grid.table(P\_result\_summary3)

names(P\_result\_summary4)[1]<-("Difference Bucket")

names(P\_result\_summary4)[2]<-("Real Draw Ratio")

names(P\_result\_summary4)[3]<-("Normalized Bookmaker Draw Ratio")

grid.table(P\_result\_summary4)

iv. Code for Table 7:

filtered\_odds=odds[betType=='ou'& bookmaker=='Pinnacle'& totalhandicap==2.5]

filtered\_odds=filtered\_odds[order(matchId,date)]

latest\_odds=filtered\_odds[,list(final\_odd=odd[.N]),by=list(matchId,oddtype)]

latest\_odds=dcast(latest\_odds,matchId~oddtype,value.var='final\_odd')

latest\_odds[,prob\_over:=1/over]

latest\_odds[,prob\_under:=1/under]

Total\_odds=latest\_odds$prob\_over+latest\_odds$prob\_under

latest\_odds[,Total\_odds:=latest\_odds$prob\_over+latest\_odds$prob\_under]

latest\_odds[,P\_over:=prob\_over/Total\_odds]

latest\_odds[,P\_under:=prob\_under/Total\_odds]

cut\_levels=c(0:20)/20

latest\_odds[,diff\_bucket:=cut(prob\_over,cut\_levels)]

matches[,TotalGoal:=HomeGoal+AwayGoal]

temp=matches[,list(matchId,date,TotalGoal)]

latest\_odds=merge(temp,latest\_odds,by='matchId')

result\_summary=latest\_odds[,

list(real\_over\_ratio=sum(TotalGoal>=3,na.rm = TRUE)/.N,