NHL Days off comparison

Malcolm Hess

March 16, 2015

Contributed by Malcolm Hess

This project is based on the win rate for NHL teams based on days between games.

Links

The code and files for this project can be found here: https://github.com/Mal-Hess/Hockey

If you enjoyed this or would like to see more of my work feel free to check out my github account for more projects.
https://github.com/Mal-Hess/

Background

```
In the NHL not all schedules are created equal.  Some teams will have more back-to-back games than others.
A team with more back-to-back games is thought to have the harder schedule.  The expectation is that having
to play two games in a row puts added strain on a team and would therefore reduce the likelihood of their v
ictory.  On top of the physical strain a team won't have the opportunity to prepare for a specific opponent
.
```

Goal

The goal of this project is to analyze if the win rate for back-to-back games is significantly different than the win rate of

games with a longer period of rest.

Process

I began by acquiring at least one full season's results. I web-scraped it off of hockey-refernce.com. There are two tables on the site, one for the regular season results and one for the playoffs. My analysis focused on the regular season games.

```
## Loading required package: bitops
##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:lubridate':
##
##     intersect, setdiff, union
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

I needed to manipulate the table to get the data into a usuable format for my analysis. The first steps I took towards cleaning the data was to transform the variables into the correct class types. There are a few games that were postponed which were were removed. I created a win and loss column for both the visitor and home team side. I then split the data into a list based on the away team and the home team. By seperating the two, when I rbind the two groups together there would be two rows for each game, one for the home team and one for the away team.

```r
#Row rename and variable transformation
names(season)[5:6] <- c("G.1", "OT")
season <- transform(season, Notes = as.character(Notes), Date = as.character(Date), G= as.character(G), G.1
= as.character(G.1))
season <- transform(season, Date = as.Date(Date))


#Get's rid of postponed games
season <- subset(season, G!= "")
rows <- nrow(season)


#puts in W/L factor for both Visitor and Home.
i<-1
while (i <= rows){

  if(season$G[i]<season$G.1[i]){
  season$V.W[i] <- "L"
  season$H.W[i] <- "W"
  } else {
  season$V.W[i] <- "W"
  season$H.W[i] <- "L"}

  i<-i+1
}


#Creates a dataframe with all games based on visiting team
visitorgames <- season[,c(1,2,3,6,8,7)]
visitorgames$Home_Away <- "A"
names(visitorgames)[2] <- "Team"
```

```
names(visitorgames)[5] <- "W"


#Creates a dataframe with all games based on home team
homegames <- season[,c(1,4,5,6,9,7)]
homegames$Home_Away <- "H"
names(homegames)[2] <- "Team"
names(homegames)[3] <- "G"
names(homegames)[5] <- "W"


#Combines visiter based games and Home based games into one dataframe
allgames <- rbind(homegames, visitorgames)
```

After arranging the data into team and date order I calculated the days off between games. I used a do call, by each team, to find the difference between one game and the next. After unlisting it, I cbined it onto my database with the rest of the data.

Lastly I subsetted the data to create a table which contained the count of wins and losses for each amount of days off.

Teams that played a game back-to-back had a 48% win rate compared to a 52% win rate for teams playing with 2 days off between games.

```
suball <- data.frame(cbind(allgames$W, allgames$daysoffvector))


table<- data.frame(table(suball))
backtoback <- table$Freq[2]/(table$Freq[2] +table$Freq[1])
onedayoff <- table$Freq[10]/(table$Freq[10] +table$Freq[9])
twodaysoff <- table$Freq[16]/(table$Freq[16] +table$Freq[15])


backtoback
```

```
## [1] 0.4807256
```

```
onedayoff
```

```
## [1] 0.499256
```

```
twodaysoff
```

```
## [1] 0.5229759
```

Conclusion

I took out the values for the back-to-back and 2 days off games and tabled them to perform a simple Chi-squared test between the two groups. With a p-value of .23 we do not reject the null-hypothesis. There is no significant difference between the win rates of teams that are playing a back-to-back game and those that have had 2 days of rest between games.

```
data <- matrix(c(table[1,3],table[2,3], table[15,3],table[16,3]), nrow=2)
table2 <- as.table(data)


chisq.test(table2)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
```

```
## 
## data:  table2
## X-squared = 1.438, df = 1, p-value = 0.2305
```

Future Work

I will go on to compare the individual chance of a team winning dependant on if for that specific game the one team has had more rest. This extra step should clear up any noise created when teams with the same amount of days off play against each other.