DA 460 - Assignment 2

Marjorie Blanco

04 July, 2018

# Part 1

Based on Handout 2 R, apply R to compare Kobe Bryant to the simulated independent shooter. Using calc\_streak, to compute the streak lengths of sim\_basket, and then answer the following questions. Make sure you include clear headings (e.g., Handout 2 R or Handout 2 SAS). For each part of the question, make sure you include the command line/code, then paste relevant output/results, and also comment on the output/results as needed (to answer the questions).

## Problem 1

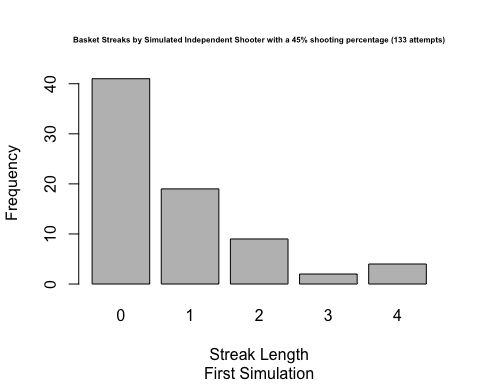
Describe the distribution of streak lengths. What is the typical streak length for this simulated independent shooter with a 45% shooting percentage? How long is the player’s longest streak of baskets in 133 shots?

# possible outcome H: hit M: missed  
outcomes <- c("H", "M")  
# number of simulated shots  
sampleSize = 133  
# shooting percentage  
shooting = 0.45

# run simulation  
sim\_basket <- sample(outcomes, size = sampleSize, replace = TRUE, prob = c(shooting, 1 - shooting))  
# view the results of this simulation  
table(sim\_basket)

## sim\_basket  
## H M   
## 59 74

# calculate the lengths of all shooting streaks  
sim\_streak <- calc\_streak(sim\_basket)  
# streak distribution  
barplot(table(sim\_streak), main = "Basket Streaks by Simulated Independent Shooter with a 45% shooting percentage (133 attempts)",  
 sub = "First Simulation", xlab = "Streak Length", ylab = "Frequency", cex.main=0.5)



The distribution of simulated independent shooter with a 45% shooting percentage is strongly skewed to the right. The typical streak length is 0 and the longest streak of baskets is 4.

## Problem 2

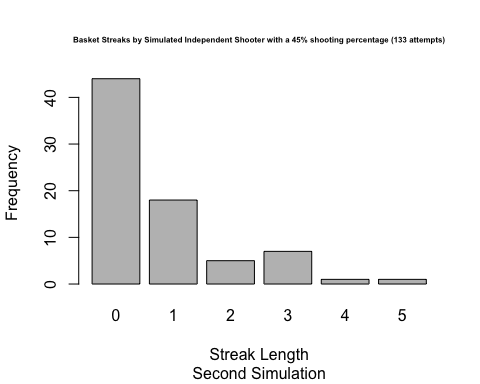
If you were to run the simulation of the independent shooter a second time, how would you expect its streak distribution to compare to the distribution from the question above? Exactly the same? Somewhat similar? Totally different? Explain your reasoning.

The expectation is for the streak distribution for the second simulation to be the somewhat similar to the distribution from the first simulation. The simulation generates random outcome that adhere to the specified probabilities; therefore the probabilities will be the same, but not necessarily the outcome will be the same.

# run simulation  
sim\_basket <- sample(outcomes, size = sampleSize, replace = TRUE, prob = c(shooting, 1 - shooting))  
# view the results of this simulation  
table(sim\_basket)

## sim\_basket  
## H M   
## 58 75

# calculate the lengths of all shooting streaks  
sim\_streak <- calc\_streak(sim\_basket)  
# streak distribution  
barplot(table(sim\_streak), main = "Basket Streaks by Simulated Independent Shooter with a 45% shooting percentage (133 attempts)",  
 sub = "Second Simulation", xlab = "Streak Length", ylab = "Frequency", cex.main=0.5)

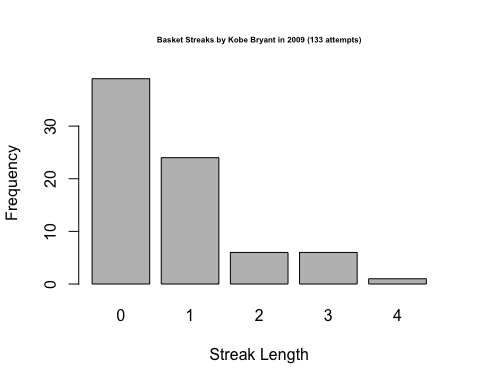


The distribution of simulated independent shooter with a 45% shooting percentage is strongly skewed to the right. The typical streak length is 0 and the longest streak of baskets is 5.

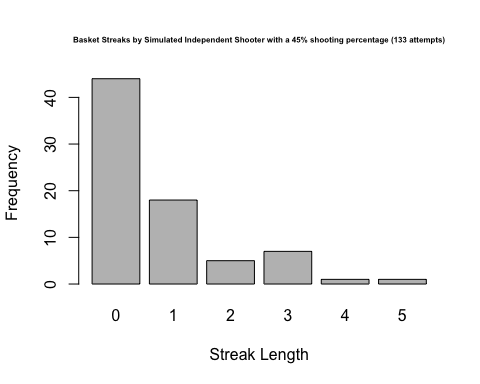
## Problem 3

How does Kobe Bryant’s distribution of streak lengths compare to the distribution of streak lengths for the simulated shooter? Using this comparison, do you have evidence that the hot hand model fits Kobe’s shooting patterns? Explain.

barplot(table(kobe\_streak), main = "Basket Streaks by Kobe Bryant in 2009 (133 attempts)",  
 xlab = "Streak Length", ylab = "Frequency", cex.main=0.5)



barplot(table(sim\_streak), main = "Basket Streaks by Simulated Independent Shooter with a 45% shooting percentage (133 attempts)",  
 xlab = "Streak Length", ylab = "Frequency", cex.main=0.5)



If Kobe’s shooting streaks diverge significantly from an independent shooter’s streaks, we can conclude that Kobe likely has a hot hand. Kobe Bryant’s distribution of streak lengths is very similar to the distribution of streak lengths for the simulated shooter. The distributions do share similar characteristics such as skewness.

Since Kobe’s streak length distribution looks very similar to the independent shooter’s simulated steak length distribution, the conclusion is that Kobe Bryant does not have a “hot hand”.

# Part 2

In [1]:

filename kobe url 'http://www.openintro.org/stat/data/kobe.csv';

proc import datafile=kobe

out=kobe(drop=time)

dbms=csv

replace;

getnames=yes;

guessingrows=max;

run;

filename cstrk url "http://www.openintro.org/stat/data/calc\_streak.sas";

%include cstrk;

SAS Connection established. Subprocess id is 25065

Out[1]:

34 ods listing close;ods html5 (id=saspy\_internal) file=stdout options(bitmap\_mode='inline') device=svg; ods graphics on /  
34 ! outputfmt=png;  
*NOTE: Writing HTML5(SASPY\_INTERNAL) Body file: STDOUT*  
35   
36 filename kobe url 'http://www.openintro.org/stat/data/kobe.csv';  
37   
38 proc import datafile=kobe  
39 out=kobe(drop=time)  
40 dbms=csv  
41 replace;  
42 getnames=yes;  
43 guessingrows=max;  
44 run;  
*NOTE: Unable to open SASUSER.PROFILE. WORK.PROFILE will be opened instead.*  
*NOTE: All profile changes will be lost at the end of the session.*  
45 /\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
46 \* PRODUCT: SAS  
47 \* VERSION: 9.4  
48 \* CREATOR: External File Interface  
49 \* DATE: 05JUL18  
50 \* DESC: Generated SAS Datastep Code  
51 \* TEMPLATE SOURCE: (None Specified.)  
52 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*/  
53 data WORK.KOBE (drop=time) ;  
54 %let \_EFIERR\_ = 0; /\* set the ERROR detection macro variable \*/  
55 infile KOBE delimiter = ',' MISSOVER DSD lrecl=32767 firstobs=2 ;  
56 informat vs $3. ;  
57 informat game best32. ;  
58 informat quarter $3. ;  
59 informat time $7. ;  
60 informat description $67. ;  
61 informat basket $1. ;  
62 format vs $3. ;  
63 format game best12. ;  
64 format quarter $3. ;  
65 format time $7. ;  
66 format description $67. ;  
67 format basket $1. ;  
68 input  
69 vs $  
70 game  
71 quarter $  
72 time $  
73 description $  
74 basket $  
75 ;  
76 if \_ERROR\_ then call symputx('\_EFIERR\_',1); /\* set ERROR detection macro variable \*/  
77 run;  
*NOTE: The infile KOBE is:*  
 *Filename=http://www.openintro.org/stat/data/kobe.csv,*  
 *Local Host Name=localhost.localdomain,*  
 *Local Host IP addr=::1,*  
 *Service Hostname Name=www.openintro.org,*  
 *Service IP addr=192.185.65.127,*  
 *Service Name=httpd,Service Portno=80,*  
 *Lrecl=32767,Recfm=Variable*  
  
*NOTE: 133 records were read from the infile KOBE.*  
 *The minimum record length was 39.*  
 *The maximum record length was 82.*  
*NOTE: The data set WORK.KOBE has 133 observations and 5 variables.*  
*NOTE: DATA statement used (Total process time):*  
 *real time 0.26 seconds*  
 *cpu time 0.01 seconds*  
  
133 rows created in WORK.KOBE from KOBE.  
   
   
   
*NOTE: WORK.KOBE data set was successfully created.*  
*NOTE: The data set WORK.KOBE has 133 observations and 5 variables.*  
*NOTE: PROCEDURE IMPORT used (Total process time):*  
 *real time 1.49 seconds*  
 *cpu time 0.10 seconds*  
  
78 filename cstrk url "http://www.openintro.org/stat/data/calc\_streak.sas";  
79 %include cstrk;  
110   
111 ods html5 (id=saspy\_internal) close;ods listing;  
  
112

## Problem 1

Describe the distribution of streak lengths. What is the typical streak length for this simulated independent shooter with a 45% shooting percentage? How long is the player’s longest streak of baskets in 133 shots?

In [2]:

data sim\_basket(keep=basket);

do i = 1 to 133;

flip = rand("Bernoulli",0.45);

if flip = 1 then basket="H"; else basket="M";

output;

end;

run;

Out[2]:

114 ods listing close;ods html5 (id=saspy\_internal) file=stdout options(bitmap\_mode='inline') device=svg; ods graphics on /  
114! outputfmt=png;  
*NOTE: Writing HTML5(SASPY\_INTERNAL) Body file: STDOUT*  
115   
116 data sim\_basket(keep=basket);  
117 do i = 1 to 133;  
118 flip = rand("Bernoulli",0.45);  
119 if flip = 1 then basket="H"; else basket="M";  
120 output;  
121 end;  
122 run;  
*NOTE: The data set WORK.SIM\_BASKET has 133 observations and 1 variables.*  
*NOTE: DATA statement used (Total process time):*  
 *real time 0.00 seconds*  
 *cpu time 0.00 seconds*  
  
123   
124 ods html5 (id=saspy\_internal) close;ods listing;  
  
125

In [3]:

%calc\_streak(dset=sim\_basket, streakvar=basket, outset=sim\_streak);

proc sgplot data=sim\_streak;

title "Basket Streaks by Simulated Independent Shooter with a 45% shooting percentage (133 attempts)";

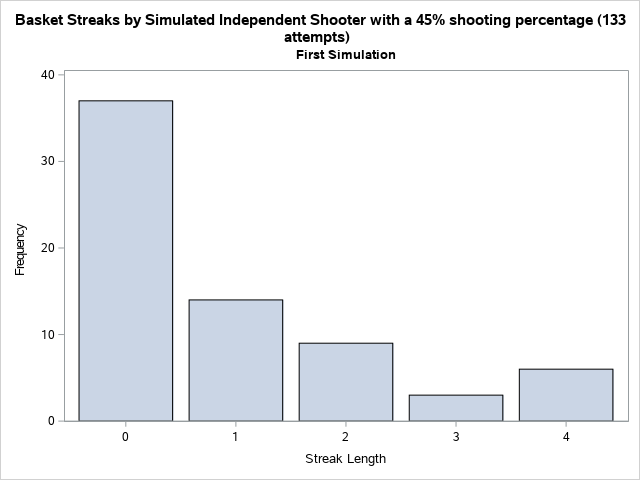
title2 "First Simulation";

xaxis label='Streak Length';

vbar streak;

run;

Out[3]:



Ans: The distribution of simulated independent shooter with a 45% shooting percentage is strongly skewed to the right. The typical streak length is 0 and the longest streak of baskets is 5.

## Problem 2

If you were to run the simulation of the independent shooter a second time, how would you expect its streak distribution to compare to the distribution from the question above? Exactly the same? Somewhat similar? Totally different? Explain your reasoning.

In [4]:

data sim\_basket(keep=basket);

do i = 1 to 133;

flip = rand("Bernoulli",0.45);

if flip = 1 then basket="H"; else basket="M";

output;

end;

run;

Out[4]:

140 ods listing close;ods html5 (id=saspy\_internal) file=stdout options(bitmap\_mode='inline') device=svg; ods graphics on /  
140! outputfmt=png;  
*NOTE: Writing HTML5(SASPY\_INTERNAL) Body file: STDOUT*  
141   
142 data sim\_basket(keep=basket);  
143 do i = 1 to 133;  
144 flip = rand("Bernoulli",0.45);  
145 if flip = 1 then basket="H"; else basket="M";  
146 output;  
147 end;  
148 run;  
*NOTE: The data set WORK.SIM\_BASKET has 133 observations and 1 variables.*  
*NOTE: DATA statement used (Total process time):*  
 *real time 0.00 seconds*  
 *cpu time 0.00 seconds*  
  
149   
150 ods html5 (id=saspy\_internal) close;ods listing;  
  
151

In [5]:

%calc\_streak(dset=sim\_basket, streakvar=basket, outset=sim\_streak);

proc sgplot data=sim\_streak;

title "Basket Streaks by Simulated Independent Shooter with a 45% shooting percentage (133 attempts)";

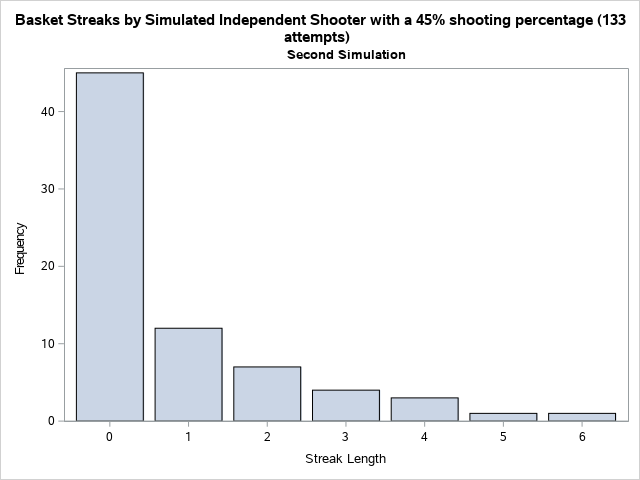
title2 "Second Simulation";

xaxis label='Streak Length';

vbar streak;

run;

Out[5]:



Ans: The expectation is for the streak distribution for the second simulation to be the somewhat similar to the distribution from the first simulation. The simulation generates random outcome that adhere to the specified probabilities; therefore, the probabilities will be the same, but not necessarily the outcome will be the same.

## Problem 3

How does Kobe Bryant’s distribution of streak lengths compare to the distribution of streak lengths for the simulated shooter? Using this comparison, do you have evidence that the hot hand model fits Kobe’s shooting patterns? Explain.

Ans: If Kobe's shooting streaks diverge significantly from an independent shooter's streaks, we can conclude that Kobe likely has a hot hand. Kobe Bryant's distribution of streak lengths is very similar to the distribution of streak lengths for the simulated shooter. The distributions do share similar characteristics such as skewness.

Since Kobe's streak length distribution looks very similar to the independent shooter's simulated steak length distribution, the conclusion is that Kobe Bryant does not have a "hot hand".