HW 14 Solutions

2024-04-06

Setup Code

```
# 1. define the values
k <- 4
p_A <- 0.55

# 2. Sequence of Length
s <- rbinom(2*k-1,1,p_A)

# 3. Determine if A or B won the series
# 1s will be team A winning

o <- ifelse(sum(s) >= 4, 'A', 'B')

# 4. Number of Games Played

# loop through the possible lengths of games
for(i in k:(2*k - 1)){
  # A is the number of games A won
  A = sum(s[1:i])
  B = i - A

  # check if A or B won 4 games yet
  if((A == 4) | (B == 4)){
    # number of games is i
    n = i

    # leave the for loop since we know the number of games now
    break
  }
}

# 5. Put into a Function called world series

world_series <- function(k, p_A){

  # create the sequence of games
  s <- rbinom(2*k-1,1,p_A)

  # get the winner of the series
  o <- ifelse(sum(s) >= 4, 'A', 'B')
```

```
  # get the number of games
  for(i in k:(2*k - 1)){
    # A is the number of games A won
    A = sum(s[1:i])
    B = i - A

    # check if A or B won 4 games yet
    if((A == 4) | (B == 4)){
        # number of games is i
        n = i

        # leave the for loop since we know the number of games now
        break
      }
  }

  # package as a vector
  return(c(o, n))

}

# 6. Create a function with inputs R, k, p_A, output an R x 2 with o in Col 1 and n in Col 2

replicate_world_series <- function(R, k, p_A){
  replicate(R, world_series(k, p_A)) |>
    t() |>
    matrix(ncol = 2, nrow = R)
}
```

1. Generate 5000 draws from the world series distribution (Use any k and p_A).

```
R <- 5000
k <- 4
p_A <- 0.55

# set the seed for reproducibility
set.seed(20302)
output <- replicate_world_series(R, k, p_A)

# make output a Data Frame
output <- output |>
  data.frame()

# df column names
colnames(output) <- c('Winner', 'NumGames')

# print first 5 rows of the output
head(output, 5)
```

```
##   Winner NumGames
```

```
## 1      A        6
## 2      B        5
## 3      A        6
## 4      B        7
## 5      A        7
```

2. With the draws, create the contingency table with number of games as the row variable and winner as the column variable. Include conditional probabilities.

```
# load gmodels for cross table command
library(gmodels)

# create the cross table
gmodels::CrossTable(output$NumGames,
                    output$Winner,
                    prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  5000
##
##
##                  | output$Winner
## output$NumGames |         A |         B | Row Total |
## ----------------|-----------|-----------|-----------|
##               4 |       458 |       223 |       681 |
##                 |     0.673 |     0.327 |     0.136 |
##                 |     0.150 |     0.115 |           |
##                 |     0.092 |     0.045 |           |
## ----------------|-----------|-----------|-----------|
##               5 |       841 |       450 |      1291 |
##                 |     0.651 |     0.349 |     0.258 |
##                 |     0.275 |     0.232 |           |
##                 |     0.168 |     0.090 |           |
## ----------------|-----------|-----------|-----------|
##               6 |       944 |       602 |      1546 |
##                 |     0.611 |     0.389 |     0.309 |
##                 |     0.309 |     0.310 |           |
##                 |     0.189 |     0.120 |           |
## ----------------|-----------|-----------|-----------|
```

```
##               7 |       814 |       668 |      1482 |
##                 |     0.549 |     0.451 |     0.296 |
##                 |     0.266 |     0.344 |           |
##                 |     0.163 |     0.134 |           |
## ----------------|-----------|-----------|-----------|
##    Column Total |      3057 |      1943 |      5000 |
##                 |     0.611 |     0.389 |           |
## ----------------|-----------|-----------|-----------|
##
##
```

3. Supposing that the world series ends in $2k - 1$ games, what is the probability that team A will win?

We can get this value from the table we produced. The probability we want is $P(\text{A}|\text{7 Games})$. From the cross table, this is the row conditional probability for 7 Games and A wins which is 0.549. This matches our value for p_A.