Homework 2

Kwasi Mensah

2022-10-21

Author: Kwasi Mensah Homwork_Number: 1 Due Date: “2022-10-21” Output: pdf_document Attribution statement:

I did the homework by myself, with help from the book and the professor #R Markdown #Run these three functions to get a clean test of homework code

```
dev.off() #Clear the graph window

## null device
##           1

cat('\014') #Clear the console

rm(list = ls()) #Clear user objects from the environment
```

#R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see http://rmarkdown.rstudio.com.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:
###

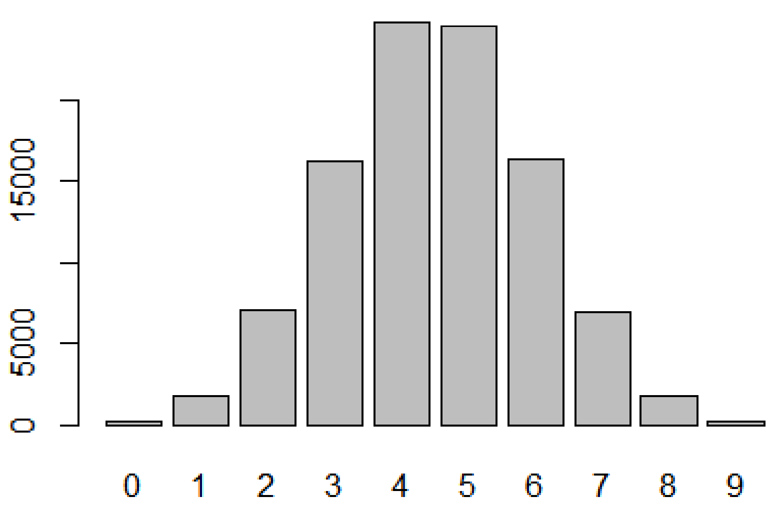
1. Flip a fair coin nine times and write down the number of heads obtained. Now repeat this process 100,000 times. Obviously, you don’t want to have to do that by hand, so create the necessary lines of R code to do it for you. Hint: You will need both the rbinom() function and the table() function. Write down the results and explain in your own words what they mean.

```
set.seed(321)
cointoss <-rbinom(n=100000, size=9, prob=.5)
# Trial
table(cointoss)

## cointoss
##     0     1     2     3     4     5     6     7     8     9
##   199  1789  7039 16270 24806 24630 16349  7007  1725   186
```

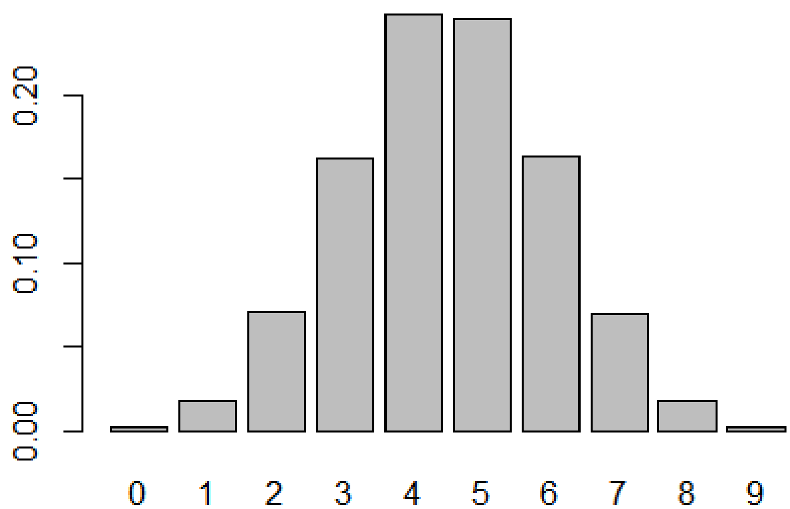
```
#Probability
table(cointoss)/length(cointoss)

## cointoss
##       0       1       2       3       4       5       6       7       8
9
## 0.00199 0.01789 0.07039 0.16270 0.24806 0.24630 0.16349 0.07007 0.01725
0.00186

# We do a trial of filliping a coin 10000 times, and we see that we get
results with  4 heads 24.8%  during the trial and 5 heads 24.6% during the
trial. If we round up we see a relation that we get a result of 4 or 5 heads
with 25% We can also see that we get all a result of all head .00186 of the
time and all tails .00199, but these result are not likely with a fair coin.
```

2. Using the output from Exercise 1, summarize the results of your 100,000 trials of nine flips each in a bar plot using the appropriate commands in R. Convert the results to probabilities and represent that in a bar plot as well. Write a brief interpretive analysis that describes what each of these barplots signifies and how the two bar plots are related. Make sure to comment on the shape of each bar plot and why you believe that the bar plot has taken that shape. Also make sure to say something about the center of the bar plot and why it is where it is.

```
#Trail Bar Plot
plot1<- barplot(table(cointoss))
```



```
#Probability Bar Plot
plot2<- barplot(table(cointoss)/length(cointoss))
```




```
#The bar plots are related in that the have the same shape in having a normal
distribution which results in a normal bell curve shape.
#The probability used is .5 or 50% because a coin only has 2 sides. The
probability of getting heads 4 or 5 times were very close to each other and
when rounded to the nearest percent they both result in 25%. Both plots
result in a normal distribution or bell curve shape, which aligns with the
law of large numbers and the central limit theorem.
```

6. One hundred students took a statistics test. Fifty of them are high school students and 50 are college students. Eighty students passed and 20 students failed. You now have enough information to create a two-by-two contingency table with all of the marginal totals specified (although the four main cells of the table are still blank). Draw that table and write in the marginal totals. I'm now going to give you one additional piece of information that will fill in one of the four blank cells: only three college students failed the test. With that additional information in place, you should now be able to fill in the remaining cells of the two-by-two table. Comment on why that one additional piece of information was all you needed in order to figure out all four of the table's main cells. Finally, create a second copy of the complete table, replacing the counts of students with probabilities. What is the pass rate for high school students? In other words, if one focuses only on high school students, what is the probability that a student will pass the test?

```r
# The additional information of 3 College students failing you could do the
simple math needed to find out the High School and College Students who
passed and failed.

students <- matrix(c(33,47,17,3),ncol=2,byrow=T)
colnames(students)<-c("High School","College")
rownames(students)<-c("Passed", "Failed")
students<- as.table(students)
students

##        High School College
## Passed          33      47
## Failed          17       3

#Grand Total
margin.table(students)

## [1] 100

#Marginal Total for Rows(Passed, Failed)
margin.table(students,1)

## Passed Failed
##     80     20

#Marginal Total for Columns(High School, College)
margin.table(students,2)

## High School     College
##          50          50

#Probability Table
testprob<-((students)/(margin.table(students)))
testprob

##        High School College
## Passed        0.33    0.47
## Failed        0.17    0.03

# Proportion Table
students[1,1]/sum(students[,1])

## [1] 0.66

prop.table(students,2)

##        High School College
## Passed        0.66    0.94
## Failed        0.34    0.06

#66% probability of the High School students passing the test.
```

7. In a typical year, 71 out of 100,000 homes in the United Kingdom is repossessed by the bank because of mortgage default (the owners did not pay their mortgage for many months). Barclays Bank has developed a screening test that they want to use to predict whether a mortgagee will default. The bank spends a year collecting test data: 93,935 households pass the test and 6,065 households fail the test. Interestingly, 5,996 of those who failed the test were actually households that were doing fine on their mortgage (i.e., they were not defaulting and did not get repossessed). Construct a complete contingency table from this information. Hint: The 5,996 is the only number that goes in a cell; the other numbers are marginal totals. What percentage of customers both pass the test and do not have their homes repossessed?

```
customers<-matrix(c(2,69,93933,5996),ncol=2,byrow=T)
colnames(customers)<-c("Passed","Failed")
rownames(customers)<-c("Default", "No Default")
customers<-as.table(customers)
prob1<-customers/margin.table(customers)
#Contingency Table
prob1

##              Passed  Failed
## Default    0.00002 0.00069
## No Default 0.93933 0.05996

Pass<-prob1[2,1]
#Passed & No Default
Pass

## [1] 0.93933

#93.933% of customers both pass the test and do not have their homes
repossessed.
```

8. Imagine that Barclays deploys the screening test from Exercise 6 on a new customer and the new customer fails the test. What is the probability that this customer will actually default on his or her mortgage? Show your work and especially show the tables that you set up to help with your reasoning.

```
Failed<-customers[1,2]
#Failed and Defaulted
Failed

## [1] 69

TotalFailed<-margin.table(customers,2)[2]
#Total Failed
TotalFailed

## Failed
##   6065
```

```r
#Failed Probability
100*Failed/TotalFailed

##   Failed
## 1.137675

#There is a probability of 1.137675 that new customer with failed tests would default on their mortgage.
```