R Competency Check #4

Scott Dolan

9/15/2020

PMF of a 5 Question True/False Test

A test consists of five true-false problems. A reasonable model for the answer given by a student who has not studied assumes that each question is marked T or F by flipping a coin. Thus, any 5-long binary sequence of 0's (incorrect) and 1's (correct), that is, points received in each of the five questions, is equally likely. Let X denote the test score of such a student. Use R to do the following:

1. Create an expanded grid of all possible outcomes (0's and 1's).

```
allOutcomes <- expand.grid(X1=0:1, X2=0:1, X3=0:1, X4=0:1, X5=0:1)
allOutcomes
```

```
##    X1 X2 X3 X4 X5
## 1   0  0  0  0  0
## 2   1  0  0  0  0
## 3   0  1  0  0  0
## 4   1  1  0  0  0
## 5   0  0  1  0  0
## 6   1  0  1  0  0
## 7   0  1  1  0  0
## 8   1  1  1  0  0
## 9   0  0  0  1  0
## 10  1  0  0  1  0
## 11  0  1  0  1  0
## 12  1  1  0  1  0
## 13  0  0  1  1  0
## 14  1  0  1  1  0
## 15  0  1  1  1  0
## 16  1  1  1  1  0
## 17  0  0  0  0  1
## 18  1  0  0  0  1
## 19  0  1  0  0  1
## 20  1  1  0  0  1
## 21  0  0  1  0  1
## 22  1  0  1  0  1
## 23  0  1  1  0  1
## 24  1  1  1  0  1
## 25  0  0  0  1  1
## 26  1  0  0  1  1
## 27  0  1  0  1  1
## 28  1  1  0  1  1
## 29  0  0  1  1  1
```

```
## 30  1  0  1  1  1
## 31  0  1  1  1  1
## 32  1  1  1  1  1
```

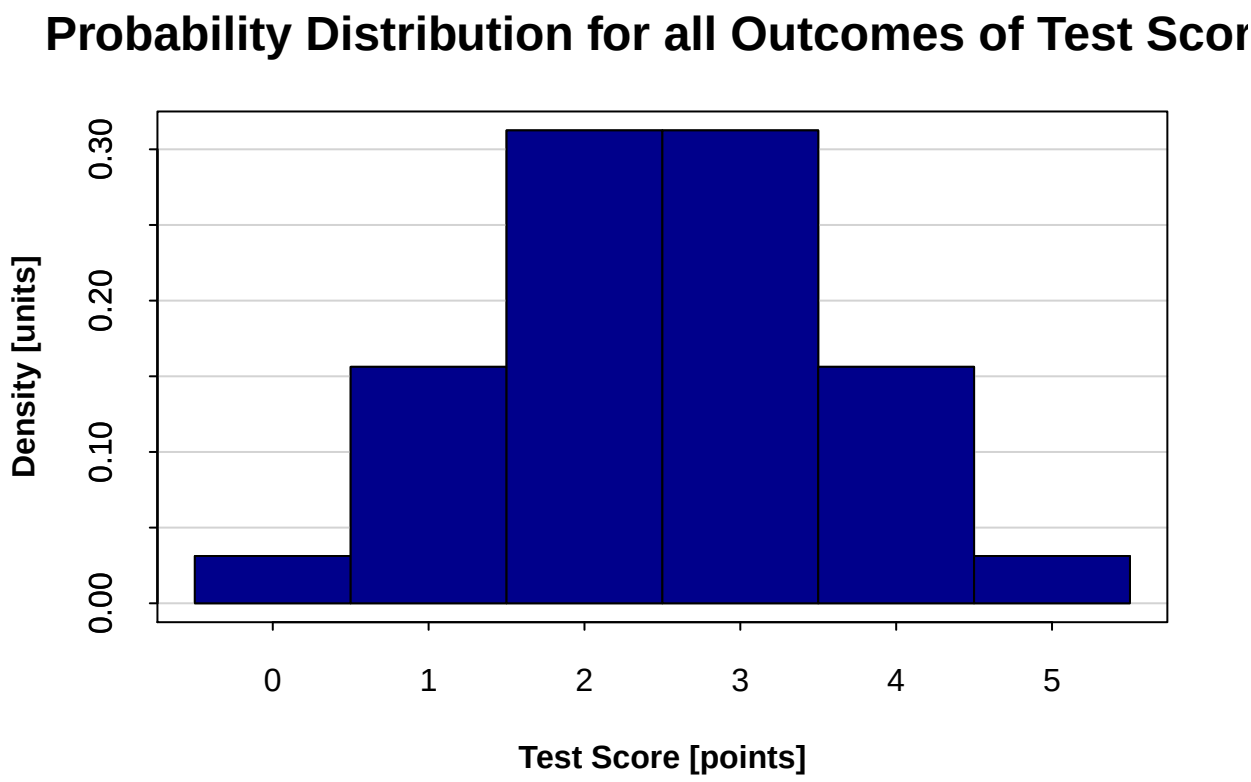
2. Create a PMF (probability mass function) for X=test score. Name this pmf.p.

```
allOutcomes %>%
  attach()
  X=X1+X2+X3+X4+X5
  pmf.p <- table(X)/length(allOutcomes)
  pmf.p
```

```
## X
##   0   1   2   3   4   5
## 0.2 1.0 2.0 2.0 1.0 0.2
```

3. Using the PMF, use R to construct a histogram of "test scores."

```
hist(X, breaks=seq(-0.5, 5.5, 1), freq=F, col="darkblue", xaxt="n", yaxt="n", ylab = 'Density [units]',
     main="Probability Distribution for all Outcomes of Test Score", cex.main=1.5, cex.axis=1, cex.lab=
axis(2, tck=1, col.ticks="light gray", lwd.ticks="1")
axis(2, tck=-0.015)
axis(1,tck=-.015)
box()
hist(X, add=TRUE, breaks=seq(-0.5, 5.5, 1), freq=F, col="darkblue", xaxt="n", yaxt="n", ylab = 'Density
     main="Probability Distribution for all Outcomes of Test Score", cex.main=1.5, cex.axis=1, cex.lab=
```



4. Use R to calculate the population mean and population standard deviation of test scores.

```
mu.p <- mean(X)
sigma.p <- sqrt(var(X)*((length(X)-1)/length(X)))
cat(paste0('The population mean is: ',mu.p))
```

```
## The population mean is: 2.5
```

```
cat(paste0('The population standard deviation is: ',sigma.p))
```

```
## The population standard deviation is: 1.11803398874989
```

5. Set the seed to (111) and based on your pmf.p, take a SRS of size 1000 of "test scores." Use this sample to create the PMF of the sample (call it pmf.s).

```
set.seed(111)
size.s = 1000
s <- sample(0:5,size=size.s,replace=T,prob=pmf.p)
pmf.s <- table(s)/length(s)
```

6. Use R to calculate the sample mean and sample standard deviation of your 1000 sampled "test scores."

```
mu.s <- mean(s)
sigma.s <- sd(s)
cat(paste0('The sample mean is: ',mu.s))
```

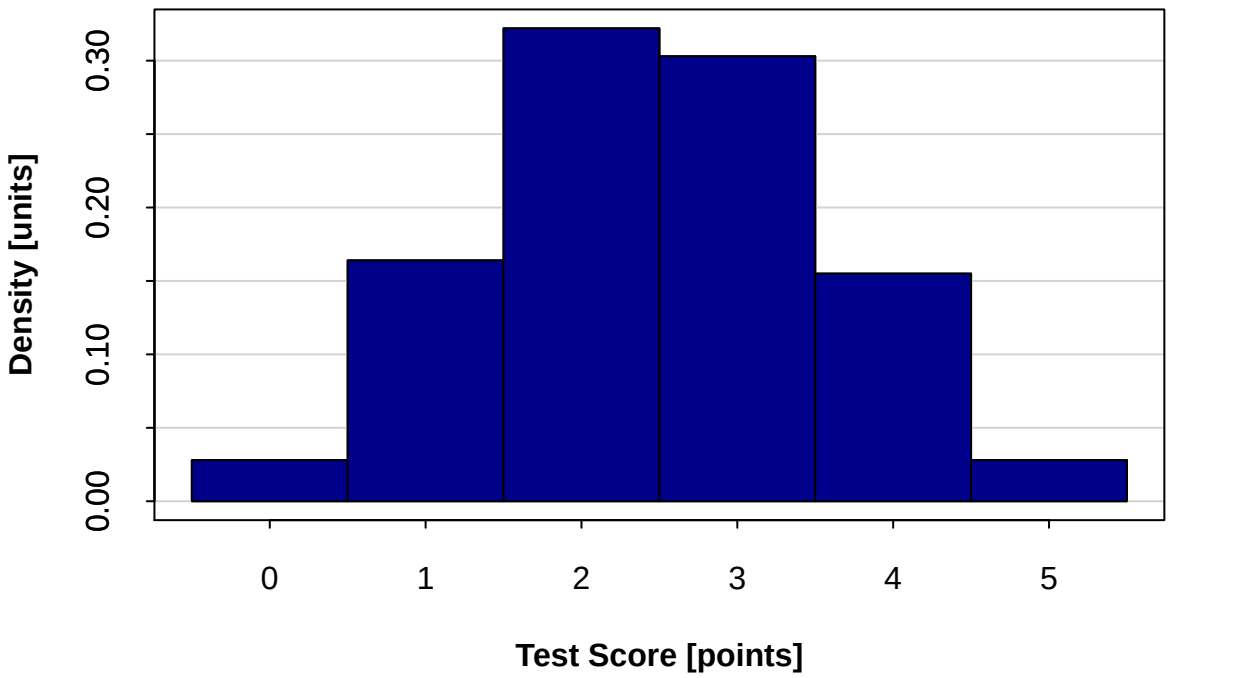
```
## The sample mean is: 2.477
```

```
cat(paste0('The sample standard deviation is: ',sigma.s))
```

```
## The sample standard deviation is: 1.10665970184863
```

7. Using the sample PMF (pmf.s), use R to construct a histogram of sampled "test scores."

```
hist(s, breaks=seq(-0.5, 5.5, 1),freq=F, col="darkblue", xaxt="n", yaxt="n", ylab = 'Density [units]', 
     main="Probability Distribution for Sample of 1000 Test Scores", cex.main=1.5, cex.axis=1, cex.lab=
axis(2, tck=1, col.ticks="light gray", lwd.ticks="1")
axis(2, tck=-0.015)
axis(1,tck=-.015)
box()
hist(s, add=TRUE, breaks=seq(-0.5, 5.5, 1),freq=F, col="darkblue", xaxt="n", yaxt="n", ylab = 'Density'
     main="Probability Distribution for Sample of 1000 Test Scores", cex.main=1.5, cex.axis=1, cex.lab=
```



8. Compare and contrast the Shape, Center and Spread of the pmf.p and the pmf.s.

The shape of both histograms show a normal distribution which appears Gaussian having a bell shaped curve. However, the sample of 1000 better shows the center to be a test score of 2 while the population is centered on a test score of 2 and 3. The spread of both appear to be very similar with highest probabilities being near the center.