Exam 1

Jonathan De Los Santos

Question 1

The data set below contains the total cost (in dollars) of attending 26 different baseball games in 2012. (8 points) Data set: baseball.csv (data = read.csv("http://tiny.cc/mbr27y (http://tiny.cc/mbr27y)")) Report mean, median, standard deviation, skewness, and kurtosis of the data.

Reading in baseball cost data:

```
baseball_data <- read.csv("http://tiny.cc/mbr27y")
```

Print the mean, median, and standard deviation of cost:

```
mean(baseball_data$cost)
```

```
## [1] 218.2917
```

```
median(baseball_data$cost)
```

```
## [1] 220.5
```

```
sd(baseball_data$cost)
```

```
## [1] 32.8825
```

To find skewness, we need to load the e1071 package from the R library:

```
library(e1071)
skewness(baseball_data$cost)
```

```
## [1] -0.1094914
```

This indicates that the distribution is skewed slightly to the left.

Finally, we can see the kurtosis of this distribution

```
kurtosis(baseball_data$cost)
```

```
## [1] -0.7752946
```

Question 2

Roll a fair die 100 times by simulation in R. A fair die is 6-sided (1,2,3,4,5,6) and all possible outcomes are equally likely Create our die and simulate 100 rolls:

```
X = c(1,2,3,4,5,6)
# Use sample to create 100 rolls of our die WITH replacement
Xsim <- sample(X, 100, replace = TRUE)
```

A. Estimate the mean and standard deviation of the outcome

```
mean(Xsim)
```

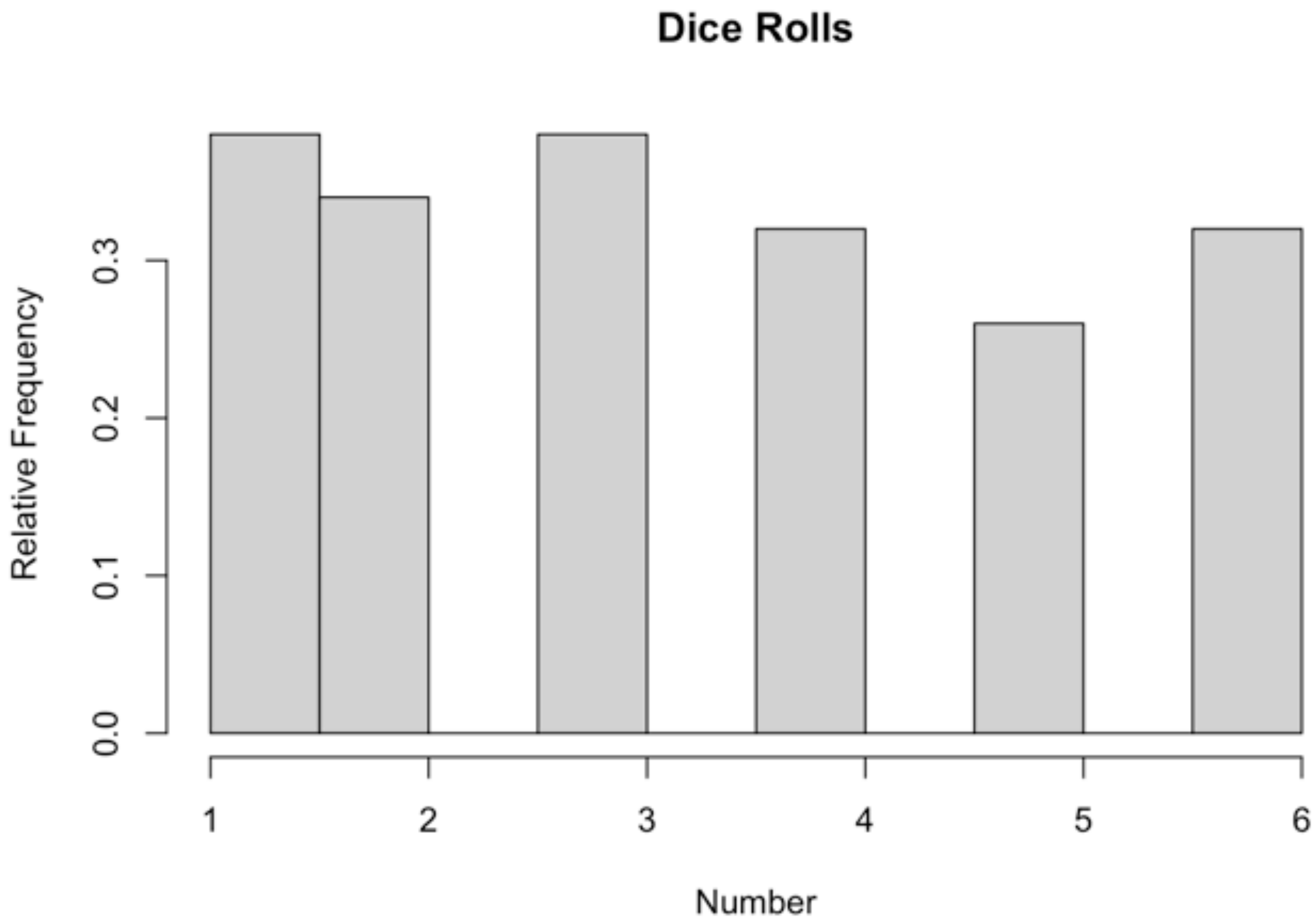
```
## [1] 3.35
```

```
sd(Xsim)
```

```
## [1] 1.719614
```

B. Create a histogram representing the relative frequency in the y-axis

```
hist(Xsim, freq = FALSE, main = "Dice Rolls", ylab = "Relative Frequency", xlab = "Nu
mber")
```



C. This time roll the fair die 10000 times and repeat parts a and b. What sample size provides more precise estimations? This time we see that increasing the simulation to 10,000 rolls gives us more equal relative frequencies across the board. The estimations of mean and standard deviation are also more accurate due to the law of large numbers (LLN).

```
X2 = c(1,2,3,4,5,6)
# Use sample to create 10000 rolls of our die WITH replacement
X2sim <- sample(X2, 10000, replace = TRUE)

# Run the mean, standard deviation, and histograms again
mean(X2sim)
```

```
## [1] 3.513
```

```
sd(X2sim)
```

```
## [1] 1.703268
```

```
hist(X2sim, freq = FALSE, main = "Dice Rolls", ylab = "Relative Frequency", xlab = "N
umber")
```