Weekly Assignment 4

Sean Leggett - BDA201 Winter 2020
February 25, 2020

Part 1:

What is the expected value thrown of a fair 12-sided dice?
What is the expected variance of a fair 12-sided dice?
Simulate 10,000 throws in R (using random function) from this dice and see if your answers match those above. Record the average value from the 10,000 throws and comment on the results

Answer:
The subjective probability would appear to be 6 as an expected value of a thrown 12-sided die.

However, empirical probability suggests we should calculate the actual answer. The expected value for calculating the expected value for a finite, discrete range, is the mean (average) of all options. In this case:

```
##establish object for reusability

die12 <- (1:12)
die12ev <- mean(die12)
die12ev
```

```
## [1] 6.5
```

The expected variance of a 12-sided die is calculated as follows:

```
die12var <- var(die12)
die12var
```

```
## [1] 13
```

Following is a simulation of 10,000 rolls of a 12-sided die:

```
sample10K <- sample(die12, 10000, replace = TRUE)
samplemean <- mean(sample10K)
samplemean
```

```
## [1] 6.4773
```

While intuition and the human mind might not appreciate fractional means of whole numbers on a die, statistics does. When running this prior to knit, we obtained a mean of 6.4768 (will be slightly different in knitted pdf version as it will run again). This is very close to our calculated mean of 6.5. Plot of our rolls shows the following:

```
tab10K <- table(sample10K)
barplot(tab10K)
```
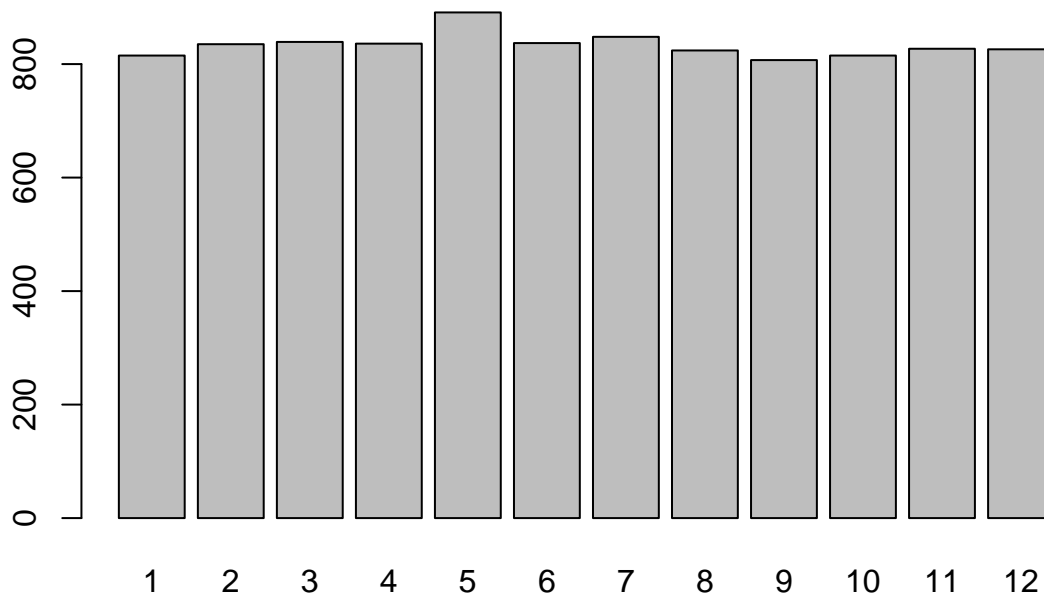
Fairly neat and uniform distributions across all possible outcomes. This is also a good illustration of the law of large numbers. Initially we could have great variability in the outcomes especially over say 5 or 10 rolls. However, after 10,000 rolls we see the outcomes reasoably even in distribution reinforcing the fact there is an equal chance on a roll or any number from 1 to 12.

Part 2:

For each year from 2000 to 2010, record the following over the 12 months of the year from "HAMILTON A" station.
- Total snowfall
- Average of the Mean temp column (the sums and averages are reported at the bottom of the table).

1. Plot these two variables against time. Superimpose the mean of all 11 years over the plot as horizontal brown line.

2. Use an appropriate plot to show and verify whether the 11 values are possibly from a normal distribution.

3. Write a function that accepts a array of values and returns a array of z_scores. Use the function to display z_scores of the 11 values.

Answer:

See code below where dataframe built with variables of annual total snowfall, average mean temperature and year.

```
## create dataframe from webpage summary data.
HamA <- data.frame("Year" = c(2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010),
                   "Snow(cm)" = c(170.9, 94.1, 138, 166.2, 175.8, 218.4, 56.6, 182.4, 243.2, 93, 70.2),
                   "AvgTemp(c)" = c(7.6, 8.8, 8.8, 7.3, 7.7, 8.2, 9.1, 8.2, 7.7, 7.3, 8.7)
                   )
HamA
```

```
##    Year Snow.cm. AvgTemp.c.
## 1  2000    170.9        7.6
## 2  2001     94.1        8.8
## 3  2002    138.0        8.8
## 4  2003    166.2        7.3
## 5  2004    175.8        7.7
## 6  2005    218.4        8.2
## 7  2006     56.6        9.1
## 8  2007    182.4        8.2
## 9  2008    243.2        7.7
## 10 2009     93.0        7.3
## 11 2010     70.2        8.7
```

```
## plot snowfall with mean abline
plot(x = HamA$Year,
     y = HamA$Snow.cm.,
     type = "l",
     main = "Annual Total Snowfall Hamilton A Station",
     xlab = "Year",
     ylab = "Snowfall (cm)")
abline(h = mean(HamA$Snow.cm.),
       lty = 1,
       col = "brown")
```

Annual Total Snowfall Hamilton A Station
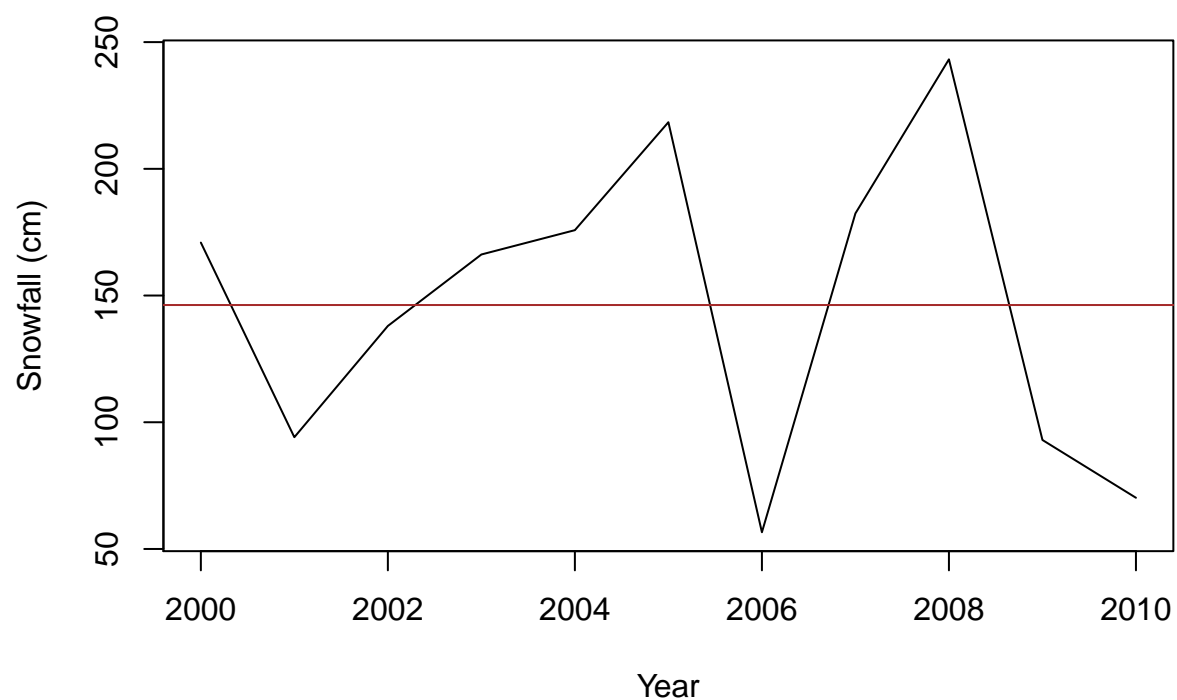

```r
## plot average temperature with mean abline
plot(x = HamA$Year,
     y = HamA$AvgTemp.c.,
     type = "l",
     main = "Annual Average Temperature Hamilton A Station",
     xlab = "Year",
     ylab = "Temperature (c)")
abline(h = mean(HamA$AvgTemp.c.),
       lty = 1,
       col = "brown")
```

Annual Average Temperature Hamilton A Station