Lab 3 Key

Anwesha Guha & Janette Avelar

1/25/2022

For lab 3, we're going to be exploring the concepts of normal distribution and sampling distribution by creating our own custom population and samples using R. We have provided the necessary code for you to create the population and pull samples, but you are expected to draw from previous labs and class materials to answer the following questions.

The first thing we'll do is read in the population data. Data from this week comes from the FiveThirtyEight story Should Travelers Avoid Flying Airlines That Have Had Crashes in the Past?.

```
airlines <- read.csv("https://raw.githubusercontent.com/fivethirtyeight/data/master/airline-safety/airli
```

For the purpose of this assignment, we will be looking at the total number of incidents from 2000 to 2014, which is the column `airlines$incidents_00_14`.

Question 1: Report the descriptive statistics for your population, `airlines$incidents_00_14`.

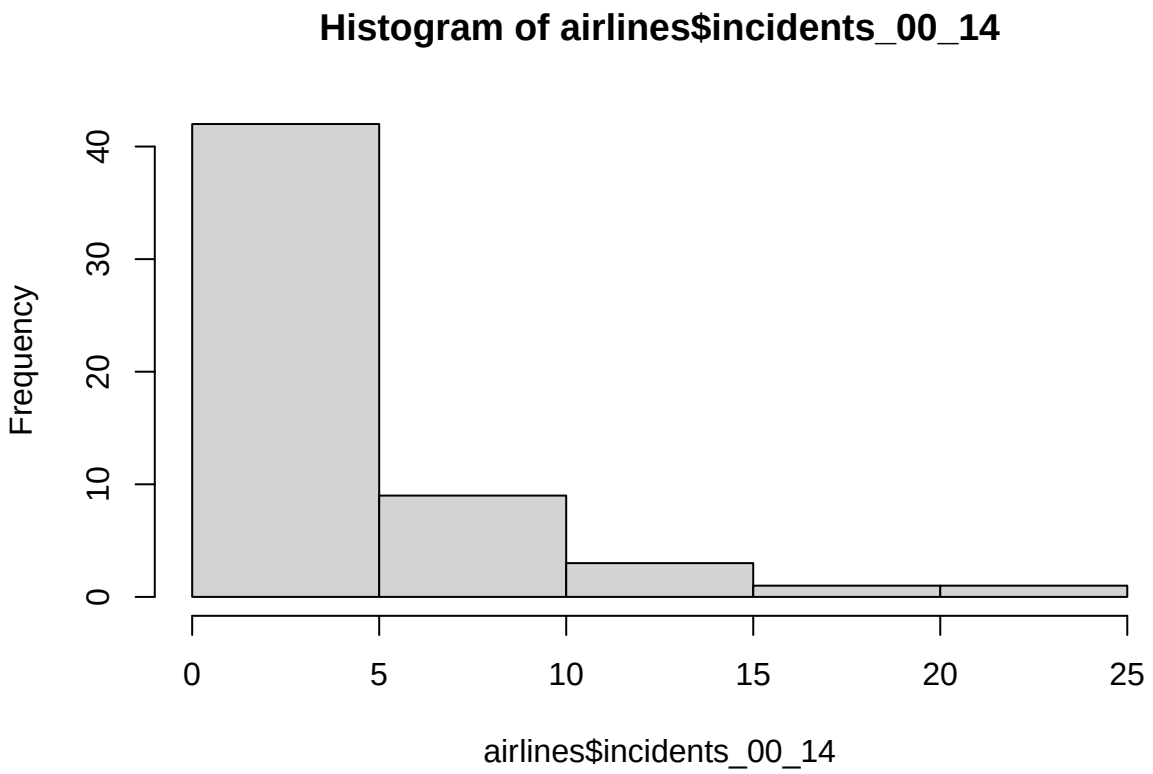
```
library(psych)
describe(airlines$incidents_00_14)
```

```
##    vars  n mean   sd median trimmed  mad min max range skew kurtosis   se
## X1    1 56 4.12 4.54      3    3.35 2.97   0  24    24 2.09     5.49 0.61
```

Mean: 4.12 Median: 3 Standard deviation: 4.54 Skew: 2.09 Kurtosis: 5.49

Question 2: Create a histogram for your population data.

```
hist(airlines$incidents_00_14)
```



Question 3: Now we're going to generate a random sample of 10 from our population (n=10).

```
set.seed(100)

sample1 <- sample(airlines$incidents_00_14, 10, replace = TRUE)

describe(sample1)
```

```
##    vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1    1 10  4.2 3.05    4.5    4.25 4.45   0   8     8 -0.17     -1.6 0.96
```

Report the descriptive statistics for your sample.
Mean: 4.2 Median: 4.5 Standard deviation: 3.05 Skew: -0.17 Kurtosis: -1.6

Question 4: We're going to generate a random sample of 10 again, but this time we'll do it 30 times.

```
set.seed(100)
sample_list <- list()

samples_30 <- replicate(30, sample(airlines$incidents_00_14, 10, replace = TRUE))
describe(samples_30)
```

```
##     vars  n mean   sd median trimmed  mad min max range  skew kurtosis   se
## X1     1 10  4.2 3.05    4.5    4.25 4.45   0   8     8 -0.17    -1.60 0.96
## X2     2 10  3.1 2.28    3.5    3.12 2.97   0   6     6 -0.11    -1.67 0.72
## X3     3 10  5.6 4.88    5.0    4.88 3.71   0  17    17  0.98     0.34 1.54
## X4     4 10  5.8 4.83    4.5    5.38 3.71   1  14    13  0.71    -1.08 1.53
## X5     5 10  2.6 2.46    2.0    2.50 2.97   0   6     6  0.26    -1.77 0.78
## X6     6 10  4.2 4.89    1.0    3.88 1.48   0  11    11  0.57    -1.70 1.55
```

```
## X7     7 10  4.4 7.20    2.5    2.50 3.71   0  24    24  1.93     2.55 2.28
## X8     8 10  5.1 4.63    3.0    4.62 3.71   0  14    14  0.58    -1.19 1.46
## X9     9 10  3.9 5.17    2.0    2.75 1.48   0  17    17  1.56     1.22 1.64
## X10   10 10  6.0 4.76    4.5    5.75 5.19   0  14    14  0.32    -1.53 1.51
## X11   11 10  5.4 6.29    2.5    4.62 0.74   0  17    17  1.12    -0.60 1.99
## X12   12 10  4.0 3.46    3.0    3.62 2.97   0  11    11  0.65    -0.92 1.10
## X13   13 10  2.3 2.21    1.5    2.12 1.48   0   6     6  0.54    -1.51 0.70
## X14   14 10  5.3 6.46    2.5    4.50 3.71   0  17    17  1.02    -0.73 2.04
## X15   15 10  2.8 1.87    2.0    2.62 1.48   1   6     5  0.53    -1.50 0.59
## X16   16 10  3.6 5.06    2.5    2.38 2.97   0  17    17  1.73     2.01 1.60
## X17   17 10  4.0 3.02    3.0    3.62 2.97   1  10     9  0.76    -0.89 0.95
## X18   18 10  3.8 3.12    4.0    3.38 1.48   0  11    11  0.86     0.29 0.99
## X19   19 10  3.3 3.20    3.0    2.88 3.71   0  10    10  0.65    -0.67 1.01
## X20   20 10  3.1 2.56    2.5    3.00 2.22   0   7     7  0.40    -1.53 0.81
## X21   21 10  7.9 8.97    5.0    6.88 5.19   0  24    24  0.98    -0.79 2.84
## X22   22 10  5.8 7.28    4.5    4.25 5.19   0  24    24  1.44     1.04 2.30
## X23   23 10  5.4 5.78    3.0    4.62 2.22   0  17    17  0.97    -0.74 1.83
## X24   24 10  6.4 7.90    3.5    5.00 3.71   0  24    24  1.19    -0.11 2.50
## X25   25 10  3.8 3.85    2.5    3.00 1.48   0  14    14  1.72     2.03 1.22
## X26   26 10  8.9 9.60    5.5    8.12 8.15   0  24    24  0.51    -1.50 3.03
## X27   27 10  3.1 3.28    2.0    2.62 2.22   0  10    10  0.89    -0.64 1.04
## X28   28 10  6.7 7.83    4.5    5.38 5.19   0  24    24  1.14    -0.20 2.48
## X29   29 10  4.1 3.45    3.5    3.75 3.71   0  11    11  0.58    -0.95 1.09
## X30   30 10  4.4 5.48    2.5    3.38 3.71   0  17    17  1.13     0.05 1.73
```

Report the following descriptive statistics for your sampling distribution.

```
mean(samples_30)
```

```
## [1] 4.633333
```

```
median(samples_30)
```

```
## [1] 3
```

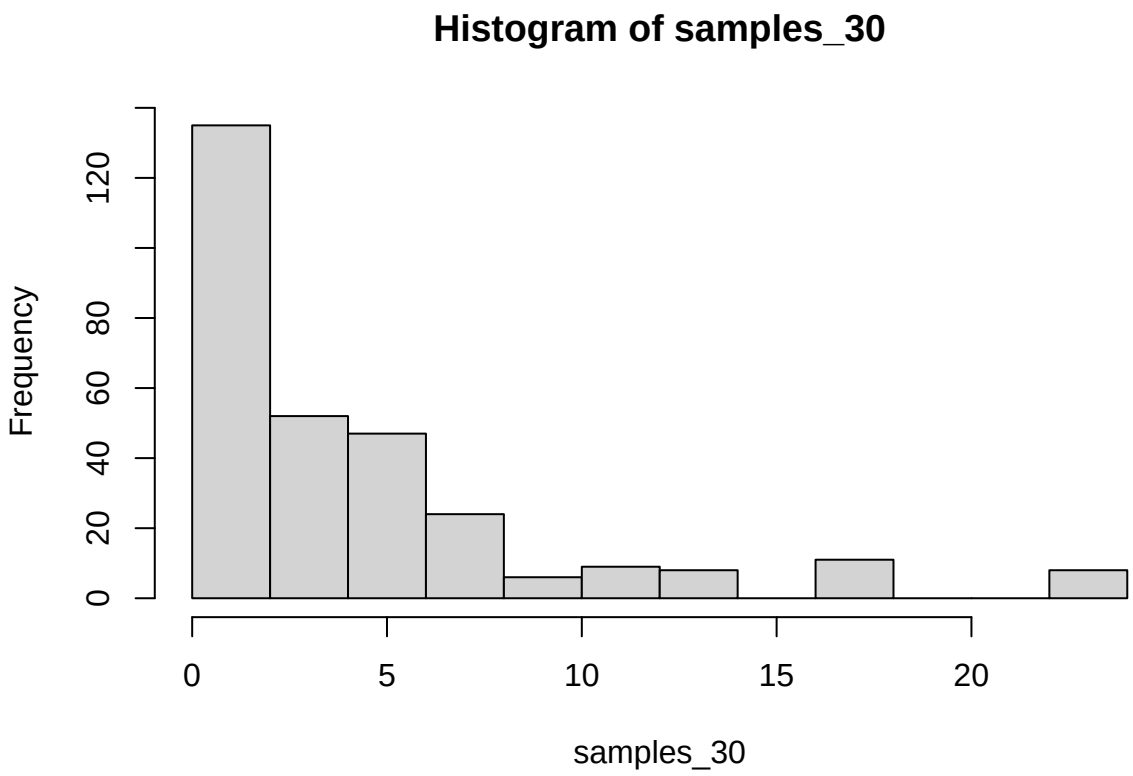
```
sd(samples_30)
```

```
## [1] 5.220767
```

Mean:
Median:
SD:

Question 5: Create a histogram for your sampling distribution.

```
hist(samples_30)
```



Question 6: Suppose that you randomly sample from your population with a size of 10 and compute the mean for each sample. You repeat this an infinite number of times. What would you expect the mean and standard deviation of your sampling distribution to be? Compute the mean and standard deviation of the hypothetical sampling distribution and explain your reasoning.

Question 7: Let's check our work by drawing 1,000,000 random samples of 10 from our population.

```
set.seed(100)
samples_mil <- replicate(1000000, sample(airlines$incidents_00_14, 10, replace = TRUE))
```

Calculate the descriptive statistics for your 1,000,000 samples. Compare them to the descriptive statistics of your sampling distribution in questions 1 and 2. Do they match or not? Explain why you think that is.

```
mean(samples_mil)
```

```
## [1] 4.123196
```

```
sd(samples_mil)
```

```
## [1] 4.501777
```

```
median(samples_mil)
```

```
## [1] 3
```