PSTAT174_Lab01

Celeste Herrera

4/1/2021

1. Let X be U(-1,1) be uniform on (-1,1), i.e., $f_X(x)$ = 1/2 when -1 < x < 1 and zero otherwise. Let Y = X^2. Calculate the correlation of X and Y ? Now, use runif() in R to generate 1000 I.I.D samples of X's from U(-1,1), and Y's, calculate the sample correlation of X and Y. Are X and Y uncorrelated? Are X and Y independent? Any conclusion on the relationship between uncorrelated and independent?

```
x <- runif(1000,min = -1, max = 1)
y <- x^2

cov(x,y)
```

```
## [1] 0.004002003
```

```
cor(x,y)
```

```
## [1] 0.02393375
```

With the following solutions x and y are conisdered to be uncorrelated and quadriatically dependent. We have the prior knowledge to infact state that all independent variables are uncorrelated. (Also see attatchment of writting for part of question 1).

2. Use runif() in R to generate 10, 100, and 1000 I.I.D samples from U(-1,1) respectively, calculate the sample means, compare these sample means with true mean? Any conclusion on the relationship between sample mean and true mean?

```
x1 <- runif(10, min = -1, max = 1)
x2 <- runif(100, min = -1, max = 1)
x3 <- runif(1000, min = -1, max = 1)

mean(x1)
```

```
## [1] 0.2136456
```

```
mean(x2)
```

```
## [1] 0.06559686
```

```
mean(x3)
```

```
## [1] 0.004712507
```

With our solutions we recognize that the true mean in this instance is known to be zero. This implies that more samples would indicate to be closer to the true mean.

Celeste Herrera

PSTAT174 Lab #1 April 12, 2021

1.) Let $X \sim U(-1,1)$ be unif $(-1,1)$

$f_x(x) = 1/2 \quad -1 < x < 1$ and 0 o.w.

$$cov(x,y) = E(xy) - E(x)E(y)$$

$$E[x] - 0 \cdot E(Y)$$

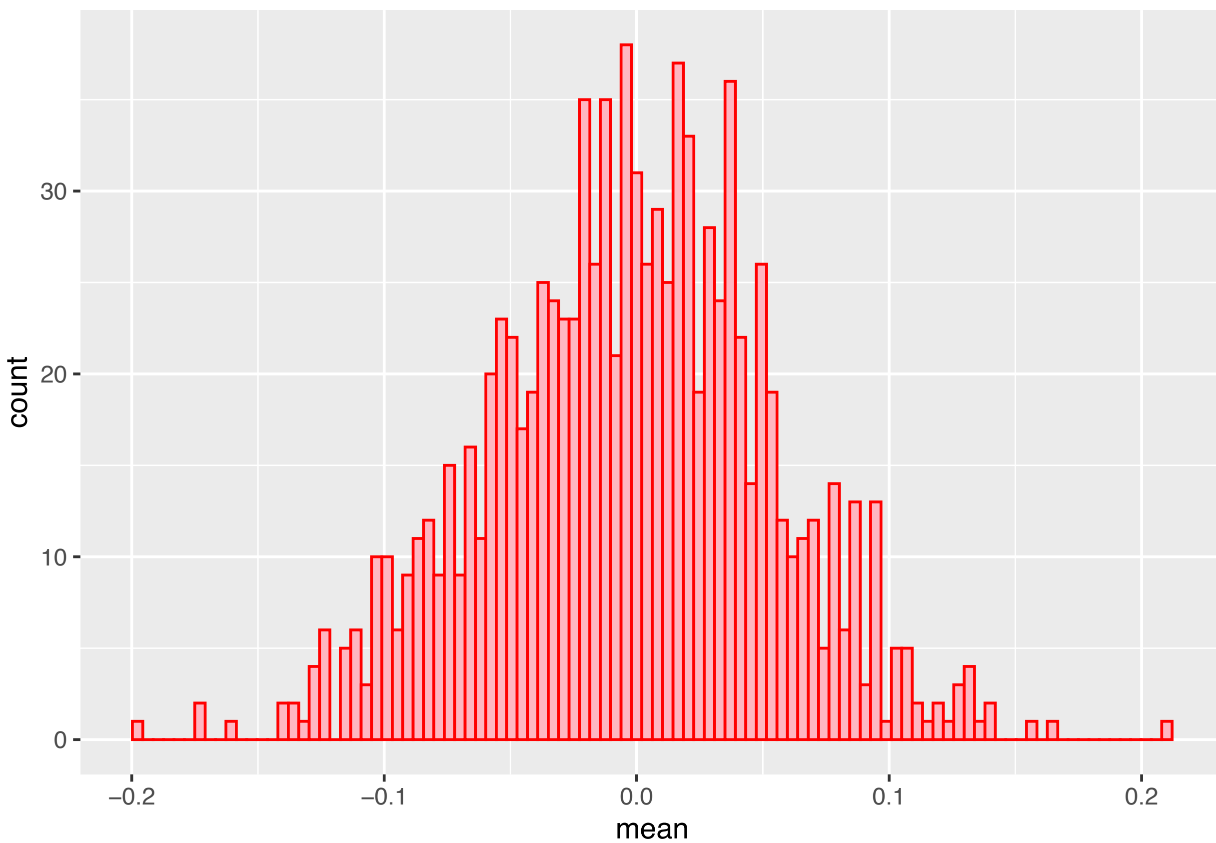
$$= \int_{-1}^{1} \frac{1}{2} x^3 \, dx$$

$$\frac{x^4}{4} \cdot \frac{1}{2} \Big|_{-1}^{1} = \frac{1}{8} - \left(\frac{1}{8}\right) = 0$$

Therefore, since we have a correlation of zero it would be safe to say Y is dependent but also uncorrelated.

3. Generate 1000 I.I.D samples of size 100 from U (-1, 1), and calculate the sample means. Now, we have 1000 sample means. Plot the histogram of these sample means. What's the asymptotic distribution of these sample means?

```
library(ggplot2)

iid = 1000
n = 100
samples = list()

mean = NULL
for(i in 1:iid){
  samples[[i]] = runif(100,-1,1)
  mean[i] = mean(samples[[i]])
}

graph = ggplot(data = data.frame(mean), aes(x = mean)) +
  geom_histogram(bins = 100, fill = "lightpink", col = "red")
graph
```



The asympototic distribution is known to be normal by the vision of the distribution it is safe to say there is a normal distribution.

4. The file uspop.txt contains US population from 1970 to 1990. Plot US population (in Millions) vs Year. Now, take the square root of US population (in Millions), and plot it vs Year. Any difference between these two graphs?

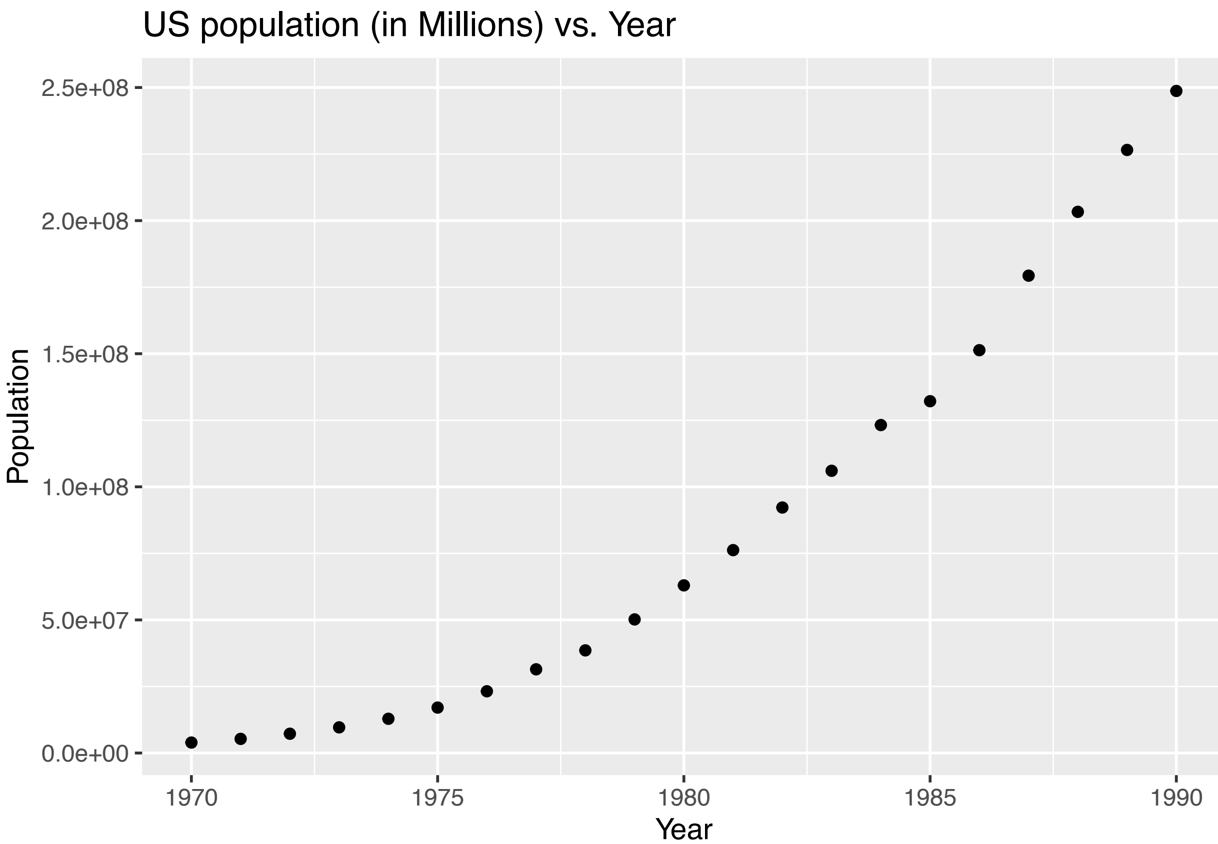
```
getwd()
```

```
## [1] "/Users/celesteherrera/Documents/PSTAT174/Lab/Lab01"
```

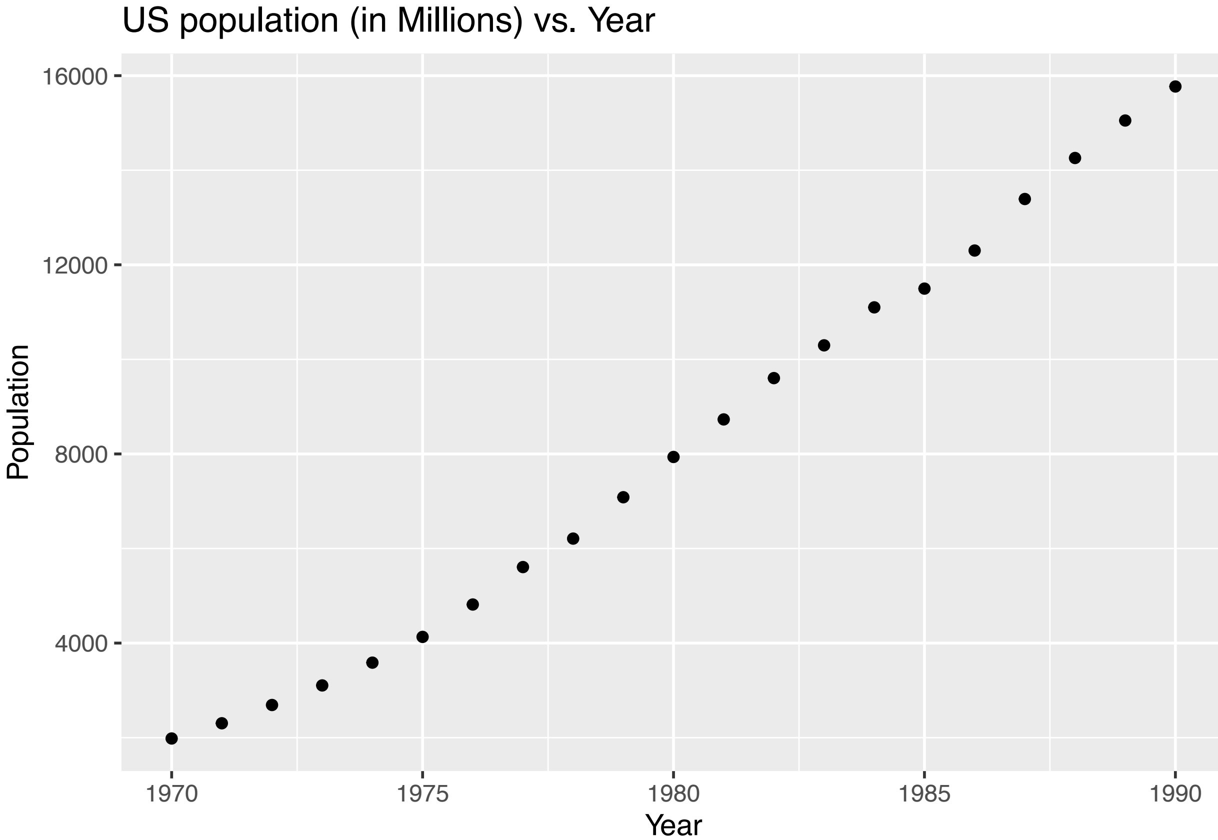
```
setwd("/Users/celesteherrera/Documents/PSTAT174/Lab/Lab01")
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
## v purrr   0.3.4

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
uspop = read.table("uspop.txt")

sqrt_uspop = sqrt(uspop$V1)
seq_val = seq(1970,1990, by = 1)

# Plot US population (in Millions) vs Year

uspop %>%
  ggplot(aes(x= seq_val, y = V1, xlab = "Population")) +
  ggtitle("US population (in Millions) vs. Year") +
  geom_point() +
  xlab("Year")+
  ylab("Population")
```




```
# the square root of US population (in Millions), and plot it vs Year

uspop %>%
  ggplot(aes(x= seq_val, y = sqrt_uspop, xlab = "Population")) +
  ggtitle("US population (in Millions) vs. Year") +
  geom_point() +
  xlab("Year")+
  ylab("Population")
```



The difference between the two graphs is that square root of US population (in Millions) vs Year is more linear. Whereas the original graph is know to have more of a curved shape.