Lab 1 Assignment

Dylan Berneman

2023-04-07

Question 1

Let X ~ 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 and 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?

$$X = U(-1,1)$$

$$\begin{aligned}
E[X] &= \int_{-1}^{1} \frac{x}{2}\, dx \\
&= \frac{1}{2}\int_{-1}^{1} x\, dx \\
&= \frac{1}{2}[x^2/2]_{-1}^{1} \\
&= \frac{1}{4}[1^2 - (-1)^2] \\
&= 0
\end{aligned}$$

$$\begin{aligned}
E[X^2] &= \int_{-1}^{1} \frac{x^2}{2} dx \\
&= \frac{1}{2}\int_{-1}^{1} x^2 dx \\
&= \frac{1}{2}[x^3/3]_{-1}^{1} \\
&= \frac{1}{6}[1^3 - (-1)^3] \\
&= \frac{1}{3}
\end{aligned}$$

$$
\begin{aligned}
Var[X] &= E[X^2] - E[X]^2 \\
&= 1/3 - 0^2 \\
&= 1/3
\end{aligned}
$$

$$
\begin{aligned}
SD[X] &= \sqrt{Var[X]} \\
&= \frac{1}{\sqrt{3}}
\end{aligned}
$$

$$
\begin{aligned}
&Y = X^2 \\
&P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) \\
&F_y(y) = \sqrt{y}, \quad for\ 0 \le y < 1 \\
&f_y(y) = \frac{dF_y(y)}{dy} \\
&\qquad = \frac{1}{2\sqrt{y}}
\end{aligned}
$$

$$
\begin{aligned}
E[Y] &= \int_0^1 \frac{y}{2\sqrt{y}}\ dy \\
&= \frac{1}{2}\int_0^1 \sqrt{y}\, dy \\
&= \frac{1}{2}[\frac{2y^{3/2}}{3}]_0^1 \\
&= \frac{1}{3}[1^{3/2}\ -\ 0^{3/2}] \\
&= \frac{1}{3}
\end{aligned}
$$

$$
\begin{aligned}
E[Y^2] &= \int_0^1 \frac{y^2}{2\sqrt{y}}\ dy \\
&= \frac{1}{2}\int_0^1 y^{3/2}\ dy \\
&= \frac{1}{2}[\frac{2y^{5/2}}{5}]_{-1}^1 \\
&= \frac{1}{5}[1^{5/2} - 0^{5/2}] \\
&= 1/5
\end{aligned}
$$

$$
\begin{aligned}
Var[Y] &= E[Y^2] - E[Y]^2 \\
&= 1/5 - (1/3)^2 \\
&= 1/5 - 1/9 \\
&= 4/45
\end{aligned}
$$

$$
\begin{aligned}
SD[Y] &= \sqrt{Var[Y]} \\
&= \sqrt{4/45} \\
&= \frac{2}{3\sqrt{5}}
\end{aligned}
$$

$$
\begin{aligned}
E[XY] &= \int_{-1}^{1} \int_{0}^{1} xy\ f_x(x) f_y(y)\ dy\ dx \\
&= \int_{-1}^{1} \int_{0}^{1} xy\ (\frac{1}{2})(\frac{1}{2\sqrt{y}})\ dy\ dx \\
&= \frac{1}{4} \int_{-1}^{1} \int_{0}^{1} x\sqrt{y}\ dy\ dx \\
&= \frac{1}{4} \int_{-1}^{1} x\ [\frac{2y^{3/2}}{3}]_0^1\ dx \\
&= \frac{1}{6} \int_{-1}^{1} x\ [1^{3/2} - 0^{3/2}]\ dx \\
&= \frac{1}{6} \int_{-1}^{1} x\ dx \\
&= \frac{1}{6} [\frac{x^2}{2}]_{-1}^{1} \\
&= \frac{1}{12} [1^2 - (-1)^2] \\
&= 0
\end{aligned}
$$

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

$$= \frac{E[XY] - E[X]E[Y]}{SD[X]\ SD[Y]}$$

$$= \frac{0 - 0(\frac{1}{3})}{\frac{1}{\sqrt{3}} * \frac{2}{3\sqrt{5}}}$$

$$= 0$$

```
set.seed(5)
X = runif(1000, -1,1)
Y = X^2
cor(X,Y)
```

```
## [1] 0.006708915
```

The sample correlation between X and Y is close to 0. Since the correlation coefficient can only be 0 when the covariance of X and Y is 0, this implies that the covariance is even closer to 0 than the correlation coefficient is. Therefore, X and Y are uncorrelated. Even though all independent variables are uncorrelated, the opposite is not true. In this case, since we already know that Y is a transformation of X, the variables X and Y cannot be independent.

Question 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?

```
set.seed(399)
X = runif(10, -1,1)
mean(X)
```

```
## [1] -0.1225091
```

```
X = runif(100, -1,1)
mean(X)
```

```
## [1] 0.05411703
```

```
X = runif(1000, -1,1)
mean(X)
```

```
## [1] -0.002858567
```

```
# True Mean
mean(dist_uniform(-1,1))
```

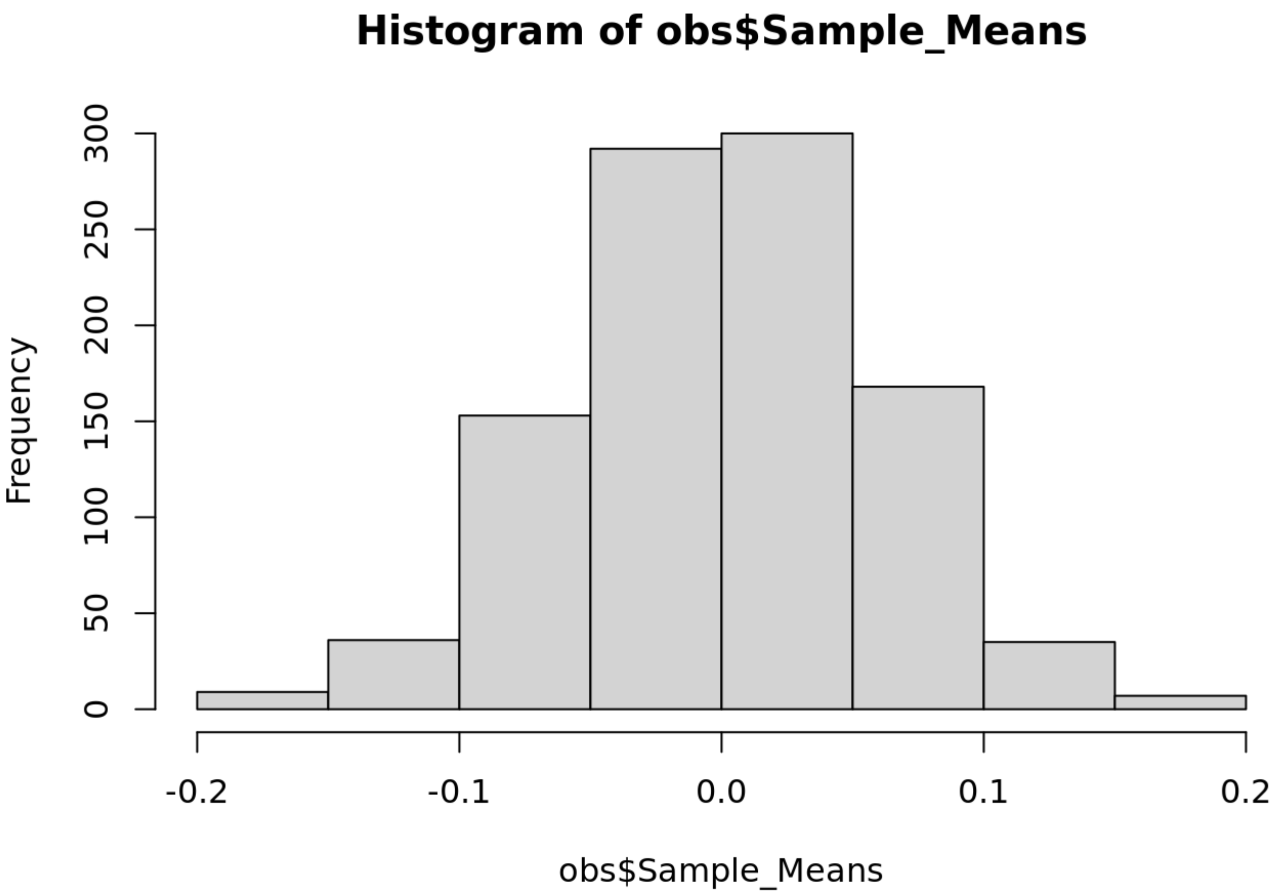
```
## [1] 0
```

As the number of observations from the sample increases, the difference between it and the true mean decreases.

Question 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?

```
obs <- data.frame(matrix(nrow=1000))
colnames(obs) <- "Sample_Means"
for(i in 1:1000){
  obs$Sample_Means[i] = mean(runif(100, -1, 1))}
hist(obs$Sample_Means)
```




```
obs
```

Sample_Means
<dbl>
7.849680e-02
1.043621e-01
-4.918774e-03
-6.255146e-02
-4.122957e-02
-7.508837e-03
1.681502e-01
7.721108e-02
-8.516522e-03

-4.435590e-02

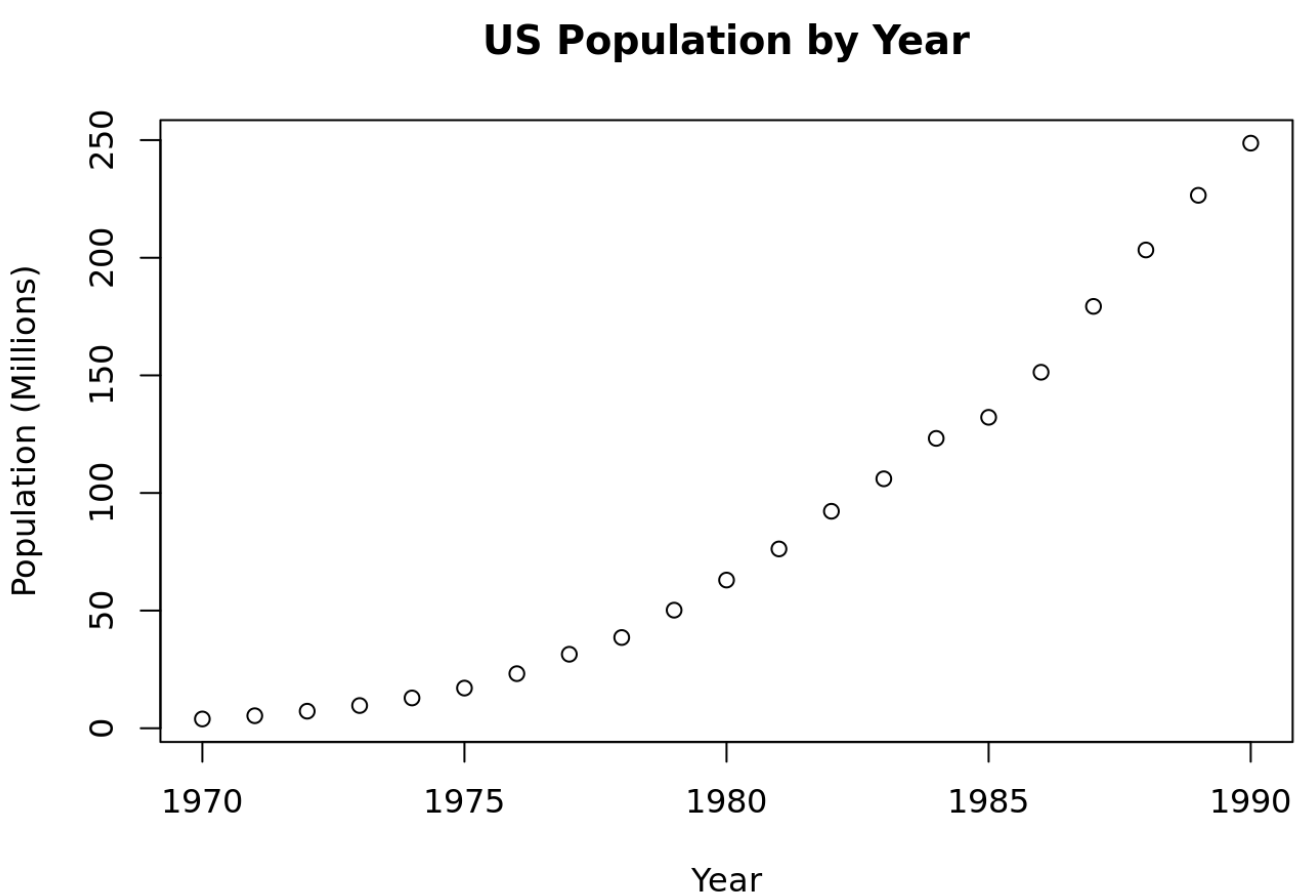
1-10 of 1,000 rows Previous **1** 2 3 4 5 6 ... 100 Next

The asymptotic distribution of the histogram of sample means is the normal distribution.

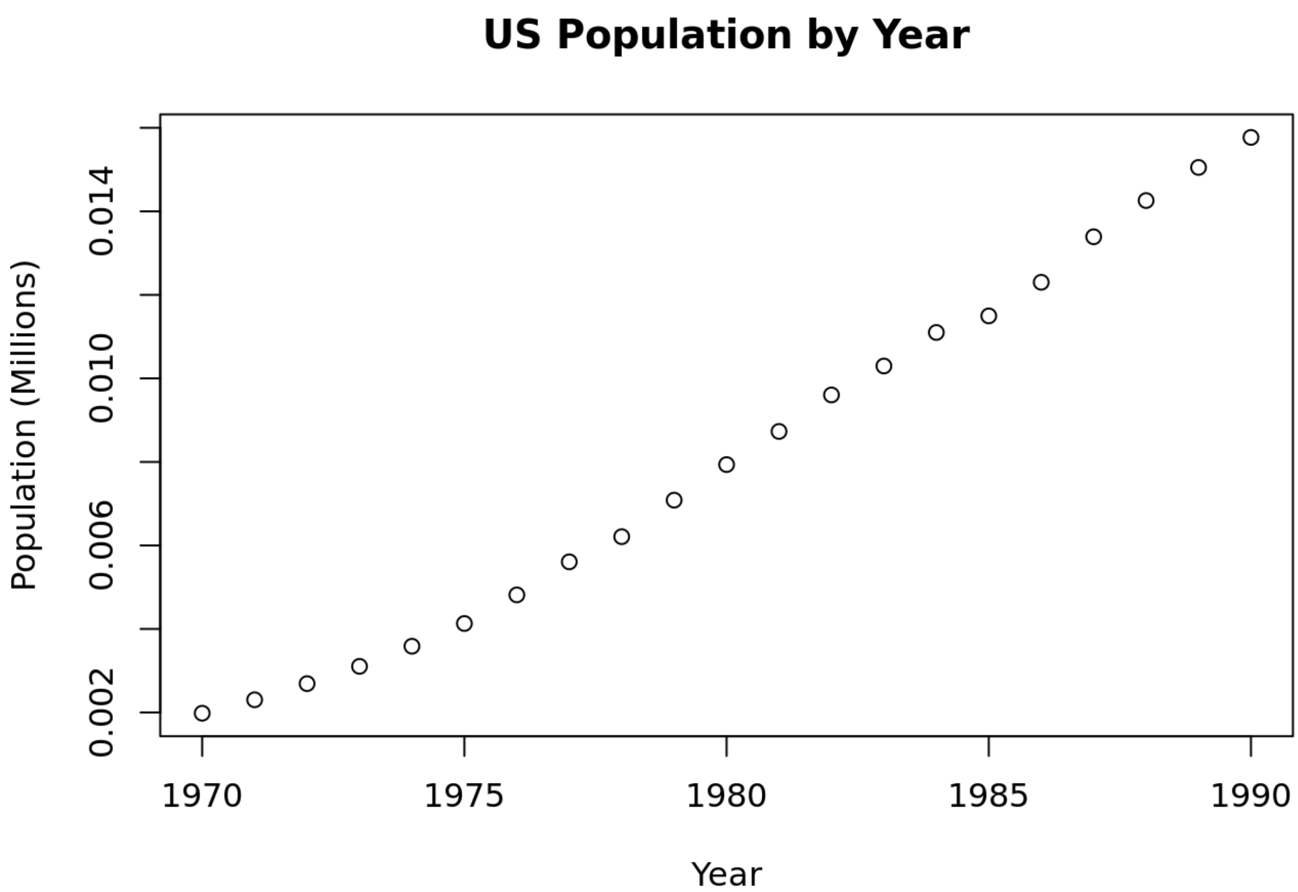
Question 4

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?

```
uspop <- read.csv("uspop.txt", header = FALSE)
colnames(uspop) <- "Population"
uspop["Year"] <- NA
uspop$Year <- c(seq(1970,1990,1))
plot(x=uspop$Year, y=uspop$Population/1000000, ylab = "Population (Millions)",
xlab="Year", main="US Population by Year")
```



```
plot(x=uspop$Year, y=sqrt(uspop$Population)/1000000, ylab = "Population (Millio
ns)", xlab="Year", main="US Population by Year")
```



The slope of the first plot starts off small, but it rapidly increases as time goes on. Meanwhile, the slope of the second plot increases a little bit in the beginning, but as time goes on, it becomes constant.