

Run the code below and answer the following questions

```
library(dplyr)
library(tidyr)
library(ggplot2)
library(gganimate)
set.seed(10)
options(scipen=999)
```

```
df1 <- sample(x=c(0,1), prob = c(1-0.04,0.04), size=100000, replace = TRUE)
```

```
df1 <- as.data.frame(df1) %>%
  rename(colon.cancer=1) %>%
  mutate(id = row_number()) %>%
  select(id, colon.cancer)
```

```
names(df1) <- c("id", "colon.cancer")
```

- a. Using the population cancer data `df1`, calculate the population mean and population variance for colon cancer.
- b. Draw a random sample ( $N=100$ ) from the population, and calculate the sample mean.
- c. Increase the sample size gradually ( $N=200, N=300$ , etc.) and re-calculate the sample mean.
- d. Draw 10 samples from the population simultaneously and calculate the mean from each sample.
- e. Calculate the mean and the variance of the sample means.
- f. Replicate d and e with 1000 samples.
- g. Draw a histogram of the sample means.