Pset2

Gabriel

October 13rd, 2025

1 Question 1

```
set.seed (123)
cor_values <- numeric(1000)
for (i in 1:1000) {
  a <- rnorm(20, mean = 0, sd = 1)
  b <- rnorm(20, mean = 0, sd = 1)
  cor_values[i] <- cor(a, b)
}
mean(cor_values)
#> mean(cor_values) [1] -0.0007532091
sd(cor_values)
#> sd(cor_values)[1] 0.2378933

plot(cor_values)
```

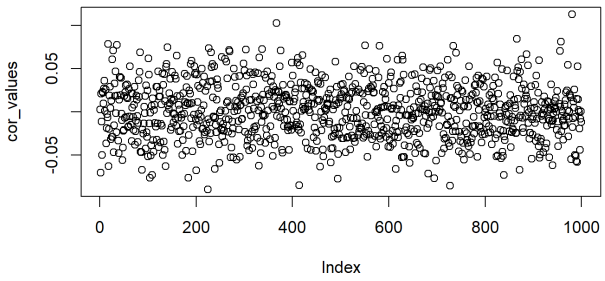

Figure 1: Plot(cor_values)question1

The standard deviation is 0.2378933; On average, the correlation between the two variables is -0.0007532091. We can say the correlation of the two random variables in population is 0. The correlation estimates of samples form a distribution around 0. The estimator is unbiased, while the standard deviation (SD 0.237) reflects the standard error.

—

2 Question 2

```
cor_values2 <- numeric(1000)
for (i in 1:1000) {
  a <- rnorm(1000, mean = 0, sd = 1)
  b <- rnorm(1000, mean = 0, sd = 1)
  cor_values2[i] <- cor(a, b)
}
mean(cor_values2)
#> mean(cor_values2) [1] 0.0002157361
sd(cor_values2)
```

```
#> sd(cor_values2)[1] 0.03222388

plot(cor_values2)
```

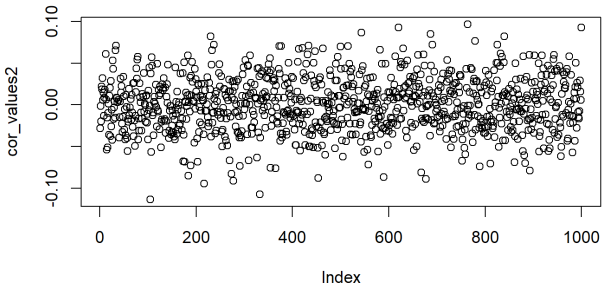

Figure 2: Plot(cor_values2)question2

The standard deviation is 0.0322, and the average sample correlation between the two variables is 0.0002. This suggests that the true population correlation is approximately 0. The sample correlation estimates form a distribution around this true value, indicating that the estimator is unbiased. The standard deviation (around 0.032) represents the standard error, reflecting the sampling variability of the estimator. As the sample size increases, both the standard deviation and the variability of the sample correlations decrease, meaning that the estimates become more stable and closer to the true population parameter.

—

3 Question 3

```
Z <- rnorm(20, mean = 0, sd = 1)
epsilonX <- rnorm(20)
epsilonY <- rnorm(20)
X <- rnorm(20) + 1.5 * Z + epsilonX
Y <- rnorm(20) + 2 * Z + epsilonY

plot(X, col = "black", pch = 16, ylim = range(c(X, Y)),
     main = "X (black) and Y (green)",
     ylab = "Values", xlab = "Index");
points(Y, col = "green", pch = 16)

print(cor(X, Y))
#> print(cor(X,Y)) [1] 0.4993344
```

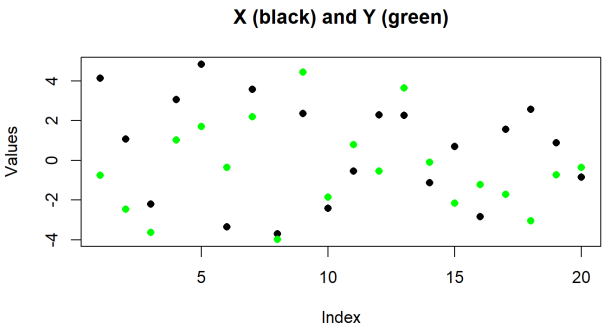

Figure 3: Plot(X,Y)question3

The correlation estimate between X and Y is 0.4993344, confirming that although X and Y do not have a direct

causal relationship, the existence of Z that causally influences both X and Y makes X and Y share a correlation. This illustrates that correlation does not imply causation.