Homework 9

Noah McIntire

Problem 1

a)

```
set.seed(07312001)
data1<-read.csv("/Users/noahmcintire/Downloads/OneDrive_1_11-8-2021/data1.csv")
cor1<-cor(data1$V1,data1$V2)
cor1
```

```
## [1] 0.524066
```

b)

```
K <- 10000
mu <- 0
pTtest <- function(df){
  #draw paired samples
  samps <- df[sample(nrow(df), 13, replace= T),]
  # test stat
  pvalue <- t.test(samps$V1, samps$V2, paired = T, mu=0, alternative="two.sided")$p.valu
  reject_null <- (pvalue < 0.05)
  reject_null
}
t1error1b <- sum(replicate(K, pTtest(data1)))/K
t1error1b
```

```
## [1] 0.0486
```

c)

```
twoSamptest <- function(df){
  #draw paired samples
  samps <- df[sample(nrow(df), 13, replace= T),]
  # test stat
  pvalue <- t.test(samps$V1, samps$V2, mu=0, alternative="two.sided")$p.value
  reject_null <- (pvalue < 0.05)
  reject_null
}
t1error1c <- sum(replicate(K, twoSamptest(data1)))/K
t1error1c
```

```
## [1] 0.0085
```

Problem 2

a)

```
data2<-read.csv("/Users/noahmcintire/Downloads/OneDrive_1_11-8-2021/data2.csv")
cor2<-cor(data2$V1,data2$V2)
cor2
```

```
## [1] -0.52036
```

b)

```
t1error2b <- sum(replicate(K, pTtest(data2)))/K
t1error2b
```

```
## [1] 0.0497
```

c)

```
t1error2c <- sum(replicate(K, twoSamptest(data2)))/K
t1error2c
```

```
## [1] 0.1085
```

Problem 3

a)

```
data3<-read.csv("/Users/noahmcintire/Downloads/OneDrive_1_11-8-2021/data3.csv")
cor3<-cor(data3$V1,data3$V2)
cor3
```

```
## [1] 0.002426237
```

b)

```
t1error3b <- sum(replicate(K, pTtest(data3)))/K
t1error3b
```

```
## [1] 0.0514
```

c)

```
t1error3c <- sum(replicate(K, twoSamptest(data3)))/K
t1error3c
```

```
## [1] 0.0486
```

Problem 4

```
problem1 <- c(cor1,t1error1b, t1error1c)
problem2 <- c(cor2,t1error2b, t1error2c)
problem3 <- c(cor3,t1error3b, t1error3c)
summary1 <- data.frame(problem1, problem2, problem3)
row.names(summary1) <- c("Correlation", "Paired T-test", "Two Sample T-Test")
summary1
```

```
##                   problem1 problem2    problem3
## Correlation       0.524066 -0.52036 0.002426237
## Paired T-test     0.048600  0.04970 0.051400000
## Two Sample T-Test 0.008500  0.10850 0.048600000
```

All three problems are based off of symmetric distributions, just with different correlations among the data sets. Even with a stronger positive correlation (seen in problem 1), we can see that the empirical type one error with a paired t-test in two vectors that are more correlated is equivalent to the dataset where the two vectors are almost not correlated at all (problem 3). Additionally, we can see with a negative correlation between the two vectors (problem 2) that the empirical type 1 error among two-sample t-test is much higher when the correlation is negative v when it is positive. This trend is also seen when the is no correlation among two vectors (problem 3), as the two-sample t-test empirical type one error is higher than that of two vectors with correlation (problem 1), but less than that of those with a negative correlation (problem 2).

Problem 5

a)

```
data4<-read.csv("/Users/noahmcintire/Downloads/OneDrive_1_11-8-2021/data4.csv")
cor5<-cor(data4$V1,data4$V2)
cor5
```

```
## [1] 0.5906402
```

b)

```
t1error5b <- sum(replicate(K, pTtest(data4)))/K
t1error5b
```

```
## [1] 0.0494
```

c)

```
t1error5c <- sum(replicate(K, twoSamptest(data4)))/K
t1error5c
```

```
## [1] 0.0161
```

Problem 6

a)

```
data5<-read.csv("/Users/noahmcintire/Downloads/OneDrive_1_11-8-2021/data5.csv")
cor6<-cor(data5$V1,data5$V2)
cor6
```

```
## [1] -0.5721193
```

b)

```
t1error6b <- sum(replicate(K, pTtest(data5)))/K
t1error6b
```

```
## [1] 0.0746
```

c)

```
t1error6c <- sum(replicate(K, twoSamptest(data5)))/K
t1error6c
```

```
## [1] 0.1142
```

Problem 7

a)

```
data6<-read.csv("/Users/noahmcintire/Downloads/OneDrive_1_11-8-2021/data6.csv")
cor7<-cor(data6$V1,data6$V2)
cor7
```

```
## [1] -0.007297158
```

b)

```
t1error7b <- sum(replicate(K, pTtest(data6)))/K
t1error7b
```

```
## [1] 0.0377
```

c)

```
t1error7c <- sum(replicate(K, twoSamptest(data6)))/K
t1error7c
```

```
## [1] 0.0324
```

Problem 8

```
problem5 <- c(cor5,t1error5b, t1error5c)
problem6 <- c(cor6,t1error6b, t1error6c)
problem7 <- c(cor7,t1error7b, t1error7c)
summary2 <- data.frame(problem5, problem6, problem7)
row.names(summary2) <- c("Correlation", "Paired T-test", "Two Sample T-Test")
summary2
```

```
##                    problem5   problem6     problem7
## Correlation       0.5906402 -0.5721193 -0.007297158
## Paired T-test     0.0494000  0.0746000  0.037700000
## Two Sample T-Test 0.0161000  0.1142000  0.032400000
```

All three problems are based off of skewed distributions, just with different correlations among the data sets. Even with a stronger positive correlation (seen in problem 5), we can see that the empirical type one error with a paired t-test in two vectors that are more correlated is equivalent to the dataset where the two vectors are almost not correlated at all (problem 7). Additionally, we can see with a negative correlation between the two vectors (problem 6) that the empirical type 1 error among two-sample t-test is much higher when the correlation is negative v when it is positive. This trend is also seen when the is no correlation among two vectors (problem 7), as the two-sample t-test empirical type one error is higher than that of two vectors with correlation (problem 5), but less than that of those with a negative correlation (problem 6).