# Homework 9

Santhosh Rajendran

## Setting Seed

```
set.seed(06072001)
```

## Problem 1

#a

```
setwd('/Users/santhoshrajendran/Documents/3rd Year 1st Sem HW Files/STAT 3080')
data1 <- read.csv("Data1.csv")
corr1 <- cor(data1$V1, data1$V2)
corr1
```

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

#b

```
pairedt <- function(df){
  samp <- df[sample(nrow(df), 13), ]
  test <- t.test(samp$V1, samp$V2, mu=0, alternative="two.sided", paired=TRUE)
  pval <- test$p.value
  pval < 0.05
}
paired1 <- sum(replicate(10000, pairedt(data1)) == TRUE) / 10000
paired1
```

```
## [1] 0.0474
```

#c

```
twosamp <- function(df){
  samp <- df[sample(nrow(df), 13), ]
  test <- t.test(samp$V1, samp$V2, mu=0, alternative="two.sided")
  pval <- test$p.value
  pval < 0.05
}
twosamp1 <- sum(replicate(10000, twosamp(data1)) == TRUE) / 10000
twosamp1
```

```
## [1] 0.0075
```

## Problem 2

#a

```
data2<-read.csv("Data2.csv")
corr2 <- cor(data2$V1, data2$V2)
corr2
```

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

#b

```
paired2 <- sum(replicate(10000, pairedt(data2)) == TRUE) / 10000
paired2
```

```
## [1] 0.0533
```

#c

```
twosamp2 <- sum(replicate(10000, twosamp(data2)) == TRUE) / 10000
twosamp2
```

```
## [1] 0.107
```

## Problem 3

#a

```
data3 <- read.csv("data3.csv")
corr3 <- cor(data3$V1, data3$V2)
corr3
```

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

#b

```
paired3 <- sum(replicate(10000, pairedt(data3)) == TRUE) / 10000
paired3
```

```
## [1] 0.0511
```

#c

```
twosamp3 <- sum(replicate(10000, twosamp(data3)) == TRUE) / 10000
twosamp3
```

```
## [1] 0.0511
```

## Problem 4

```
correlation <- c(corr1, corr2, corr3)
pairedtest <- c(paired1, paired2, paired3)
twosampletest <- c(twosamp1, twosamp2, twosamp3)
data.frame(correlation, pairedtest, twosampletest)
```

```
##    correlation pairedtest twosampletest
## 1  0.524066001     0.0474        0.0075
## 2 -0.520360008     0.0533        0.1070
## 3  0.002426237     0.0511        0.0511
```

One observation that can be made is the relationship between correlation and the Type 1 error probability for a two sample t test; with more negative correlation values, a larger probability for a Type 1 Error occurs. Another observation is that the Type 1 error for the paired t test was not affected by the sign before the correlation value. The probability values for the paired test stayed consistent around .05.

## Problem 5

#a

```
data5 <- read.csv("data4.csv")
corr5 <- cor(data5$V1, data5$V2)
corr5
```

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

#b

```
paired5 <- sum(replicate(10000, pairedt(data5)) == TRUE) / 10000
paired5
```

```
## [1] 0.0468
```

#c

```
twosamp5 <- sum(replicate(10000, twosamp(data5)) == TRUE) / 10000
twosamp5
```

```
## [1] 0.0139
```

## Problem 6

#a

```
data6 <- read.csv("data5.csv")
corr6 <- cor(data6$V1, data6$V2)
corr6
```

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

#b

```
paired6 <- sum(replicate(10000, pairedt(data6)) == TRUE) / 10000
paired6
```

```
## [1] 0.0717
```

#c

```
twosamp6 <- sum(replicate(10000, twosamp(data6)) == TRUE) / 10000
twosamp6
```

```
## [1] 0.115
```

## Problem 7

#a

```
data7 <- read.csv("data6.csv")
corr7 <- cor(data7$V1, data7$V2)
corr7
```

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

#b

```
paired7 <- sum(replicate(10000, pairedt(data7)) == TRUE) / 10000
paired7
```

```
## [1] 0.0386
```

#c

```
twosamp7 <- sum(replicate(10000, twosamp(data7)) == TRUE) / 10000
twosamp7
```

```
## [1] 0.0292
```

## Problem 8

```
correlation <- c(corr5, corr6, corr7)
pairedtest <- c(paired5, paired6, paired7)
twosampletest <- c(twosamp5, twosamp6, twosamp7)
data.frame(correlation, pairedtest, twosampletest)
```

```
##    correlation pairedtest twosampletest
## 1  0.590640226     0.0468        0.0139
## 2 -0.572119341     0.0717        0.1150
## 3 -0.007297158     0.0386        0.0292
```

One conclusion that can be noted is the increasing probability of a Type 1 error for two sample t tests when correlation values reach progressively more negative values. For paired t tests, the probability of a Type 1 error increases when correlation values change from ~-.007 to ~-.572 but not from ~0.591 to -0.007.