

Homework 8

Will Scheib

```
set.seed(06072000)
```

Problem 1

a

```
data1 <- read.csv("data/data1.csv")
pop.cor.data1 <- cor(data1$V1, data1$V2)
pop.cor.data1
```

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

b

```
ttest.paired <- function(data) {
  samp <- data[sample(nrow(data), 13),]
  t.test(samp[,1], samp[,2], mu=0, alternative="two.sided", paired=TRUE)$p.value < 0.05
}
paired.data1 <- mean(replicate(10000, ttest.paired(data1)))
paired.data1
```

```
## [1] 0.0506
```

c

```
ttest.two.samp <- function(data) {
  samp <- data[sample(nrow(data), 13),]
  t.test(samp[,1], samp[,2], mu=0, alternative="two.sided")$p.value < 0.05
}
two.samp.data1 <- mean(replicate(10000, ttest.two.samp(data1)))
two.samp.data1
```

```
## [1] 0.0077
```

Problem 2

a

```
data2 <- read.csv("data/data2.csv")
pop.cor.data2 <- cor(data2$V1, data2$V2)
pop.cor.data2
```

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

b

```
paired.data2 <- mean(replicate(10000, ttest.paired(data2)))
paired.data2
```

```
## [1] 0.0515
```

c

```
two.samp.data2 <- mean(replicate(10000, ttest.two.samp(data2)))
two.samp.data2
```

```
## [1] 0.1116
```

Problem 3

a

```
data3 <- read.csv("data/data3.csv")
pop.cor.data3 <- cor(data3$V1, data3$V2)
pop.cor.data3
```

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

b

```
paired.data3 <- mean(replicate(10000, ttest.paired(data3)))
paired.data3
```

```
## [1] 0.0464
```

c

```
two.samp.data3 <- mean(replicate(10000, ttest.two.samp(data3)))
two.samp.data3
```

```
## [1] 0.0505
```

Problem 4

```
results.table <- cbind(  
  data1=c(pop.cor=pop.cor.data1, paired=paired.data1, two.samp=two.samp.data1),  
  data2=c(pop.cor=pop.cor.data2, paired=paired.data2, two.samp=two.samp.data2),  
  data3=c(pop.cor=pop.cor.data3, paired=paired.data3, two.samp=two.samp.data3)  
)  
results.table
```

```
##           data1    data2    data3  
## pop.cor  0.524066 -0.52036 0.002426237  
## paired   0.050600  0.05150 0.046400000  
## two.samp 0.007700  0.11160 0.050500000
```

For two data sets that are symmetric and positively correlated, the Type I error for a two sample t-test will be smaller than 0.05. When they are negatively correlated, the Type I error for a two sample t-test will be larger than 0.05. When they are not correlated at all, the Type I error for a two sample t-test will be around 0.05. In all cases, the Type I error for a paired t-test is around 0.05.

Problem 5

a

```
data4 <- read.csv("data/data4.csv")
pop.cor.data4 <- cor(data4$V1, data4$V2)
pop.cor.data4
```

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

b

```
paired.data4 <- mean(replicate(10000, ttest.paired(data4)))
paired.data4
```

```
## [1] 0.0491
```

c

```
two.samp.data4 <- mean(replicate(10000, ttest.two.samp(data4)))
two.samp.data4
```

```
## [1] 0.016
```

Problem 6

a

```
data5 <- read.csv("data/data5.csv")
pop.cor.data5 <- cor(data5$V1, data5$V2)
pop.cor.data5
```

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

b

```
paired.data5 <- mean(replicate(10000, ttest.paired(data5)))
paired.data5
```

```
## [1] 0.0752
```

c

```
two.samp.data5 <- mean(replicate(10000, ttest.two.samp(data5)))
two.samp.data5
```

```
## [1] 0.1111
```

Problem 7

a

```
data6 <- read.csv("data/data6.csv")
pop.cor.data6 <- cor(data6$V1, data6$V2)
pop.cor.data6
```

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

b

```
paired.data6 <- mean(replicate(10000, ttest.paired(data6)))
paired.data6
```

```
## [1] 0.0368
```

c

```
two.samp.data6 <- mean(replicate(10000, ttest.two.samp(data6)))
two.samp.data6
```

```
## [1] 0.0331
```

Problem 8

```
results.table <- cbind(  
  data4=c(pop.cor=pop.cor.data4, paired=paired.data4, two.samp=two.samp.data4),  
  data5=c(pop.cor=pop.cor.data5, paired=paired.data5, two.samp=two.samp.data5),  
  data6=c(pop.cor=pop.cor.data6, paired=paired.data6, two.samp=two.samp.data6)  
)  
results.table
```

##	data4	data5	data6
## pop.cor	0.5906402	-0.5721193	-0.007297158
## paired	0.0491000	0.0752000	0.036800000
## two.samp	0.0160000	0.1111000	0.033100000

For two data sets that are skewed and positively correlated or not correlated, the Type I error for a two sample t-test will be smaller than 0.05. When they are negatively correlated, the Type I error for a two sample t-test will be larger than 0.05.

For two data sets that are skewed and positively correlated, the Type I error for a paired t-test will be around 0.05. When they are negatively correlated, the Type I error for a paired t-test will be larger than 0.05. When they are not correlated at all, the Type I error for a paired t-test will be smaller than 0.05.