

STA305A1

Qizheng Wang

2/4/2020

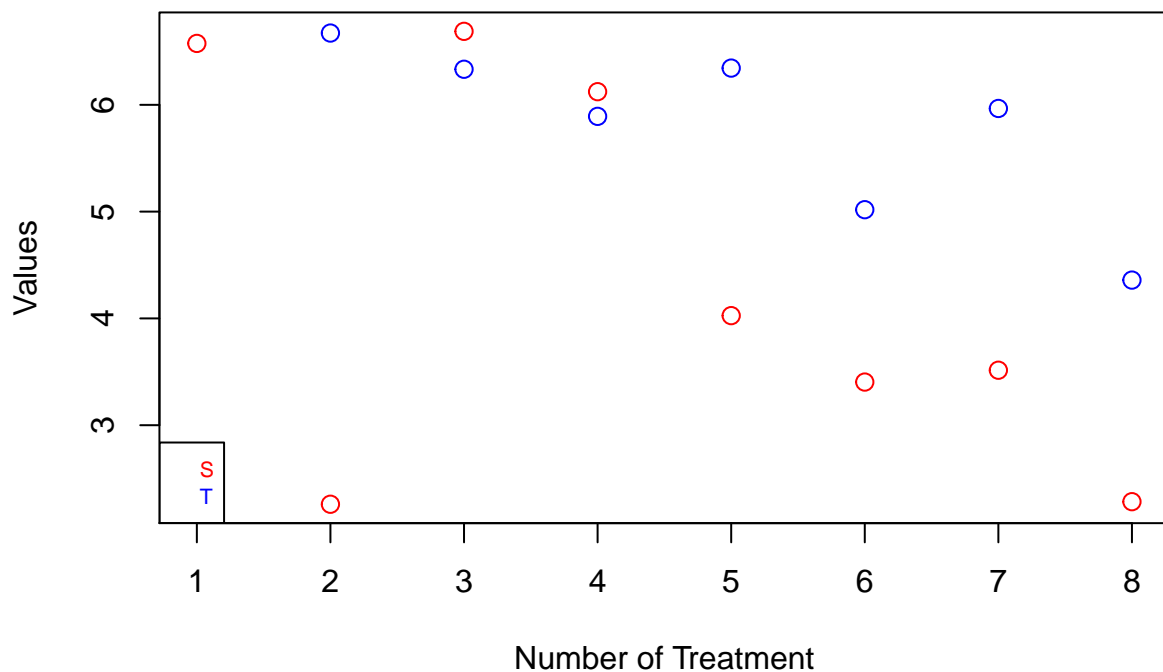
STA305 Assignment 1, student

Student number 1002079877

Question A

```
set.seed(9877)
Number<-c(1:8)
S<-round(rnorm(8,mean=5,sd=2),3)
T<-round(rnorm(8,mean=7,sd=2),3)
diff<-S-T
compD<-data.frame(Number,S,T,diff)

plot(compD$Number,compD$S,col="red",cex=1.2,xlab="Number of Treatment",ylab="Values")
points(compD$Number,compD$T,col="blue",cex=1.2)
legend("bottomleft",legend=c("S","T"),text.col=c("red","blue"),cex = 0.7)
```



compD

##	Number	S	T	diff
## 1	1	6.575	8.244	-1.669
## 2	2	2.260	6.672	-4.412
## 3	3	6.688	6.333	0.355
## 4	4	6.123	5.892	0.231
## 5	5	4.026	6.344	-2.318
## 6	6	3.404	5.018	-1.614
## 7	7	3.515	5.966	-2.451
## 8	8	2.284	4.359	-2.075

(1)-(4)

The seed is 9877.

The first treatment S is 6.575, 2.26, 6.688, 6.123, 4.026, 3.404, 3.515, 2.284.

The second treatment T is 8.244, 6.672, 6.333, 5.892, 6.344, 5.018, 5.966, 4.359.

The observation is compD.

Question B

Question B(i)

```
## Randomized pair design
meandiff<-mean(diff)
ST<-c(S,T)
N <- 2^8
res <- numeric(N)
LR=list(c(-1,1))
trtassign<-expand.grid(rep(LR,8))
for(i in 1:N)
{
  res[i]<-mean(as.numeric(trtassign[i,])*diff)
}
summary(res)
```



```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -1.8906 -0.6408   0.0000   0.0000  0.6408   1.8906
```

```
##Completely randomized design
comp<-c(S,T)
N <- choose(16,8)
res2 <- numeric(N)
observed=mean(S)-mean(T)
index <-combn(1:16,8)
for (i in 1:N){
  res2[i] <- mean(comp[index[,i]])-mean(comp[-index[,i]])
}
```

The first design is the randomized pair design.

There are $2^8=256$ values in this distribution, and the probability of the observed treatment allocation is $\frac{1}{2^8}=0.0039062$.

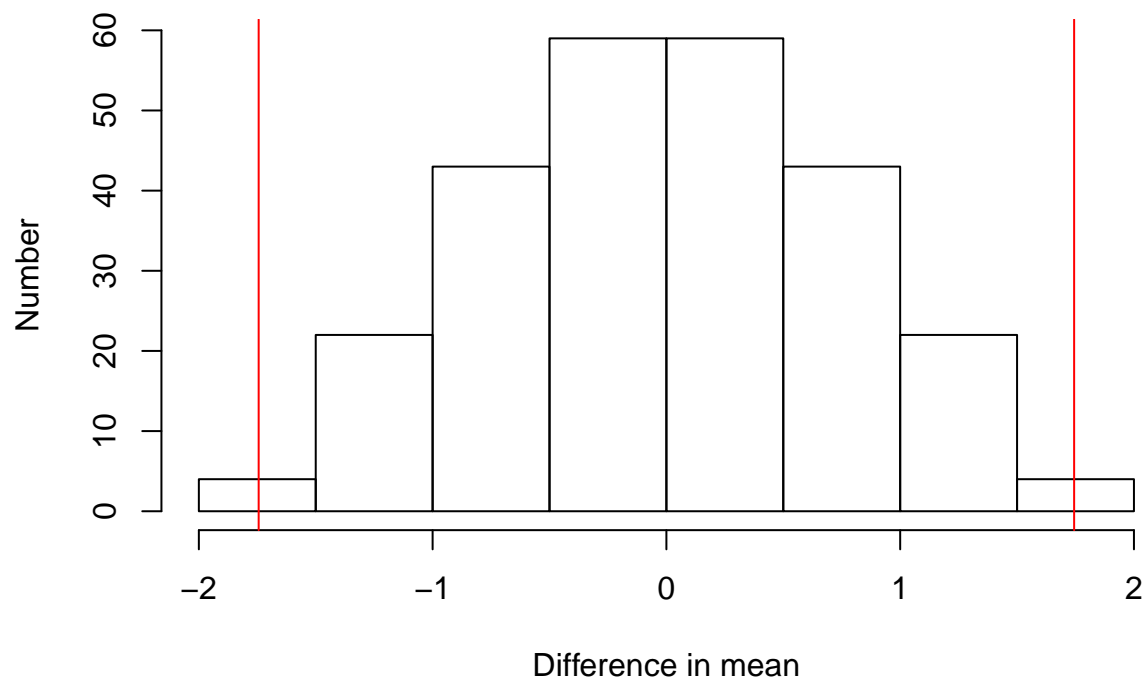
The second design is the Completely randomized design.

There are $\binom{16}{8}=1.287 \times 10^4$ values in this distribution, and the probability of the observed treatment allocation is $\frac{1}{\binom{16}{8}}=7.7700078 \times 10^{-5}$.

Question B(ii)

```
## Randomized paired design
hist(res,xlab='Difference in mean',ylab="Number",main="Randomization Distribution for Paired Comparison")
abline(v = meandiff, col = 'red')
abline(v = -meandiff, col = 'red')
```

Randomization Distribution for Paired Comparison

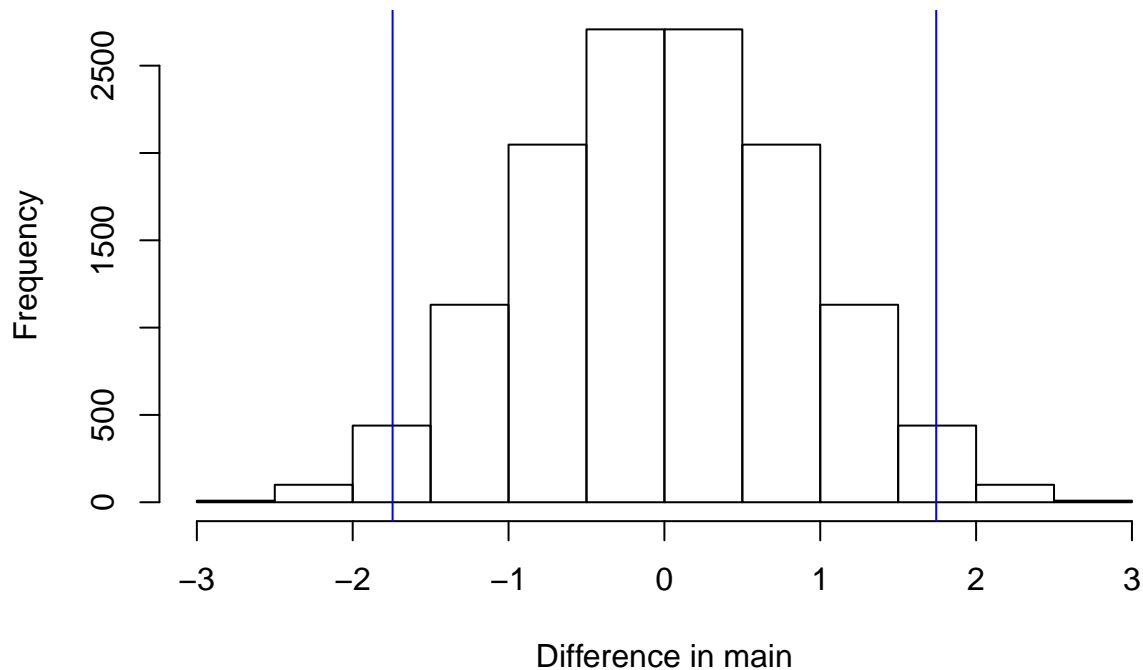


```
tbar<-mean(res)
p_val<-sum(abs(res-tbar)>=abs(meandiff-tbar))/N
round(p_val,4)
```

```
## [1] 6e-04
```

```
## Completely randomized design
hist(res2,xlab = 'Difference in main', main = 'Randomization Distribution for Completely Randomized Des
abline(v=observed, col='blue')
abline(v=-observed, col='blue')
```

Randomization Distribution for Completely Randomized Design



```
tbar2<-mean(res2)
p_val2<-sum(abs(res2-tbar2)>=abs(observed-tbar2))/N
round(p_val2,4)
```

```
## [1] 0.0421
```

By doing a two-sided randomization test, the p-value for the randomized paired design is 0.0421134. The significance level I choose will be 0.05 based on the rule of thumb. Thus 0.0421134 is smaller than 0.05, thus reject the null hypothesis, meaning there are difference in means for two treatments.

Question C

```
#paired t test
t.test(S,T, paired=TRUE)
```

```
##
## Paired t-test
##
## data: S and T
## t = -3.2243, df = 7, p-value = 0.01457
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.0232277 -0.4650223
```

```
## sample estimates:
## mean of the differences
## -1.744125
```

```
#two sample t test
t.test(S,T)
```

```
##
## Welch Two Sample t-test
##
## data: S and T
## t = -2.2654, df = 11.733, p-value = 0.04326
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -3.42584816 -0.06240184
## sample estimates:
## mean of x mean of y
## 4.359375 6.103500
```

Question C(i)

For the completely randomized design, because the population distribution is normal, thus an appropriate test will be two sample t-test.

H_0 : the mean of two sample is equal to zero.

H_a : true difference in means is not equal to 0.

P-value at here is 0.0432617. The p-value here means under the assumption that the population means are equal, the probability that the difference between the sample means is at least as large as what has been observed.

For the paired design, paired t test should be conducted.

H_0 : the mean of two sample is equal to zero.

H_a : true difference in means is not equal to 0.

P-value at here is 0.0145665.

Question C(ii)

Because the distribution is normal and two samples are independent, the normality and independence are satisfied for the two-sample t test. But since the parameters are unknown, the variance may not be the same.

Question C(iii)

The two-sided randomization test has a p-value of 0.0421134, while the p-value of t-test is 0.0432617, it's almost identical, and they are all smaller than $\alpha = 0.05$, thus the results are consistent.