Statistics 3080
Homework 08
Rahul Zalkikar

Problem 01

```
> set.seed(5041998)
>
> K <- 10000
> alpha <- 0.05
> n<-8
> samp.8 <- replicate(K, rnorm(n, mean = 121.8, sd = 34.7))
>
> diff.8 <- samp.8 - 121.8
> sign.8 <- ifelse(diff.8 > 0, 1, 0)
> rank.8 <- apply(abs(diff.8),2,rank)
> ts.8 <- apply((sign.8 * rank.8),2,sum)
> ts.8 <- (ts.8 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.8 <- 2*pnorm(-abs(ts.8))
> Reject_Null.8 <- as.logical(p_value.8<=alpha)
> samp.8_vect <- as.integer(Reject_Null.8)
> samp.8_sum <- sum(samp.8_vect)
> Prop_Reject_Null.8.1 <- samp.8_sum / 10000
> Prop_Reject_Null.8.1
[1] 0.0566
> n<-24
> samp.24 <- replicate(K, rnorm(n, mean = 121.8, sd = 34.7))
>
> diff.24 <- samp.24 - 121.8
> sign.24 <- ifelse(diff.24 > 0, 1, 0)
> rank.24 <- apply(abs(diff.24),2,rank)
> ts.24 <- apply((sign.24 * rank.24),2,sum)
> ts.24 <- (ts.24 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.24 <- 2*pnorm(-abs(ts.24))
> Reject_Null.24 <- as.logical(p_value.24<=alpha)
> samp.24_vect <- as.integer(Reject_Null.24)
> samp.24_sum <- sum(samp.24_vect)
> Prop_Reject_Null.24.1 <- samp.24_sum / 10000
> Prop_Reject_Null.24.1
[1] 0.0465
> n<-48
> samp.48 <- replicate(K, rnorm(n, mean = 121.8, sd = 34.7))
>
> diff.48 <- samp.48 - 121.8
```

```
> sign.48 <- ifelse(diff.48 > 0, 1, 0)
> rank.48 <- apply(abs(diff.48),2,rank)
> ts.48 <- apply((sign.48 * rank.48),2,sum)
> ts.48 <- (ts.48 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.48 <- 2*pnorm(-abs(ts.48))
> Reject_Null.48 <- as.logical(p_value.48<=alpha)
> samp.48_vect <- as.integer(Reject_Null.48)
> samp.48_sum <- sum(samp.48_vect)
> Prop_Reject_Null.48.1 <- samp.48_sum / 10000
> Prop_Reject_Null.48.1
[1] 0.0488
```

Problem 02

```
> n<-8
> samp.8 <- replicate(K, rchisq(n,df=2))
>
> diff.8 <- samp.8 - 1.386294
> sign.8 <- ifelse(diff.8 > 0, 1, 0)
> rank.8 <- apply(abs(diff.8),2,rank)
> ts.8 <- apply((sign.8 * rank.8),2,sum)
> ts.8 <- (ts.8 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.8 <- 2*pnorm(-abs(ts.8))
> Reject_Null.8 <- as.logical(p_value.8<=alpha)
> samp.8_vect <- as.integer(Reject_Null.8)
> samp.8_sum <- sum(samp.8_vect)
> Prop_Reject_Null.8.2 <- samp.8_sum / 10000
> Prop_Reject_Null.8.2
[1] 0.0663
> n<-24
> samp.24 <- replicate(K, rchisq(n,df=2))
>
> diff.24 <- samp.24 - 1.386294
> sign.24 <- ifelse(diff.24 > 0, 1, 0)
> rank.24 <- apply(abs(diff.24),2,rank)
> ts.24 <- apply((sign.24 * rank.24),2,sum)
> ts.24 <- (ts.24 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.24 <- 2*pnorm(-abs(ts.24))
> Reject_Null.24 <- as.logical(p_value.24<=alpha)
> samp.24_vect <- as.integer(Reject_Null.24)
> samp.24_sum <- sum(samp.24_vect)
> Prop_Reject_Null.24.2 <- samp.24_sum / 10000
> Prop_Reject_Null.24.2
[1] 0.1162
> n<-48
```

```
> samp.48 <- replicate(K, rchisq(n,df=2))
>
> diff.48 <- samp.48 - 1.386294
> sign.48 <- ifelse(diff.48 > 0, 1, 0)
> rank.48 <- apply(abs(diff.48),2,rank)
> ts.48 <- apply((sign.48 * rank.48),2,sum)
> ts.48 <- (ts.48 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.48 <- 2*pnorm(-abs(ts.48))
> Reject_Null.48 <- as.logical(p_value.48<=alpha)
> samp.48_vect <- as.integer(Reject_Null.48)
> samp.48_sum <- sum(samp.48_vect)
> Prop_Reject_Null.48.2 <- samp.48_sum / 10000
> Prop_Reject_Null.48.2
[1] 0.2001
```

Problem 03

```
> one.samp.t.Prop.Reject_Null.8 <- 0.0516
> one.samp.t.Prop.Reject_Null.24 <- 0.0511
> one.samp.t.Prop.Reject_Null.48 <- 0.0513
>
> #Comparisons
> Prop_Reject_Null.8.1 - one.samp.t.Prop.Reject_Null.8
[1] 0.005
> Prop_Reject_Null.24.1 - one.samp.t.Prop.Reject_Null.24
[1] -0.0046
> Prop_Reject_Null.48.1 - one.samp.t.Prop.Reject_Null.48
[1] -0.0025
> Prop_Reject_Null.8.2 - one.samp.t.Prop.Reject_Null.8
[1] 0.0147
> Prop_Reject_Null.24.2 - one.samp.t.Prop.Reject_Null.24
[1] 0.0651
> Prop_Reject_Null.48.2 - one.samp.t.Prop.Reject_Null.48
[1] 0.1488
> sum_table <- matrix(c(Prop_Reject_Null.8.1, Prop_Reject_Null.8.2,
+                       one.samp.t.Prop.Reject_Null.8, Prop_Reject_Null.24.1,
+                       Prop_Reject_Null.24.2, one.samp.t.Prop.Reject_Null.24,
+                       Prop_Reject_Null.48.1, Prop_Reject_Null.48.2,
+                       one.samp.t.Prop.Reject_Null.48),ncol=3,
+                     byrow=TRUE)
> colnames(sum_table) <- c("Prob1Type1","Prob2Type1","HW7Type1")
> rownames(sum_table) <- c("Sample Size 8", "Sample Size 24", "Sample Size 48")
> sum_table <- as.table(sum_table)
> sum_table
               Prob1Type1 Prob2Type1 HW7Type1
```

```
Sample Size 8      0.0566     0.0663   0.0516
Sample Size 24     0.0465     0.1162   0.0511
Sample Size 48     0.0488     0.2001   0.0513
> #This leads to the conclusion that the sign rank test
> #is more accurate for the chi square distribution,
> #since the chisq distribution is skewed (because
> #the median is a better estimate for predicting
> #the center of a skewed  distribution). For the
> #normal distribution both tests are equally accurate.
```

Problem 04

```
> library(pwr)

Warning:  package 'pwr' was built under R version 3.4.4

> K <- 10000
> alpha <- 0.05
> n<-8
> samp.8 <- replicate(K, rnorm(n, mean = 121.8, sd = 34.7))
>
> diff.8 <- samp.8 - 115
> sign.8 <- ifelse(diff.8 > 0, 1, 0)
> rank.8 <- apply(abs(diff.8),2,rank)
> ts.8 <- apply((sign.8 * rank.8),2,sum)
> ts.8 <- (ts.8 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.8 <- 2*pnorm(-abs(ts.8))
> Power_Bool <- as.logical(p_value.8<alpha)
> samp.8_vect <- as.integer(Power_Bool)
>
> samp.8_sum <- sum(samp.8_vect)
> Power.8.4 <- samp.8_sum / 10000
> Power.8.4
[1] 0.0873
> n<-24
> samp.24 <- replicate(K, rnorm(n, mean = 121.8, sd = 34.7))
>
> diff.24 <- samp.24 - 115
> sign.24 <- ifelse(diff.24 > 0, 1, 0)
> rank.24 <- apply(abs(diff.24),2,rank)
> ts.24 <- apply((sign.24 * rank.24),2,sum)
> ts.24 <- (ts.24 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.24 <- 2*pnorm(-abs(ts.24))
> Power_Bool <- as.logical(p_value.24<alpha)
> samp.24_vect <- as.integer(Power_Bool)
```

```
>
> samp.24_sum <- sum(samp.24_vect)
> Power.24.4 <- samp.24_sum / 10000
> Power.24.4
[1] 0.1437
> n<-48
> samp.48 <- replicate(K, rnorm(n, mean = 121.8, sd = 34.7))
>
> diff.48 <- samp.48 - 115
> sign.48 <- ifelse(diff.48 > 0, 1, 0)
> rank.48 <- apply(abs(diff.48),2,rank)
> ts.48 <- apply((sign.48 * rank.48),2,sum)
> ts.48 <- (ts.48 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.48 <- 2*pnorm(-abs(ts.48))
> Power_Bool <- as.logical(p_value.48<alpha)
> samp.48_vect <- as.integer(Power_Bool)
>
> samp.48_sum <- sum(samp.48_vect)
> Power.48.4 <- samp.48_sum / 10000
> Power.48.4
[1] 0.2555
```

Problem 05

```
> K <- 10000
> alpha <- 0.05
> n<-8
> samp.8 <- replicate(K, rnorm(n, mean = 121.8, sd = 34.7))
>
> diff.8 <- samp.8 - 115
> sign.8 <- ifelse(diff.8 > 0, 1, 0)
> rank.8 <- apply(abs(diff.8),2,rank)
> ts.8 <- apply((sign.8 * rank.8),2,sum)
> ts.8 <- (ts.8 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.8 <- 2*pt(-abs(ts.8), df=n-1)
> Power_Bool <- as.logical(p_value.8<alpha)
> samp.8_vect <- as.integer(Power_Bool)
>
> samp.8_sum <- sum(samp.8_vect)
> Power.8.5 <- samp.8_sum / 10000
> Power.8.5
[1] 0.0265
> n<-24
> samp.24 <- replicate(K, rnorm(n, mean = 121.8, sd = 34.7))
>
```

```
> diff.24 <- samp.24 - 115
> sign.24 <- ifelse(diff.24 > 0, 1, 0)
> rank.24 <- apply(abs(diff.24),2,rank)
> ts.24 <- apply((sign.24 * rank.24),2,sum)
> ts.24 <- (ts.24 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.24 <- 2*pt(-abs(ts.24), df=n-1)
> Power_Bool <- as.logical(p_value.24<alpha)
> samp.24_vect <- as.integer(Power_Bool)
>
> samp.24_sum <- sum(samp.24_vect)
> Power.24.5 <- samp.24_sum / 10000
> Power.24.5
[1] 0.1173
> n<-48
> samp.48 <- replicate(K, rnorm(n, mean = 121.8, sd = 34.7))
>
> diff.48 <- samp.48 - 115
> sign.48 <- ifelse(diff.48 > 0, 1, 0)
> rank.48 <- apply(abs(diff.48),2,rank)
> ts.48 <- apply((sign.48 * rank.48),2,sum)
> ts.48 <- (ts.48 - ((n*(n+1))/4))/(sqrt((n*(n+1)*((2*n)+1))/24))
> p_value.48 <- 2*pt(-abs(ts.48), df=n-1)
> Power_Bool <- as.logical(p_value.48<alpha)
> samp.48_vect <- as.integer(Power_Bool)
>
> samp.48_sum <- sum(samp.48_vect)
> Power.48.5 <- samp.48_sum / 10000
> Power.48.5
[1] 0.2348
```

Problem 06

```
> sum_table <- matrix(c(Prop_Reject_Null.8.1, Prop_Reject_Null.8.2,
+                       one.samp.t.Prop.Reject_Null.8, Power.8.4, Power.8.5,
+                       Prop_Reject_Null.24.1, Prop_Reject_Null.24.2,
+                       one.samp.t.Prop.Reject_Null.24, Power.24.4, Power.24.5,
+                       Prop_Reject_Null.48.1, Prop_Reject_Null.48.2,
+                       one.samp.t.Prop.Reject_Null.48, Power.48.4, Power.48.5),ncol=5,
+                     byrow=TRUE)
> colnames(sum_table) <- c("Prob1Type1","Prob2Type1","HW7Type1", "Prob4Pwr", "Prob5Pwr")
> rownames(sum_table) <- c("Sample Size 8", "Sample Size 24", "Sample Size 48")
> sum_table <- as.table(sum_table)
> sum_table
               Prob1Type1 Prob2Type1 HW7Type1 Prob4Pwr Prob5Pwr
Sample Size 8      0.0566     0.0663   0.0516   0.0873   0.0265
```

```
Sample Size 24     0.0465     0.1162   0.0511   0.1437   0.1173
Sample Size 48     0.0488     0.2001   0.0513   0.2555   0.2348
> #No, this information does not change my conclusions.
```

Problem 07

```
> #They are not comparable because the chi square
> #distribution is skewed, making the calculated power
> #values of the corresponding one-samp t-test less
> #accurate.
```

References:

- Hypothesis testing in R.R
- Simulations in R.R
- http://stat.ethz.ch/R-manual/R-devel/library/stats/html/Chisquare.html