

# Homework 11

*Max Ryoo*

## Problem 1

### Part A

```
set.seed(12181998)
k <- 10000
population_mean <- 62.9
population_sd <- 13.3
for(i in c(7,28,49)) {
  samp <- replicate(k, rnorm(i, population_mean, population_sd))
  gre <- samp > 62.9
  vec <- apply(gre, 2, sum)
  result <- lapply(vec, function(x) binom.test(x, i, alternative="two.sided"))
  pval <- sapply(1:k, function(x) result[[x]]$p.value)
  result1 <- ifelse(pval <= 0.05, "R", "F")
  len <- length(which(result1 == "R"))/k
  assign(paste("A", i, sep="."), len)
}
print(A.7)
## [1] 0.0165
print(A.28)
## [1] 0.0381
print(A.49)
## [1] 0.0469
```

I constructed a for loop to conduct a normal distribution of sample sizes of 7,28,49, which were the index (i) values. To do a sign test I used the binom.test function then found which proportion of the p-value is smaller than 0.05

### Part B

```
for(i in c(7,28,49)) {
  samp <- replicate(k, rnorm(i, population_mean, population_sd))
  r <- lapply(1:k, function(x)
    wilcox.test(samp[,x], mu=62.9, alternative = "two.sided"))
  pval <- lapply(1:k, function(x) r[[x]]$p.value)
```

```

    result <- ifelse(pval <= 0.05, "R", "F")
    len <- length(which(result == "R"))/k
    assign(paste("B", i, sep="."), len)
  }
print(B.7)
## [1] 0.0482
print(B.28)
## [1] 0.049
print(B.49)
## [1] 0.051

```

I constructed the same for loop for this problem. In order to do a sign-rank test I used the wilcox.test function with a mu of 62.9 given. Then i used teh lapply function to find the p values. I then found what the proportion was smaller than 0.05 to find the type 1 error

## Problem 2

```

table2 <- rbind("Two-tailed sign test" = c(A.7, A.28, A.49),
               "Sign rank test" = c(B.7, B.28, B.49),
               "Problem 2 HW7" = c(0.0544,0.0487,0.0520))
colnames(table2) <- c("Sample Size 7", "Sample Size 28", "Sample Size 49")
print(table2)

```

##	Sample Size 7	Sample Size 28	Sample Size 49
## Two-tailed sign test	0.0165	0.0381	0.0469
## Sign rank test	0.0482	0.0490	0.0510
## Problem 2 HW7	0.0544	0.0487	0.0520

It seems that for the two tailed sign test as sample size gets bigger it gets closer to the 0.05 the prediction of proportion of type 1 error. However fof sign rank test it seems to be consistently near 0.05. The results from hw 7 problem 2 also seemed to be around the expected of 0.05 for all the number of n

## Problem 3

### Part A

```

for(i in c(7,28,49)) {
  samp <- replicate(k, rnorm(i, population_mean, population_sd))
  gre <- samp > 59
  vec <- apply(gre, 2, sum)
  result <- lapply(vec, function(x) binom.test(x, i, alternative="two.sided"))
}

```

```

    pval <- sapply(1:k, function(x) result[[x]]$p.value)
    result1 <- ifelse(pval <= 0.05, "R", "F")
    len <- length(which(result1 == "R"))/k
    assign(paste("TA", i, sep="."), len)
  }
print(TA.7)
## [1] 0.0351
print(TA.28)
## [1] 0.1826
print(TA.49)
## [1] 0.3478

```

I took the similar approach as part a of problem one. I calculated the power by finding which proportion of p-values is less than 0.05, for the median of 59.

## Part B

```

for(i in c(7,28,49)) {
  samp <- replicate(k, rnorm(i, population_mean, population_sd))
  r <- lapply(1:k, function(x) t.test(samp[,x], mu=59, alternative = "two.sided"))
  pval <- lapply(1:k, function(x) r[[x]]$p.value)
  result <- ifelse(pval <= 0.05, "R", "F")
  len <- length(which(result == "R"))/k
  assign(paste("TC", i, sep="."), len)
}
print(TC.7)
## [1] 0.1032
print(TC.28)
## [1] 0.3213
print(TC.49)
## [1] 0.5243

```

I used a similar approach to part b in problem 1 to solve this problem. Instead of 62.9 I used it with a mu of 59 and calculated which proportion of p-value is smaller than 0.05.

## Part C

```

for(i in c(7,28,49)) {
  samp <- replicate(k, rnorm(i, population_mean, population_sd))

```

```

r <- lapply(1:k, function(x) wilcox.test(samp[,x], mu=59, alternative = "two.sided"))
pval <- lapply(1:k, function(x) r[[x]]$p.value)
result <- ifelse(pval <= 0.05, "R", "F")
len <- length(which(result == "R"))/k
assign(paste("TB", i, sep="."), len)
}
print(TB.7)
## [1] 0.0965
print(TB.28)
## [1] 0.2992
print(TB.49)
## [1] 0.507

```

I used a similar approach to part b in problem 3. Instead of the wilcox.test I used a t.test function to conduct a two tailed t-test. I then calculated the proportion of p-values lower than 0.05.

## Problem 4

```

top <- c("two-tailed sign test", "sign rank test")
table4 <- rbind("Two-tailed sign test" = c(TA.7, TA.28, TA.49),
               "Sign rank test" = c(TB.7, TB.28, TB.49),
               "T-test" = c(TC.7, TC.28, TC.49))
colnames(table4) <- c("Sample Size 7", "Sample Size 28", "Sample Size 49")
print(table4)

##                Sample Size 7 Sample Size 28 Sample Size 49
## Two-tailed sign test      0.0351      0.1826      0.3478
## Sign rank test           0.0965      0.2992      0.5070
## T-test                   0.1032      0.3213      0.5243

```

We can conclude that for all different test as sample size increased the power increased as well. Since power is the prob of correctly rejecting null we can say that the bigger the sample size the test will better predict that the null hypothesis is false.

## Problem 5

### Part A

```

for(i in c(7,28,49)) {
  samp <- replicate(k, rchisq(i, df=2))
  gre <- samp > 1.386294
}

```

```

vec <- apply(gre, 2, sum)
result <- lapply(vec, function(x) binom.test(x, i, alternative="two.sided"))
pval <- sapply(1:k, function(x) result[[x]]$p.value)
result1 <- ifelse(pval <= 0.05, "R", "F")
len <- length(which(result1 == "R"))/k
assign(paste("FA", i, sep="."), len)
}
print(FA.7)
## [1] 0.0151
print(FA.28)
## [1] 0.0369
print(FA.49)
## [1] 0.0436

```

I used the same approach as Problem 1 Part A, but this time I did `rchisq(i, df=2)`. This is a chi-squared distribution with a degree of freedom of 2.

## Part B

```

for(i in c(7,28,49)) {
  samp <- replicate(k, rchisq(i, df=2))
  r <- lapply(1:k, function(x)
    wilcox.test(samp[,x], mu=1.386294, alternative = "two.sided"))
  pval <- lapply(1:k, function(x) r[[x]]$p.value)
  result <- ifelse(pval <= 0.05, "R", "F")
  len <- length(which(result == "R"))/k
  assign(paste("FB", i, sep="."), len)
}
print(FB.7)
## [1] 0.0531
print(FB.28)
## [1] 0.1247
print(FB.49)
## [1] 0.2035

```

I did the same thing as part b from problem 1 instead with a `rchisq` distribution with a degree of freedom of 2.

## Problem 6

```
table6 <- rbind("Two-tailed sign test" = c(FA.7, FA.28, FA.49),
               "Sign rank test" = c(FB.7, FB.28, FB.49))
colnames(table6) <- c("Sample Size 7", "Sample Size 28", "Sample Size 49")
print(table6)
```

##	Sample Size 7	Sample Size 28	Sample Size 49
## Two-tailed sign test	0.0151	0.0369	0.0436
## Sign rank test	0.0531	0.1247	0.2035

As shown in the table we can see that as the sample size goes up for both tests the proportion of type 1 error is increasing. However, we can see that the Sign rank test increases at a much bigger rate than the two-tailed sign test. Therefore, I think I would not use the sign rank test for a chi squared distribution.

## Problem 7

### Part A

```
for(i in c(7,28,49)) {
  samp <- replicate(k, rchisq(i, df=2))
  gre <- samp > 0.5
  vec <- apply(gre, 2, sum)
  result <- lapply(vec, function(x) binom.test(x, i, alternative="two.sided"))
  pval <- sapply(1:k, function(x) result[[x]]$p.value)
  result1 <- ifelse(pval <= 0.05, "R", "F")
  len <- length(which(result1 == "R"))/k
  assign(paste("SA", i, sep="."), len)
}
print(SA.7)
## [1] 0.1773
print(SA.28)
## [1] 0.8506
print(SA.49)
## [1] 0.9855
```

This procedure is the same as previous problems part A. I simply changed the median to 0.5 as the problem suggests.

## Part B

```
for(i in c(7,28,49)) {
  samp <- replicate(k, rchisq(i, df=2))
  r <- lapply(1:k, function(x)
    wilcox.test(samp[,x], mu=0.5, alternative = "two.sided"))
  pval <- lapply(1:k, function(x) r[[x]]$p.value)
  result <- ifelse(pval <= 0.05, "R", "F")
  len <- length(which(result == "R"))/k
  assign(paste("SB", i, sep="."), len)
}
print(SB.7)
## [1] 0.4579
print(SB.28)
## [1] 0.9964
print(SB.49)
## [1] 1
```

I redid part b from problem 5 with a different median of 0.5

## Problem 8

```
table8 <- rbind("Two-tailed sign test" = c(SA.7, SA.28, SA.49),
               "Sign rank test" = c(SB.7, SB.28, SB.49))
colnames(table8) <- c("Sample Size 7", "Sample Size 28", "Sample Size 49")
print(table8)

##               Sample Size 7 Sample Size 28 Sample Size 49
## Two-tailed sign test      0.1773      0.8506      0.9855
## Sign rank test           0.4579      0.9964      1.0000
```

The table shows that both tests will have greater power as the sample size gets bigger. Both end up at a really high rate while at a small sample size it seems that the sign rank test is higher power.

## References

1. <https://www.statisticshowto.datasciencecentral.com/standardized-test-statistic/>