

DSC 462: Assignment 6

Shouman Das

December 13, 2017

Q1

D_i	$ D_i $	Rank	+/-
-5.8	5.8	5	-
-6.5	6.5	6	-
+4.1	4.1	2.5	+
-7.7	7.7	7	-
-4.1	4.1	2.5	-
+2.9	2.9	1	+
+5.1	5.1	4	+

Now our hypothesis:

$$H_o : \mu_{\tilde{D}} = 0$$

$$H_a : \mu_{\tilde{D}} < 0$$

with $\alpha = 0.05$. From the above table, we have $T_+ = 1 + 2.5 + 4 = 7.5$ and $T_- = 2.5 + 5 + 6 + 7 + 20.5 = 20.5$. So our test statistic $T_{obs} = \min(T_+, T_-) = 7.5$. And $\alpha_{obs} = P(T \leq 7.5) = 0.1484$. So we cannot reject H_o with $\alpha = 0.05$.

Now we use normal approximation. The mean and standard deviation of the positive or negative rank sums are

$$\mu_T = \frac{n(n+1)}{4} = 14 \text{ and } \sigma_T = \sqrt{\frac{n(n+1)(2n+1)}{24}} = 5.91608$$

This gives z-score

$$Z = \frac{T_{obs} - \mu_T}{\sigma_T} = (7.5 - 14)/5.91608 = -1.0987$$

with p-value $\alpha_{obs} = P(Z \leq -1.0987) = 0.1378$. So we cannot reject H_o . Verify the answer:

```
before = c(91.4, 101.1, 97.7, 95.5, 100.7, 102.6, 84)
after = c(85.6, 94.6, 101.8, 87.8, 96.6, 105.5, 89)
# exact signed rank test
wilcox.test(after-before, exact = T, alternative = "less")
```

```
##
## Wilcoxon signed rank test
##
## data: after - before
## V = 7, p-value = 0.1484
## alternative hypothesis: true location is less than 0
```

```
# normal approximation
wilcox.test(after-before, exact = F, alternative = "less", correct = F)
```

```
##  
## Wilcoxon signed rank test  
##  
## data: after - before  
## V = 7, p-value = 0.1184  
## alternative hypothesis: true location is less than 0
```

Q2

Calculate the rank sum:

```
x1 = c(31.7, 20.9, 23.2, 30.2, 34.4, 31.0, 26.9)
x2 = c(35, 40.8, 39.8, 30.2, 40.9, 32.7, 38.9, 35.3, 35.2, 38.9)
df = data.frame(val = c(x1,x2), sample = c(rep(1,7),rep(2,10)))
aggregate(rank(val)~sample, df, sum)
```

```
##      sample rank(val)
## 1         1      32.5
## 2         2     120.5
```

So $T_{obs} = \min(T_1, T_2) = T_1 = 32.5$. The mean and standard deviation of T_1 are

$$\mu_1 = n_1(n_1 + n_2 + 1)/2 = 63, \sigma_W = \sqrt{7.10(18)/12} = 10.24695$$

Z-score = $(32.5 - 63)/10.24695 = -2.976495$ with p-value = $2P(Z < -2.976495) = 0.0029$. So we can reject H_o .

Verifying the answer:

```
wilcox.test(x1,x2, paired = F, correct = F, exact = F)
```

```
##
## Wilcoxon rank sum test
##
## data:  x1 and x2
## W = 4.5, p-value = 0.002881
## alternative hypothesis: true location shift is not equal to 0
```

So we see that the p-value is less than 0.05 so we can reject H_o . If there are ties, then `wilcox.test()` uses normal approximation.

Q3

We will use R as a calculator.

```
obs = c(8,19,31,66)
prob = c(1/15,2/15,4/15,8/15)
exp_count = sum(obs)*prob
# without Yate's correction
chi.sq = sum((obs-exp_count)^2/exp_count)
chi.sq
```

```
## [1] 0.5060484
```

```
qchisq(df = 3, p = 0.95)
```

```
## [1] 7.814728
```

```
# p-value
1-pchisq(df=3, .506484)
```

```
## [1] 0.9174646
```

Since $X^2 \leq \chi_{3,0.05}^2$, we can not reject H_o .

Verify:

```
chisq.test(obs, p = prob)
```

```
##
## Chi-squared test for given probabilities
##
## data:  obs
## X-squared = 0.50605, df = 3, p-value = 0.9176
```

Q4

Note that the degree of freedom is $(2 - 1)(2 - 1) = 1$. Now Use R as an calculator:

```
Ob = matrix(c(2597, 3128, 425, 350), nrow = 2, ncol = 2)
rowtotal = rbind(3022, 3478)
coltotal = cbind(5775, 775)
N = 6500
E = rowtotal %*% coltotal/N
chi.sq = sum((abs(Ob-E)-.5)^2/E)
chi.sq
```

```
## [1] 24.66841
```

```
qchisq(df = 1, p = 1-0.05)
```

```
## [1] 3.841459
```

Since $X^2 > \chi_{1,0.05}^2$, we cannot reject H_o .

Veryfying:

```
chisq.test(Ob, correct = T)
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  Ob
## X-squared = 24.261, df = 1, p-value = 8.414e-07
```

This implies that row and columns are independent so there is not enough evidence that males are more prone to be left-handed.

Q5

(a) Let $x = 1:n$ and $y = \text{sample}(n)$. Since $\text{sample}(n)$ function is random permutation this x, y pair will give us $\rho = 0$.

(b), (c), (d)

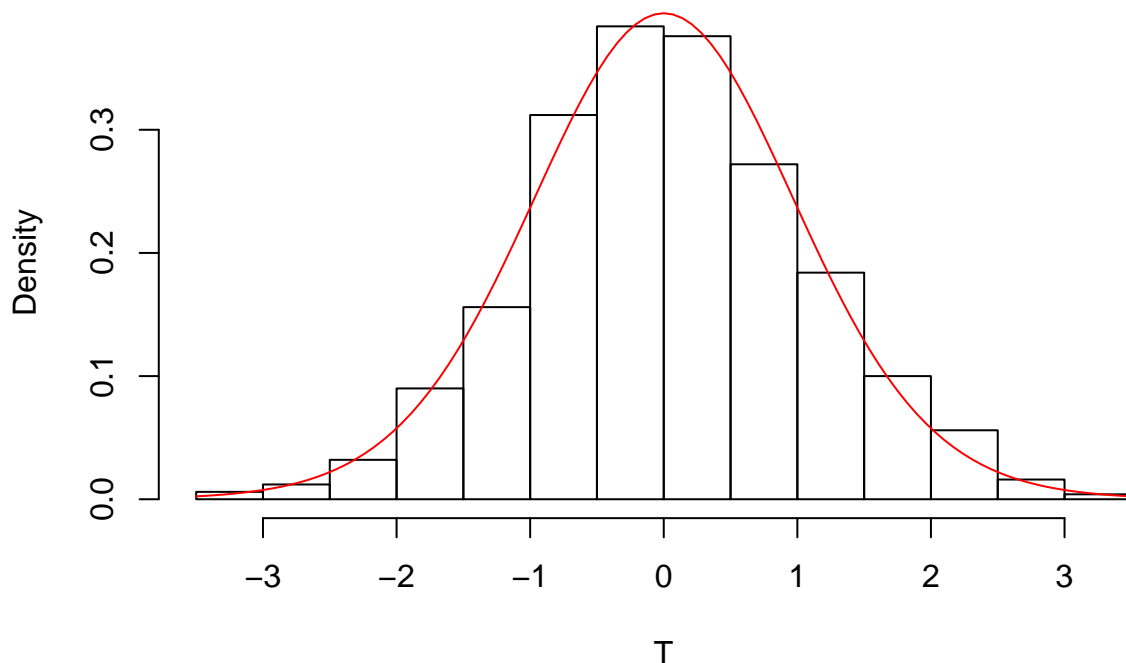
```
set.seed(345471)
n = 25
sim = function(size){
  x = 1:n
  res = 1:size
  for(i in 1:size){
    res[i] = cor(x, sample(n))
  }
  res
}
# simulated value of correlation
r = sim(1000)

# transformation for t-value
T = r/sqrt((1-r^2)/(23))

# histogram
hist(T, probability = TRUE, breaks = 20)

x = seq(-3.5, 3.5, 0.1)
#plot(x, dt(x, df = 23), type = 'l', col = 'blue')
curve(dt(x, df = 23), add = TRUE, col = 'red')
```

Histogram of T



We can

see that our density plot is very similar to the shape of the histograms.