Homework 2

Non-Parametric Statistics

Langdon Holmes

Question 1

```r
urban <- c(1,0,1,1,0,0,1,1,1,8,1,1,1,0,1,1,2)
rural <- c(3,2,1,1,2,1,3,2,2,2,2,5,1,4,1,1,1,1,6,2,2,2,1,1)

get_ranks <- function(a, b){
  m <- length(a)
  n <- length(b)
  combined_ranks <- rank(c(a, b), ties.method="average")
  a_ranks <- combined_ranks[1:m]
  b_ranks <- combined_ranks[(m+1):(m+n)]
  return(list(a_ranks, b_ranks))
}

R <- get_ranks(urban, rural)
R1 <- sapply(R[1], sum)
R2 <- sapply(R[2], sum)

cat("R1:", R1, "\nR2:", R2)
```

```
R1: 246.5
R2: 614.5
```

The critical values for a two-tailed Wilcoxon Rank Sum test with sample sizes 17 and 24 are $W \leq 282$ or $W \geq 432$.

Based on the above calculations, there is sufficient evidence to reject the null hypothesis that the means are equal. Since R_1 (corresponding to urban sibling ranks) is below the critical value, rural folk are more likely to have more siblings than urbanites.

```r
est_var <- function(x) {
  return(sum((x - mean(x))^2)/(length(x)-1))
}

pool_var <- function(a, b) {
  numerator <- (length(a)-1)*est_var(a) + (length(b)-1)*est_var(b)
  denominator <- length(a) + length(b)-2
  return(numerator/denominator)
}

student_t_test <- function(a, b) {
  pooled_var_est <- pool_var(a,b)
  t = (mean(a) - mean(b)) / sqrt(pooled_var_est*(1/length(a) + 1/length(b)))
  return(t)
}

cat(student_t_test(urban, rural), "not less than", qt(p=.05, df=39), "\n")
```

```
-1.638335 not less than -1.684875
```

```r
#equivalent to
t.test(urban, rural, var.equal=TRUE, "less")
```

```

    Two Sample t-test

data:  urban and rural
t = -1.6383, df = 39, p-value = 0.0547
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
       -Inf 0.02290674
sample estimates:
mean of x mean of y
 1.235294  2.041667
```

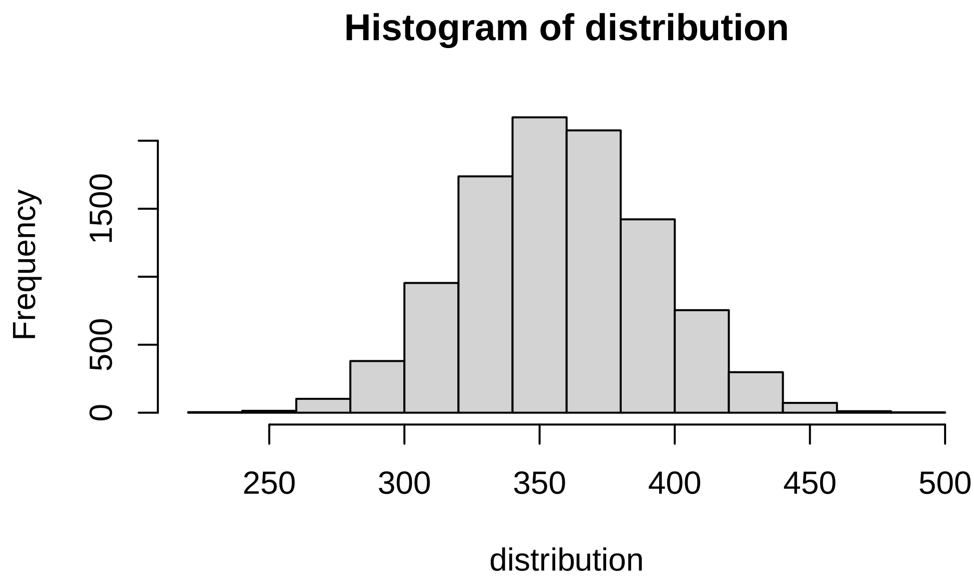
Assuming equal variance between the two samples, $t_{.05,39} = 1.64$, which is not extreme enough to reject the null hypothesis that the means are equal between the two groups. This is true for both one-sided and two-sided tests.

Question 2

```r
library("combinat")
```

```
Attaching package: 'combinat'

The following object is masked from 'package:utils':

    combn
```

```r
permutation.test <- function(a, b, n_permutations, obs_value){
  distribution=c()

  m <- length(a)

  combined_ranks <- rank(c(a, b), ties.method="average")

  for(i in 1:n_permutations){
    distribution[i]=sum(sample(combined_ranks, size=m, replace=FALSE))
  }

  result = quantile(distribution, c(.05,.95))

  hist(distribution)

  p_value = sum(distribution<obs_value)/n_permutations

  return(list(result, p_value))
}

permutation.test(urban, rural, 10000, 246.5)
```




```
[[1]]
     5%     95%
300.475 415.500

[[2]]
[1] 6e-04
```

Using a permutation test, I found that there is a very low probably of finding such a low sum of ranks for urbanites under H_0. This analysis is closer to the rank sum test than the T-test for this data; however, the permutation test appears to have higher power.

Question 3 Large Sample Approximation

asd

Question 4 Permutation and T-test

```
C1 <- c(1.0,5.3,1.2,3.9,8.3,6.3,2.2,9.8,2.8,2.6)
C1 <- c(5.1,6.0,8.0,8.2,7.3,4.4,7.4,7.5,6.4,4.5,8.9)
```

Question 5 Equality of Variance F-test, Siegel-Tukey, and Higgins

Question 6 F-test and one-way ANOVA

```
G1 <- c(2.9736, 0.9448, 1.6394, 0.0389, 1.2958)
G2 <- c(0.7681, 0.8027, 0.2156, 0.0740, 1.5076)
G3 <- c(4.8249, 2.2516, 1.5609, 2.0452, 1.0959)
```

Question 7 Kruskal-Wallis