df\_students<-read.csv('df\_students.csv')

Difference Between U-Test and T-Test 1. Definition & Purpose T-Test: A parametric test used to compare the means of two groups to determine if they are significantly different. It assumes that the data follows a normal distribution. U-Test (Mann-Whitney U Test): A non-parametric test used to compare differences between two independent groups when the data does not necessarily follow a normal distribution. 2. Underlying Principle T-Test: Based on the assumption that the sample data is normally distributed and uses means and variances to assess differences between groups. U-Test: Ranks all observations from both groups together and compares the sum of ranks instead of actual data values.

Since U-test and T-test only focus on binary data, so some category data like ‘Age Arrive USA’, continous data ‘quiz time(categorized by 10 minutes)’ don’t work.

1. quiz time(categorized by 10 minutes) //We didn’t test because it is not binary variable.//
2. English Proficiency
3. Race
4. Home Language

#### Mann-Whitney U Test

1. Accomodations
2. Age Arrive USA //We didn’t test because it is not binary variable.//
3. Sex
4. Born in USA

In T and U test, We only consider whether P-value is smaller than 0.05 or not. If P-value is greater than 0.05, these two different groups have no such a significant difference in the persepective of response variable.

T-test 1.English Proiciency

group1 <- df\_students$`Total.Score`[df\_students$`English.Proficiency` == 0]  
group2 <- df\_students$`Total.Score`[df\_students$`English.Proficiency` == 1]  
  
t\_test\_result <- t.test(group1, group2, var.equal = TRUE)   
print(t\_test\_result)

##   
## Two Sample t-test  
##   
## data: group1 and group2  
## t = 1.0918, df = 29, p-value = 0.2839  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.06362701 0.20935350  
## sample estimates:  
## mean of x mean of y   
## 0.4805556 0.4076923

2.Race

group1 <- df\_students$`Total.Score`[df\_students$`Race.Ethnicity` == 0]  
group2 <- df\_students$`Total.Score`[df\_students$`Race.Ethnicity` == 1]  
  
t\_Race <- t.test(group1, group2, var.equal = TRUE)   
print(t\_Race)

##   
## Two Sample t-test  
##   
## data: group1 and group2  
## t = -0.097293, df = 29, p-value = 0.9232  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.1458685 0.1326206  
## sample estimates:  
## mean of x mean of y   
## 0.4461538 0.4527778

3.Home Language

group1 <- df\_students$`Total.Score`[df\_students$`Home.Language` == 0]  
group2 <- df\_students$`Total.Score`[df\_students$`Home.Language` == 1]  
  
t\_HomeLanguage <- t.test(group1, group2, var.equal = TRUE)   
print(t\_HomeLanguage)

##   
## Two Sample t-test  
##   
## data: group1 and group2  
## t = -0.098565, df = 29, p-value = 0.9222  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.1478624 0.1342659  
## sample estimates:  
## mean of x mean of y   
## 0.4473684 0.4541667

4.Sex

group1 <- df\_students$`Total.Score`[df\_students$Sex == 0]  
group2 <- df\_students$`Total.Score`[df\_students$Sex == 1]  
  
t\_test\_result <- t.test(group1, group2, var.equal = TRUE)   
print(t\_test\_result)

##   
## Two Sample t-test  
##   
## data: group1 and group2  
## t = 1.4789, df = 29, p-value = 0.1499  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.04193559 0.26095733  
## sample estimates:  
## mean of x mean of y   
## 0.5312500 0.4217391

5.Born USA

group1 <- df\_students$`Total.Score`[df\_students$`Born.USA` == 0]  
group2 <- df\_students$`Total.Score`[df\_students$`Born.USA` == 1]  
  
t\_test\_result <- t.test(group1, group2, var.equal = TRUE)   
print(t\_test\_result)

##   
## Two Sample t-test  
##   
## data: group1 and group2  
## t = 0.57732, df = 29, p-value = 0.5682  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -0.1172949 0.2095569  
## sample estimates:  
## mean of x mean of y   
## 0.4604167 0.4142857

6.Accomodations

#group1 <- df\_students$`Total.Score`[df\_students$Accomodations == 0]  
#group2 <- df\_students$`Total.Score`[df\_students$Accomodations == 1]  
  
#t\_test\_result <- t.test(group1, group2, var.equal = TRUE)   
#print(t\_test\_result)

U-test 1.Sex

u\_test\_sex <- wilcox.test(df\_students$`Total.Score` ~ df\_students$Sex, exact = FALSE)  
  
print(u\_test\_sex)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: df\_students$Total.Score by df\_students$Sex  
## W = 125, p-value = 0.139  
## alternative hypothesis: true location shift is not equal to 0

Sex is 0.139, which is not verified to the U test from client.

2.Born USA

u\_test\_born <- wilcox.test(df\_students$`Total.Score` ~ df\_students$`Born.USA`, exact = FALSE)  
  
print(u\_test\_born)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: df\_students$Total.Score by df\_students$Born.USA  
## W = 96.5, p-value = 0.5675  
## alternative hypothesis: true location shift is not equal to 0

The p-value of Born in USA is also different 3.Accomodations

u\_test\_accomodations <- wilcox.test(df\_students$`Total.Score` ~ df\_students$Accomodations, exact = FALSE)  
  
print(u\_test\_accomodations)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: df\_students$Total.Score by df\_students$Accomodations  
## W = 29, p-value = 1  
## alternative hypothesis: true location shift is not equal to 0

4.English

u\_test\_english <- wilcox.test(df\_students$`Total.Score` ~ df\_students$`English.Proficiency`, exact = FALSE)  
  
print(u\_test\_english)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: df\_students$Total.Score by df\_students$English.Proficiency  
## W = 144.5, p-value = 0.2758  
## alternative hypothesis: true location shift is not equal to 0

5.Race

u\_test\_race <- wilcox.test(df\_students$`Total.Score` ~ df\_students$`Race.Ethnicity`, exact = FALSE)  
  
print(u\_test\_race)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: df\_students$Total.Score by df\_students$Race.Ethnicity  
## W = 114.5, p-value = 0.9357  
## alternative hypothesis: true location shift is not equal to 0

6.Home language

u\_test\_homelanguage <- wilcox.test(df\_students$`Total.Score` ~ df\_students$`Home.Language`, exact = FALSE)  
  
print(u\_test\_homelanguage)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: df\_students$Total.Score by df\_students$Home.Language  
## W = 112, p-value = 0.9511  
## alternative hypothesis: true location shift is not equal to 0

In order to get a comprehensive result, We also calculated t test for those none-normal distributed and the u test for those normal distributed data.

1.race\_T

group1 <- df\_students$`Quiz.Time`[df\_students$`English.Proficiency` == 0]  
group2 <- df\_students$`Quiz.Time`[df\_students$`English.Proficiency` == 1]  
t\_test\_result <- t.test(group1, group2, var.equal = TRUE)   
print(t\_test\_result)

##   
## Two Sample t-test  
##   
## data: group1 and group2  
## t = 1.586, df = 29, p-value = 0.1236  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -3061.283 24208.992  
## sample estimates:  
## mean of x mean of y   
## 12362.778 1788.923

2.Enlgish\_T

group1 <- df\_students$`Quiz.Time`[df\_students$`English.Proficiency` == 0]  
group2 <- df\_students$`Quiz.Time`[df\_students$`English.Proficiency` == 1]  
  
t\_test\_result <- t.test(group1, group2, var.equal = TRUE)   
print(t\_test\_result)

##   
## Two Sample t-test  
##   
## data: group1 and group2  
## t = 1.586, df = 29, p-value = 0.1236  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -3061.283 24208.992  
## sample estimates:  
## mean of x mean of y   
## 12362.778 1788.923

3.BornUS\_T

group1 <- df\_students$`Quiz.Time`[df\_students$`English.Proficiency` == 0]  
group2 <- df\_students$`Quiz.Time`[df\_students$`English.Proficiency` == 1]  
  
t\_test\_result <- t.test(group1, group2, var.equal = TRUE)   
print(t\_test\_result)

##   
## Two Sample t-test  
##   
## data: group1 and group2  
## t = 1.586, df = 29, p-value = 0.1236  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -3061.283 24208.992  
## sample estimates:  
## mean of x mean of y   
## 12362.778 1788.923

1.race\_U

u\_test\_race\_quiztime <- wilcox.test(df\_students$`Quiz.Time` ~ df\_students$`Race.Ethnicity`, exact = FALSE)  
  
print(u\_test\_race\_quiztime)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: df\_students$Quiz.Time by df\_students$Race.Ethnicity  
## W = 128, p-value = 0.6742  
## alternative hypothesis: true location shift is not equal to 0

2.English\_U

u\_test\_english\_quiztime <- wilcox.test(df\_students$`Quiz.Time` ~ df\_students$`English.Proficiency`, exact = FALSE)  
  
print(u\_test\_english\_quiztime)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: df\_students$Quiz.Time by df\_students$English.Proficiency  
## W = 155, p-value = 0.1333  
## alternative hypothesis: true location shift is not equal to 0

3.BornUS\_U

u\_test\_born\_quiztime <- wilcox.test(df\_students$`Quiz.Time` ~ df\_students$`Born.USA`, exact = FALSE)  
  
print(u\_test\_born\_quiztime)

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: df\_students$Quiz.Time by df\_students$Born.USA  
## W = 92, p-value = 0.7231  
## alternative hypothesis: true location shift is not equal to 0

chisq.test(table(df\_students$`Quiz.Time.Group`, df\_students$`English.Proficiency`))

## Warning in chisq.test(table(df\_students$Quiz.Time.Group,  
## df\_students$English.Proficiency)): Chi-squared approximation may be incorrect

##   
## Pearson's Chi-squared test  
##   
## data: table(df\_students$Quiz.Time.Group, df\_students$English.Proficiency)  
## X-squared = 2.4247, df = 2, p-value = 0.2975

We can find in all those different grouping method, the P-value is larger than 0.05, which means there is no such a huge difference between diffient groups.