3480final

Shaoran Sun

May 7, 2016

Adult <- read.csv("~/Desktop/3480 Final/Adult.csv")  
attach(Adult)  
  
occ = Adult$Occupation  
edu = Adult$EducationNum  
  
occ<-as.integer(occ)  
  
##c("Tech-support","?","Adm-clerical","Farming-fishing","Other-service","Sales","Exec-managerial","Protective-serv","Prof-specialty","Craft-repair","Handlers-cleaners","Machine-op-inspct","Transport-moving"))  
  
library(jmuOutlier)

## Warning: package 'jmuOutlier' was built under R version 3.2.4

rmd.test(occ, edu, all.perms = TRUE, num.sim = 20000)

## [[1]]  
## [1] "p-value was estimated based on 20000 simulations."  
##   
## $alternative  
## [1] "two.sided"  
##   
## $rmd.hat  
## [1] 1.387978  
##   
## $p.value  
## [1] 0.0129

ks.test(occ, edu)

## Warning in ks.test(occ, edu): cannot compute exact p-value with ties

##   
## Two-sample Kolmogorov-Smirnov test  
##   
## data: occ and edu  
## D = 0.66292, p-value < 2.2e-16  
## alternative hypothesis: two-sided

1. One question that we can solve using the Materials from Chapter 0, 1, and 2 Explanation / Motivation: We are interested to find out if adults with different education levels may result in government occupation or not. To be more specific, we are testing if adults with above college degree have more government occupation than adults without college degree.

Method / Statistical Tests to use:  
RMD test (to test if two attributes have equal variance) Wilcoxon Rank Sum (test if two distributions are equal)

We plan to use the RMD Test to test for the equal variances first, with assumptions:

. Since we want to compare the test scores from Occupation with the test scores with EducationNum, we will divide the observations into two groups. Both groups contain 89 observations. RMD Test is used to test for deviances. We will use the sample medians to obtain the deviances for each group.

For this case, The RMD reject the null hypothesis and we have unequal variances. In another words, . We could not use the WRS Test to determine if adults with above college degree have more government occupation than adults without college degree. We should therefore apply Kolmogorov-Smirnov (K-S) test. This K-S test is the nonparametric analog to the two-sample t-test with unequal variances. Its assumptions are independent samples with identical distributions. Our data fulfill these requirements and we can apply the K-S Test. As for the hypothesis, the null hypothesis is:

It stands for that adults education level has the same distribution with adult occupation in general. The alternative hypothesis is:

. It means that adults education level has different distribution with adult occupation in general.

# Discussion of statistical tests

The RMD Test gave us the p-value of . We reject the null hypothesis that there are equal variances between our state and the surrounding states at a significance level of 0.05. We came to the conclusion that the two groups have unequal variances and we can not use the WRS Test.

Therefore, we apply the Kolmogorov-Smirnov Test. The KS Test gave us the p-value of less than 2.2e-16. We reject the null hypothesis that adults education level has the same distribution with adult occupation in general at a significance level of 0.05. We came to the conclusion that adults education level has different distribution with adult occupation.