Ward_abigail_HW8

Abigail Ward

8/10/2021

1

1.a

Please compute the expected value of χ_1^2, the χ^2 distribution with 1 degree of freedom.

$$Var[\chi] = E[\chi^2] - (E[\chi])^2$$

$$E(\chi_1^2) = E(Z^2) = Var[Z] + (E[Z])^2$$

$Var[Z] = 1, E[Z] = 0,$ so $(E[Z])^2 = 0$

$$E(\chi_1^2) = 1 + 0 = 1$$

1.b

Based on your answer to 1.a, please compute the expected value of χ_n^2, the χ^2 distribution with n degrees of freedom.

$$\chi_n^2 = \sum_{i=1}^{n} Z_i^2$$

$$E(\chi_n^2) = \sum_{i=1}^{n} E\,[Z_i^2] = \sum_{i=1}^{n} 1 = n$$

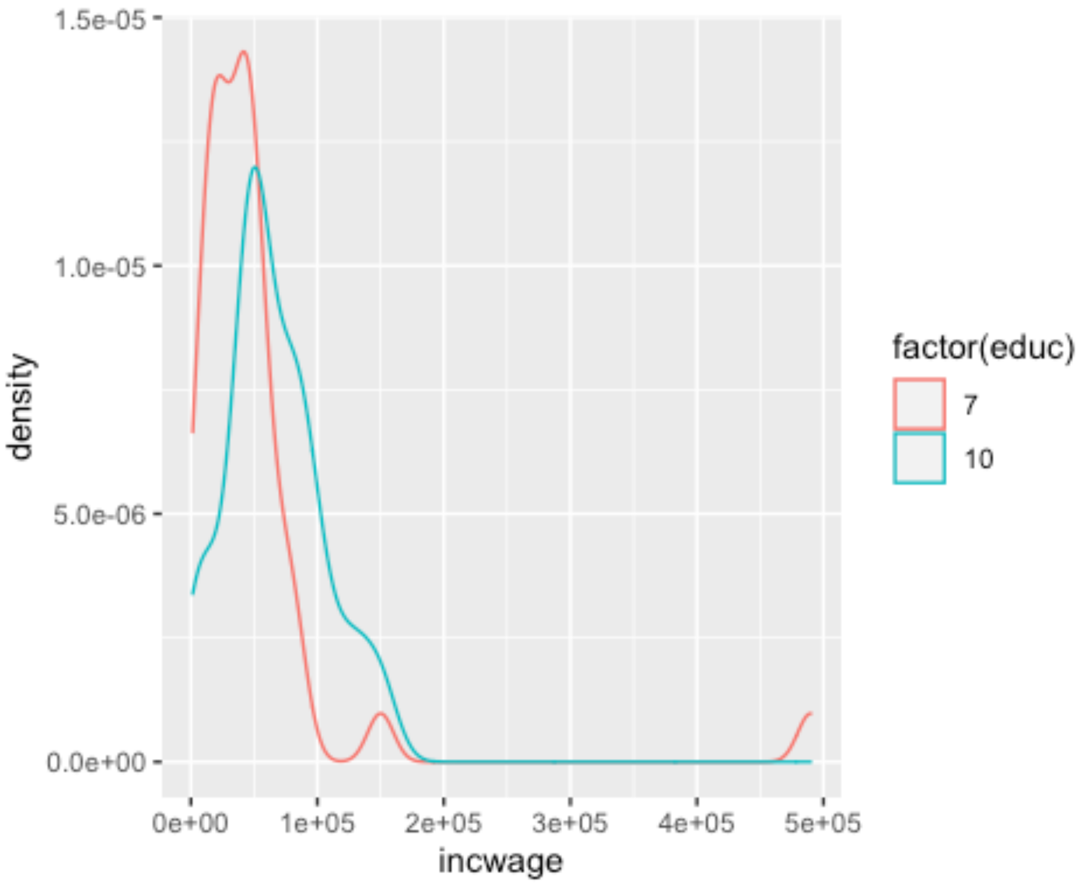
2

2.a

The Mann-Whitney U tests the nulll hypothesis that there is no difference in the incwage between the educ=7 group and the educ=10 group. The test showed that there was a significant difference (W = 467.5, p-value = 0.001395) in the income wage between the two education level groups.

Looking at the density plots, it is not reasonable to say that the red curve is the same shape as the turquoise, but shifted right. Therefore we reject that the two samples come from the same distribution and that the two samples come from distributions where one is a shift of the other.

```r
load("dat_7_10.RData")

wilcox.test(incwage~educ, data=dat.7.10)

## Warning in wilcox.test.default(x = c(78000L, 78000L, 48000L, 60000L,
45000L, :
## cannot compute exact p-value with ties

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  incwage by educ
## W = 467.5, p-value = 0.001395
## alternative hypothesis: true location shift is not equal to 0

wilcox.test(incwage~educ, data=dat.7.10, paired=TRUE)

## Warning in wilcox.test.default(x = c(78000L, 78000L, 48000L, 60000L,
45000L, :
## cannot compute exact p-value with ties

## Warning in wilcox.test.default(x = c(78000L, 78000L, 48000L, 60000L,
45000L, :
## cannot compute exact p-value with zeroes

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  incwage by educ
## V = 189, p-value = 0.005139
## alternative hypothesis: true location shift is not equal to 0

ggplot(dat.7.10,aes(x=incwage,color=factor(educ)))+geom_density()
```



2.b

There is no difference in the results of the log test vs the original test because the Mann-Whitney U test is a rank test so the actual values of the data do not matter, only the rank each data point is labelled. Therefore, taking the log of each data point did not change the rank of each data point because the same operation was applied to every data point.

```
wilcox.test(log(incwage)~educ, data=dat.7.10)

## Warning in wilcox.test.default(x = c(11.2644641056717, 11.2644641056717, :
## cannot compute exact p-value with ties

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  log(incwage) by educ
## W = 467.5, p-value = 0.001395
## alternative hypothesis: true location shift is not equal to 0
```

3

```
load("dat_sub.RData")
t<-table(dat.sub$F_AGECAT,dat.sub$NATPROBS_b_W69)
t<-t[1:4,1:4] # Drop the "Refused" row and column
```

```r
round(100*t/rowSums(t),0)
```

```
##
##         A very big problem A moderately big problem A small problem
##   18-29                 66                       18              16
##   30-49                 39                       47              12
##   50-64                 63                       23              12
##   65+                   63                       28               9
##
##         Not a problem at all
##   18-29                    0
##   30-49                    2
##   50-64                    2
##   65+                      0
```

3.a

Chi squared is not an appropriate test of the independence of this data because all of the "Not a problem at all" values are less than 5 which breaks the criteria needed to run a chi squared test accurately.

```r
chisq.test(t)
```

```
## Warning in chisq.test(t): Chi-squared approximation may be incorrect

##
##  Pearson's Chi-squared test
##
## data:  t
## X-squared = 16.261, df = 9, p-value = 0.06163
```

3.b

The fisher test is an appropriate test for the data because it works when the values are less than 5 and for categorical data. With a p-value of 0.03326, we reject the null hypothesis that the two categories are independent.

```r
fisher.test(t)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  t
## p-value = 0.03326
## alternative hypothesis: two.sided
```