Problem Set 6

```
rm(list=ls())
```

#Question 1 A new three-sided die takes on values of 1, 2, and 3. It is supposed to be fair, but you expect Pr(x = 1) != 1/3. You have data from 5 rolls of the die. Use an exact test

#1A Describe the results that would allow you to accept and reject the null hypothesis of a fair die with a two-sided test at the 95% confidence level

```
#Null Hypotheses Pr(X ==1) == 1/3
#Expected or Alternate Hypotheses Pr(X==1) != 1/3

#Number of rolls = 5
#Given Confidence level = 95% (0.95)

rolls <- 5
confidence_level = 0.95

for (k in 0:rolls){
  sample = binom.test(k, n = rolls, p = 1/3 ,alternative = "two.sided", conf.level = confidence_level)
  if (sample$p.value >= (1 - confidence_level)){
    print(paste("Null Hypotheses not rejected, successes=successes=", k))
    print(paste("p.value = ",sample$p.value))
    print(paste("Confidence interval=[", sample$conf.int[1], ", ", sample$conf.int[2], "]" ) )

  }else{
    print(paste("Null Hypotheses rejected, successes = ", k))
    print(paste("p.value = ",sample$p.value))
    print(paste("Confidence interval=[", sample$conf.int[1], ", ", sample$conf.int[2], "]" ) )
  }
}
```

```
## [1] "Null Hypotheses not rejected, successes=successes= 0"
## [1] "p.value =  0.176954732510288"
## [1] "Confidence interval=[ 0 ,  0.521823750104981 ]"
## [1] "Null Hypotheses not rejected, successes=successes= 1"
## [1] "p.value =  1"
## [1] "Confidence interval=[ 0.00505076337946806 ,  0.716417936118089 ]"
## [1] "Null Hypotheses not rejected, successes=successes= 2"
## [1] "p.value =  1"
## [1] "Confidence interval=[ 0.0527449505263169 ,  0.853367200365327 ]"
## [1] "Null Hypotheses not rejected, successes=successes= 3"
## [1] "p.value =  0.341563786008231"
## [1] "Confidence interval=[ 0.146632799634673 ,  0.947255049473683 ]"
## [1] "Null Hypotheses rejected, successes =  4"
## [1] "p.value =  0.0452674897119341"
## [1] "Confidence interval=[ 0.283582063881911 ,  0.994949236620532 ]"
```

```
## [1] "Null Hypotheses rejected, successes =  5"
## [1] "p.value =  0.00411522633744856"
## [1] "Confidence interval=[ 0.478176249895019 ,  1 ]"
```

```
#1B:  A new three-sided die takes on values of 1, 2, and 3. It is supposed to be fair, but you expect P


#Null hypotheses Pr(x = 1) == 1/3
#Expected or Alternative Hypotheses Pr(x = 1) < 1/3



#One-sided less test
rolls <- 5
confidence_level = 0.95


for (k in 0:rolls){
  sample = binom.test(k, n = rolls, p = 1/3 ,alternative = "less", conf.level = confidence_level)
  if (sample$p.value >= (1 - confidence_level)){
    print(paste("Null Hypotheses not rejected, successes=", k))
    print(paste( "p.value=",sample$p.value))
    print(paste("Confidence interval=[", sample$conf.int[1], ", ", sample$conf.int[2], "]" ) )

  }else{
    print(paste("Null Hypotheses rejected, successes=", k))
    print(paste("p.value=",sample$p.value))
    print(paste("Confidence interval=[", sample$conf.int[1], ", ", sample$conf.int[2], "]" ) )

  }
}
```

```
## [1] "Null Hypotheses not rejected, successes= 0"
## [1] "p.value= 0.131687242798354"
## [1] "Confidence interval=[ 0 ,  0.450719728346941 ]"
## [1] "Null Hypotheses not rejected, successes= 1"
## [1] "p.value= 0.460905349794239"
## [1] "Confidence interval=[ 0 ,  0.657408318001139 ]"
## [1] "Null Hypotheses not rejected, successes= 2"
## [1] "p.value= 0.790123456790123"
## [1] "Confidence interval=[ 0 ,  0.810744622562229 ]"
## [1] "Null Hypotheses not rejected, successes= 3"
## [1] "p.value= 0.954732510288066"
## [1] "Confidence interval=[ 0 ,  0.923559608587671 ]"
## [1] "Null Hypotheses not rejected, successes= 4"
## [1] "p.value= 0.995884773662551"
## [1] "Confidence interval=[ 0 ,  0.989793781686989 ]"
## [1] "Null Hypotheses not rejected, successes= 5"
## [1] "p.value= 1"
## [1] "Confidence interval=[ 0 ,  1 ]"
```

```
#2A:  Repeat the exercise in Question 1A assuming you have 25 observations.
```

```r
rolls = 25
confidence_level = 0.95

for (k in 0:rolls){
 sample = binom.test(k, n = rolls, p = 1/3 ,alternative = "two.sided", conf.level = confidence_level)
  if (sample$p.value >= (1 - confidence_level)){
    print(paste("Null Hypotheses not rejected, successes=", k))
    print(paste( "pvalue = ",sample$p.value))
    print(paste("Confidence interval=[", sample$conf.int[1], ", ", sample$conf.int[2], "]" ) )
  }else{
    print(paste("Null Hypotheses rejected, successes=", k))
    print(paste( "p.value = ",sample$p.value))
    print(paste("Confidence interval=[", sample$conf.int[1], ", ", sample$conf.int[2], "]" ) )
  }
}
```

```
## [1] "Null Hypotheses rejected, successes= 0"
## [1] "p.value =  5.5248049458346e-05"
## [1] "Confidence interval=[ 0 ,  0.137185171530713 ]"
## [1] "Null Hypotheses rejected, successes= 1"
## [1] "p.value =  0.000949681412014937"
## [1] "Confidence interval=[ 0.00101219969931085 ,  0.203516913922414 ]"
## [1] "Null Hypotheses rejected, successes= 2"
## [1] "p.value =  0.00515437100691227"
## [1] "Confidence interval=[ 0.00983959001879751 ,  0.260305842105214 ]"
## [1] "Null Hypotheses rejected, successes= 3"
## [1] "p.value =  0.0312645587203331"
## [1] "Confidence interval=[ 0.0254653966477332 ,  0.312190307286235 ]"
## [1] "Null Hypotheses not rejected, successes= 4"
## [1] "pvalue =  0.0877145110352151"
## [1] "Confidence interval=[ 0.0453794523717096 ,  0.360828454459272 ]"
## [1] "Null Hypotheses not rejected, successes= 5"
## [1] "pvalue =  0.203745458913329"
## [1] "Confidence interval=[ 0.068311464012484 ,  0.407037432278677 ]"
## [1] "Null Hypotheses not rejected, successes= 6"
## [1] "pvalue =  0.39952461202746"
## [1] "Confidence interval=[ 0.0935644393317429 ,  0.451288017816686 ]"
## [1] "Null Hypotheses not rejected, successes= 7"
## [1] "pvalue =  0.674665037476179"
## [1] "Confidence interval=[ 0.120716688504067 ,  0.493876821806255 ]"
## [1] "Null Hypotheses not rejected, successes= 8"
## [1] "pvalue =  1"
## [1] "Confidence interval=[ 0.14949542261357 ,  0.535000717497372 ]"
## [1] "Null Hypotheses not rejected, successes= 9"
## [1] "pvalue =  0.832684876416322"
## [1] "Confidence interval=[ 0.179716820583655 ,  0.574793650446151 ]"
## [1] "Null Hypotheses not rejected, successes= 10"
## [1] "pvalue =  0.525940483179574"
## [1] "Confidence interval=[ 0.211254806465142 ,  0.61334650374316 ]"
## [1] "Null Hypotheses not rejected, successes= 11"
## [1] "pvalue =  0.289938098335225"
## [1] "Confidence interval=[ 0.244023665147208 ,  0.650718366008664 ]"
## [1] "Null Hypotheses not rejected, successes= 12"
```

```
## [1] "pvalue =  0.137993550697988"
## [1] "Confidence interval=[ 0.277968009669947 ,  0.686942955542968 ]"
## [1] "Null Hypotheses not rejected, successes= 13"
## [1] "pvalue =  0.0564040785517194"
## [1] "Confidence interval=[ 0.313057044457032 ,  0.722031990330053 ]"
## [1] "Null Hypotheses rejected, successes= 14"
## [1] "p.value =  0.0198789469081528"
## [1] "Confidence interval=[ 0.349281633991336 ,  0.755976334852792 ]"
## [1] "Null Hypotheses rejected, successes= 15"
## [1] "p.value =  0.00910486698041584"
## [1] "Confidence interval=[ 0.38665349625684 ,  0.788745193534858 ]"
## [1] "Null Hypotheses rejected, successes= 16"
## [1] "p.value =  0.0021842114037348"
## [1] "Confidence interval=[ 0.425206349553849 ,  0.820283179416345 ]"
## [1] "Null Hypotheses rejected, successes= 17"
## [1] "p.value =  0.000454654811485359"
## [1] "Confidence interval=[ 0.464999282502628 ,  0.85050457738643 ]"
## [1] "Null Hypotheses rejected, successes= 18"
## [1] "p.value =  0.000127867460735985"
## [1] "Confidence interval=[ 0.506123178193745 ,  0.879283311495933 ]"
## [1] "Null Hypotheses rejected, successes= 19"
## [1] "p.value =  1.56459214159798e-05"
## [1] "Confidence interval=[ 0.548711982183314 ,  0.906435560668257 ]"
## [1] "Null Hypotheses rejected, successes= 20"
## [1] "p.value =  2.26866144378318e-06"
## [1] "Confidence interval=[ 0.592962567721323 ,  0.931688535987516 ]"
## [1] "Null Hypotheses rejected, successes= 21"
## [1] "p.value =  2.62072447953684e-07"
## [1] "Confidence interval=[ 0.639171545540728 ,  0.95462054762829 ]"
## [1] "Null Hypotheses rejected, successes= 22"
## [1] "p.value =  2.31928055930297e-08"
## [1] "Confidence interval=[ 0.687809692713765 ,  0.974534603352267 ]"
## [1] "Null Hypotheses rejected, successes= 23"
## [1] "p.value =  1.47647446933388e-09"
## [1] "Confidence interval=[ 0.739694157894786 ,  0.990160409981202 ]"
## [1] "Null Hypotheses rejected, successes= 24"
## [1] "p.value =  6.01920047450266e-11"
## [1] "Confidence interval=[ 0.796483086077586 ,  0.998987800300689 ]"
## [1] "Null Hypotheses rejected, successes= 25"
## [1] "p.value =  1.18023538715738e-12"
## [1] "Confidence interval=[ 0.862814828469287 ,  1 ]"
```

#2B: Repeat the exercise in Question 1B assuming you have 25 observations.

```
rolls <- 25
confidence_level = 0.95

for (k in 0:rolls){
  sample = binom.test(k, n = rolls, p = 1/3 ,alternative = "less", conf.level = confidence_level)
  if (sample$p.value >= (1 - confidence_level)){
    print(paste("Null Hypotheses not rejected, successes=", k))
    print(paste( "p.value = ",sample$p.value))
    print(paste("Confidence intervel=[", sample$conf.int[1], ", ", sample$conf.int[2], "]" ) )
  }else{
```

```
    print(paste("Null Hypotheses rejected, successes=", k))
    print(paste( "p.value = ",sample$p.value))
    print(paste("Confidence interval=[", sample$conf.int[1], ", ", sample$conf.int[2], "]" ) )
  }
}
```

```
## [1] "Null Hypotheses rejected, successes= 0"
## [1] "p.value =  3.96021280423662e-05"
## [1] "Confidence interval=[ 0 ,  0.112928145006843 ]"
## [1] "Null Hypotheses rejected, successes= 1"
## [1] "p.value =  0.000534628728571944"
## [1] "Confidence interval=[ 0 ,  0.176120710604518 ]"
## [1] "Null Hypotheses rejected, successes= 2"
## [1] "p.value =  0.00350478833174942"
## [1] "Confidence interval=[ 0 ,  0.231039933958144 ]"
## [1] "Null Hypotheses rejected, successes= 3"
## [1] "p.value =  0.0148904001439297"
## [1] "Confidence interval=[ 0 ,  0.281722507977714 ]"
## [1] "Null Hypotheses rejected, successes= 4"
## [1] "p.value =  0.0462008326274255"
## [1] "Confidence interval=[ 0 ,  0.329608298065589 ]"
## [1] "Null Hypotheses not rejected, successes= 5"
## [1] "p.value =  0.111952740842767"
## [1] "Confidence intervel=[ 0 ,  0.375405135124878 ]"
## [1] "Null Hypotheses not rejected, successes= 6"
## [1] "p.value =  0.221539254535002"
## [1] "Confidence intervel=[ 0 ,  0.419520007210026 ]"
## [1] "Null Hypotheses not rejected, successes= 7"
## [1] "p.value =  0.370263808831607"
## [1] "Confidence intervel=[ 0 ,  0.462208921703645 ]"
## [1] "Null Hypotheses not rejected, successes= 8"
## [1] "p.value =  0.537578932415285"
## [1] "Confidence intervel=[ 0 ,  0.503641562379504 ]"
## [1] "Null Hypotheses not rejected, successes= 9"
## [1] "p.value =  0.695598771355428"
## [1] "Confidence intervel=[ 0 ,  0.543933213922907 ]"
## [1] "Null Hypotheses not rejected, successes= 10"
## [1] "p.value =  0.822014642507542"
## [1] "Confidence intervel=[ 0 ,  0.583161963015353 ]"
## [1] "Null Hypotheses not rejected, successes= 11"
## [1] "p.value =  0.908207281929438"
## [1] "Confidence intervel=[ 0 ,  0.621378436607113 ]"
## [1] "Null Hypotheses not rejected, successes= 12"
## [1] "p.value =  0.95848632159221"
## [1] "Confidence intervel=[ 0 ,  0.658611336570926 ]"
## [1] "Null Hypotheses not rejected, successes= 13"
## [1] "p.value =  0.983625841423597"
## [1] "Confidence intervel=[ 0 ,  0.694870320348841 ]"
## [1] "Null Hypotheses not rejected, successes= 14"
## [1] "p.value =  0.994399921351334"
## [1] "Confidence intervel=[ 0 ,  0.730146941293753 ]"
## [1] "Null Hypotheses not rejected, successes= 15"
## [1] "p.value =  0.998350417324837"
```

```
## [1] "Confidence intervel=[ 0 ,  0.764413869598757 ]"
## [1] "Null Hypotheses not rejected, successes= 16"
## [1] "p.value =  0.999584947316557"
## [1] "Confidence intervel=[ 0 ,  0.797622220145338 ]"
## [1] "Null Hypotheses not rejected, successes= 17"
## [1] "p.value =  0.999911734667306"
## [1] "Confidence intervel=[ 0 ,  0.829696346026529 ]"
## [1] "Null Hypotheses not rejected, successes= 18"
## [1] "p.value =  0.999984354078584"
## [1] "Confidence intervel=[ 0 ,  0.86052469342107 ]"
## [1] "Null Hypotheses not rejected, successes= 19"
## [1] "p.value =  0.999997731338556"
## [1] "Confidence intervel=[ 0 ,  0.889943800338206 ]"
## [1] "Null Hypotheses not rejected, successes= 20"
## [1] "p.value =  0.999999737927552"
## [1] "Confidence intervel=[ 0 ,  0.917709100138754 ]"
## [1] "Null Hypotheses not rejected, successes= 21"
## [1] "p.value =  0.999999976807194"
## [1] "Confidence intervel=[ 0 ,  0.943437440979946 ]"
## [1] "Null Hypotheses not rejected, successes= 22"
## [1] "p.value =  0.999999998523526"
## [1] "Confidence intervel=[ 0 ,  0.966480405010495 ]"
## [1] "Null Hypotheses not rejected, successes= 23"
## [1] "p.value =  0.999999999939808"
## [1] "Confidence intervel=[ 0 ,  0.985596802020826 ]"
## [1] "Null Hypotheses not rejected, successes= 24"
## [1] "p.value =  0.99999999999882"
## [1] "Confidence intervel=[ 0 ,  0.997950371587379 ]"
## [1] "Null Hypotheses not rejected, successes= 25"
## [1] "p.value =  1"
## [1] "Confidence intervel=[ 0 ,  1 ]"
```

```
#3. (25 points) Using data from Homework 4, test the hypothesis that African American women and white w


library(ggplot2)
library(moments)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data <- read.csv("E:/Autumn'21/Advanced_Stats/ProblemSets/4/ppha312x2021.csv")

data <- data %>% filter(inctot >= 0)
```

```r
#New column: Hispanic

data %>% count(hispan)
```

```
##         hispan    n
## 1        Cuban    4
## 2      Mexican  212
## 3 Not Hispanic 9851
## 4        Other   52
## 5 Puerto Rican   21
```

```r
data$isHispanic <- ifelse(data$hispan == 'Not Hispanic', 0, 1)
data <- data %>% mutate(isHispanic = as.factor(isHispanic))
summary(data$isHispanic)
```

```
##    0    1
## 9851  289
```

```r
summary(data$hispan)
```

```
##    Length     Class      Mode
##     10140 character character
```

```r
summary(data)
```

```
##       year        statefip           met2013              perwt
##  Min.   :2019   Length:10140       Length:10140       Min.   :   2.0
##  1st Qu.:2019   Class :character   Class :character   1st Qu.:  57.0
##  Median :2019   Mode  :character   Mode  :character   Median :  85.0
##  Mean   :2019                                         Mean   : 115.6
##  3rd Qu.:2019                                         3rd Qu.: 133.0
##  Max.   :2019                                         Max.   :1977.0
##      sex                age                race              hispan
##  Length:10140       Length:10140       Length:10140       Length:10140
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      bpl                educd              empstat            uhrswork
##  Length:10140       Length:10140       Length:10140       Length:10140
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      inctot            incwage       isHispanic
##  Min.   :      0   Min.   :     0   0:9851
##  1st Qu.:  11775   1st Qu.:     0   1: 289
##  Median :  35000   Median : 25000
##  Mean   :1744935   Mean   :209709
##  3rd Qu.:  94000   3rd Qu.: 88250
##  Max.   :9999999   Max.   :999999
```

```r
#New column: African American
data$isAfricanAmerican <- ifelse(data$race == 'Black/African American/Negro', 1, 0)
data <- data %>% mutate(isAfricanAmerican = as.factor(isAfricanAmerican))
summary(data$isAfricanAmerican)
```

```
##    0    1
## 6241 3899
```

```r
summary(data)
```

```
##       year        statefip           met2013             perwt
##  Min.   :2019   Length:10140       Length:10140       Min.   :   2.0
##  1st Qu.:2019   Class :character   Class :character   1st Qu.:  57.0
##  Median :2019   Mode  :character   Mode  :character   Median :  85.0
##  Mean   :2019                                         Mean   : 115.6
##  3rd Qu.:2019                                         3rd Qu.: 133.0
##  Max.   :2019                                         Max.   :1977.0
##      sex                 age                race              hispan
##  Length:10140       Length:10140       Length:10140       Length:10140
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      bpl                educd              empstat            uhrswork
##  Length:10140       Length:10140       Length:10140       Length:10140
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      inctot            incwage       isHispanic isAfricanAmerican
##  Min.   :      0   Min.   :     0   0:9851     0:6241
##  1st Qu.:  11775   1st Qu.:     0   1: 289     1:3899
##  Median :  35000   Median : 25000
##  Mean   :1744935   Mean   :209709
##  3rd Qu.:  94000   3rd Qu.: 88250
##  Max.   :9999999   Max.   :999999
```

```r
#Filtering only white, non-Hispanic or African American, non-Hispanic ;

data <- data %>% filter((race == 'White' & isHispanic == 0) |
                          (isAfricanAmerican == 1 & isHispanic == 0))
summary(data)
```

```
##       year        statefip           met2013             perwt
##  Min.   :2019   Length:9632        Length:9632        Min.   :   2.0
##  1st Qu.:2019   Class :character   Class :character   1st Qu.:  57.0
##  Median :2019   Mode  :character   Mode  :character   Median :  84.0
##  Mean   :2019                                         Mean   : 114.2
##  3rd Qu.:2019                                         3rd Qu.: 131.0
##  Max.   :2019                                         Max.   :1977.0
```

```
##      sex                 age                 race              hispan
##  Length:9632        Length:9632        Length:9632        Length:9632
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      bpl                 educd              empstat            uhrswork
##  Length:9632        Length:9632        Length:9632        Length:9632
##  Class :character   Class :character   Class :character   Class :character
##  Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##
##      inctot            incwage       isHispanic isAfricanAmerican
##  Min.   :      0   Min.   :     0   0:9632     0:5738
##  1st Qu.:  11800   1st Qu.:     0   1:   0     1:3894
##  Median :  35000   Median : 24000
##  Mean   :1606598   Mean   :196023
##  3rd Qu.:  86000   3rd Qu.: 80000
##  Max.   :9999999   Max.   :999999
```

```
#Filtering to  women and age [25:55]

data$isFemale <- ifelse(data$sex == 'Female', 1, 0)
data <- data %>% mutate(isFemale = as.factor(isFemale))
data <- data %>% mutate(age = as.numeric(age))
```

```
## Warning in mask$eval_all_mutate(quo): NAs introduced by coercion
```

```
data <- data %>% filter(isFemale == 1) %>% filter(age >= 25 & age <= 55)
```

```
data_employed_by_race <-
  data %>%
  group_by(race, empstat) %>% count(race, empstat) %>% group_by(race) %>%
  mutate(race_count = sum(n), perc_freq = round(n / sum(n), 3)*100) %>% filter(empstat == 'Employed')
data_employed_by_race
```

```
## # A tibble: 2 x 5
## # Groups:   race [2]
##   race                          empstat      n race_count perc_freq
##   <chr>                         <chr>    <int>      <int>     <dbl>
## 1 Black/African American/Negro Employed   599        801      74.8
## 2 White                         Employed   791       1055      75
```

#Fisher's test:

```
#Two-tail test
#Null Hypotheses:Probability of being employed is same for both White woman and African-American women
#Alternative Hypotheses:Probability of being employed is not same for both White woman and African-Amer
test_mat <- matrix(c(791,264,599,202), nrow=2)
fisher.test(test_mat, alternative = "two.sided", conf.level = 0.95)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  test_mat
## p-value = 0.9569
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.8125879 1.2553781
## sample estimates:
## odds ratio
##   1.010405
```

#Observations: #African-American Women and White women with age range [25-55] have the same probability of getting employed. This is because, the p-value = 0.9569 we got from the two-tail test is greater than the standard 0.05 for the 95% confidence level and thus the null hypotheses is not rejected.

```
#One tail test
#Null Hypotheses:Probability of being employed is same for both White woman and African-American women
#Alternative Hypotheses:Probability of being employed is not same for White woman over African-American
fisher.test(test_mat, alternative='greater', conf.level = 0.95)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  test_mat
## p-value = 0.4828
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  0.840848      Inf
## sample estimates:
## odds ratio
##   1.010405
```

#Observations #African-American Women and White women with age range [25-55] have the same probability of getting employed. This is because, the p-value = 0.4828 we got from the test is greater than the standard 0.05 for the 95% confidence level and thus the null hypotheses is not rejected.

```
# chi-square test:
#Null Hypotheses:Probability of being employed is same for both White woman and African-American women
#Alternative Hypotheses:Probability of being employed is not same for White woman over African-American

chisq.test(test_mat)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  test_mat
## X-squared = 0.001748, df = 1, p-value = 0.9667
```

#Observations #African-American Women and White women with age range [25-55] have the same probability of getting employed. This is because, the p-value = 0.9667 we got from the test is greater than the standard 0.05 for the 95% confidence level and thus the null hypotheses is not rejected.

#####4444444444444444444444444444444444444444444444444######

4: Using data from Homework 4, test the hypothesis that African American women and white women ages 25-55 (inclusive) have the same total income. Use a two-sided t-test with a 95% confidence level. How confident of this test are you? Why?

```
data <- data %>%
  mutate(isEmployed = ifelse(data$empstat == 'Employed', 1, 0))
data <- data %>%
  mutate(isEmployed = as.factor(isEmployed))
data_income_AA_White <- data %>%
  filter(data$isEmployed == 1) %>%
  group_by(race) %>%
  summarize(count_employed = n(),
            total_income = sum(inctot, na.rm = TRUE),
            mean_income = mean(inctot, na.rm = TRUE))


data_income_AA_White
```

```
## # A tibble: 2 x 4
##   race                          count_employed total_income mean_income
##   <chr>                                  <int>        <int>       <dbl>
## 1 Black/African American/Negro             599     24083374      40206.
## 2 White                                    791     44241734      55931.
```

#Two-sided t-test with a 95% confidence level #Null hypotheses:True difference in Mean is 0 for African American women and White women with age range [25-55] #Alternative hypotheses: True difference in Mean is not 0 for African American women and white women with age range [25-55]

```
data_AA_employed <- data %>%
  filter(data$isEmployed == 1 & data$isAfricanAmerican == 1 )

data_totalIncome_AA_employed <- c(data_AA_employed$inctot)

data_White_employed <- data %>%
  filter(data$isEmployed == 1 & data$isAfricanAmerican == 0 )
data_totalIncome_White_employed <- c(data_White_employed$inctot)

t.test(data_totalIncome_AA_employed, data_totalIncome_White_employed, alternative = "two.sided",
       var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  data_totalIncome_AA_employed and data_totalIncome_White_employed
## t = -6.7861, df = 1333.4, p-value = 1.728e-11
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -20271.37 -11179.49
## sample estimates:
## mean of x mean of y
##  40205.97  55931.40
```

#Observations #Observed p-value = 1.728e-11 for the two-sided t-test with a 95% confidence, is less than the standard 0.05 and thus we can reject the null hypothes. Thus, we cannot reject the alternative hypotheses that the true difference in means for White Women and African-American Womnan with age range [25:55] is not 0.

```
t.test(data_totalIncome_AA_employed, data_totalIncome_White_employed, alternative = "less",
       var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  data_totalIncome_AA_employed and data_totalIncome_White_employed
## t = -6.7861, df = 1333.4, p-value = 8.64e-12
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -11911.17
## sample estimates:
## mean of x mean of y
##  40205.97  55931.40
```

#The one sided p value is 8.64e-12 which is less than 0.05. Thus, we can reject the null hypotheses and not reject thealternate hypotheses.

```
t.test(data_totalIncome_AA_employed, data_totalIncome_White_employed, alternative = "greater",
       var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  data_totalIncome_AA_employed and data_totalIncome_White_employed
## t = -6.7861, df = 1333.4, p-value = 1
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -19539.69       Inf
## sample estimates:
## mean of x mean of y
##  40205.97  55931.40
```

#Looking at the p values in the above tests, we cannot reject the hypothesis that the the total income of the African American women ages 25-55 (inclusive) is less that the white women age range [25:55]