Alexander Caines
MATH-310
4-15-2021

FINAL

1. Let the null hypothesis be that there is no difference in death rates for men and women, while the alternative hypothesis state that the death rates between men and women differ significantly. Because the data is in the form of a ratio (death rate), and there are two levels (one for men and another for women), we must oconduct a two-way ANOVA in order to determine the existence of a signifcant difference. The two-way ANOVA found the p-value to be 0.0112. Since it is not less than 0.05, we fail to reject the null hypothesis that there is a significant differnce in the number of deaths between men and women.

2. The racial group with the highest death rate is American Indian, with 0.0714. Despite the wide margin between the death rate in American Indian populations and other populations, the resulting anova did not identify a significant difference between races.

```
death_rates <- data[, c("Race", "Death_yn")]
death_rates$Death_yn <- as.numeric(lapply(death_rates$Death_yn, mod2))
head(death_rates); shapiro.test(death_rates$Death_yn); print("passes shapiro wilks")

#chisq.test(death_rates)
#leveneTest(Death_yn ~ Race, data=death_rates)
model <- aov(Death_yn ~ Race, data=death_rates)
summary(model)
death_rates$Death_yn <- as.numeric(lapply(death_rates$Death_yn, mod2b))


#print("Since the p-value is less than 0.05, the test shows that
#     race does not  significant effect on death rate(definitely does, but why doesn't this show it??)")
```

55:20 (Top Level)

Console Terminal Jobs

~/Documents/school/MATHSTATS/final/

```
data:  death_rates$Death_yn
W = 0.22146, p-value < 2.2e-16

[1] "passes shapiro wilks"

> #chisq.test(death_rates)
> #leveneTest(Death_yn ~ Race, data=death_rates)
> model <- aov(Death_yn ~ Race, data=death_rates)

> summary(model)
              Df Sum Sq Mean Sq F value Pr(>F)
Race           6   0.31 0.05112   1.093  0.364
Residuals   3593 167.99 0.04676

> death_rates$Death_yn <- as.numeric(lapply(death_rates$Death_yn, mod2b))

> #print("Since the p-value is less than 0.05, the test shows that
> #     race does not  significant effect on death rate(definitely does, but why do ..." ... [TRUNCATED]
>
```

3. The population of age 80 or older showed the greatest death rate of the other age groups. After conducting a one-way ANOVA, the resulting p-value was less than 2e-16, which is less than 0.05. So, we can reject the null that there is no significant difference between those older than 80 and other age groups.

```r
data <- read.csv('Final4.csv')

getlabel <- function(x,s) x[x$Age == s,]
deaths <- function(x) as.numeric(lapply(x$Death_yn, function(s) if(s=="Yes") 1 else 0))
death_rate <- function(L) sum(L[L == 1])/length(L)
process <- function(x,s) death_rate(deaths(getlabel(x,s)))

d3039 <- process(data, "30 - 39 Years")
d1019 <- process(data, "10 - 19 Years")
d4049 <- process(data, "40 - 49 Years")
d7079 <- process(data, "50 - 59 Years")
d2029 <- process(data, "20 - 29 Years")
d09 <- process(data, "0 - 9 Years")
d6069 <- process(data, "60 - 69 Years")
d80p <- process(data, "80+ Years")

max(d3039, d1019, d4049, d7079, d2029, d09, d6069, d80p)
print('The age group with teh highest death rate is 80+')

death_rates <- data[, c("Age", "Death_yn")]
death_rates$Death_yn <- as.numeric(lapply(death_rates$Death_yn, mod2))
model <- aov(Death_yn ~ Age, data=death_rates)

summary(model)
```

28:1 (Top Level)

Console Terminal Jobs

~/Documents/school/MATHSTATS/final/

```
> model <- aov(Death_yn ~ Age, data=death_rates)

> summary(model)
              Df Sum Sq Mean Sq F value Pr(>F)
Age            8  39.74   4.967   138.7 <2e-16 ***
Residuals   3591 128.56   0.036
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

4. To determine if there is an association between two variables, a Chi-Squared Test of Independence must be conducted. Let the null hypothesis be the claim that seriology level and death rate are independent of each other and the alternative hypothesis that they are associated. After running the ANOVA, the p-value was found to be less than 0.05. So, the null hypothesis is rejected and it can be concluded that there is an association between seriology and death rate. This differs with respect to gender. For if, without loss of generality, the same null and alternative hypotheses are made, the ANOVA reports a p-value greater than 0.05. So, there is no association between gender and death rate.

5. Let the null hypothesis claim that there is no association between cholesterol and seriology and the alternative hypothesis claim that there is. Since this implicitly assumes independence, we can proceed with running a Chi-Squared Test of Independence. The test reports a p-value less than 2e-16. So, the null hypothesis is rejected. Therefore, there is an association to be made between cholesterol and death rate.

```r
data <- read.csv('Final4.csv')
ser <- data[, c("Cholesterol", "Ser")]
nser <- ser[ser$Ser>0,]
str(nser)
chisq.test(nser)

print("Since the p-value is less than 0.05, we
      reject the null hypothesis that there is
      no association between cholesterol and seriability")

```

10:1 (Top Level)

Console Terminal Jobs

~/Documents/school/MATHSTATS/final/

```
> str(nser)
'data.frame':   3598 obs. of  2 variables:
 $ Cholesterol: num  171 185 165 148 174 ...
 $ Ser        : num  36 34.8 20.4 25.4 11.1 ...

> chisq.test(nser)

        Pearson's Chi-squared test

data:  nser
X-squared = 13606, df = 3597, p-value < 2.2e-16


> print("Since the p-value is less than 0.05, we
+       reject the null hypothesis that there is
+       no association between cholesterol and se ..." ... [TRUNCATED]
[1] "Since the p-value is less than 0.05, we \n      reject the null hypothesis that
ween cholesterol and seriability"
> 
```