

## 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 diffrence 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.

The screenshot shows the RStudio interface. The main editor contains R code for data manipulation and statistical testing. The Environment pane on the right shows the global environment with variables like 'multiple\_deaths', 'nas', 'nathawaiian\_death', 'new\_f\_deaths', 'new\_m\_deaths', 'ohny', 'wdr', and 'white\_deaths'. The Console pane at the bottom shows the output of the code, including a Chi-squared test result and a summary of an ANOVA model.

```

57 #}
58
59 death_rates <- data[, c("Race", "Death_yn")]
60 death_rates$Death_yn <- as.numeric(lapply(death_rates$Death_yn, mod2))
61 head(death_rates)
62 shapiro.test(death_rates$Death_yn)
63 print("passes shapiro wilks")
64
65 chisq.test(drs)
66 #leveneTest(Death_yn ~ Race, data=death_rates)
67 model <- aov(Death_yn ~ Race, data=death_rates)
68 summary(model)
69
70
71 (Top Level)

```

Console Output:

```

~/Documents/school/MATHSTATS/final />
Chi-squared test for given probabilities

data:  drs
X-squared = 0.019888, df = 5, p-value = 1

> #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

> #print("Since the p-value is less than 0.364, the death rate is not
> #      significant")
Warning message:
In chisq.test(drs) : Chi-squared approximation may be incorrect
>

```

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.
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.

The screenshot shows the RStudio IDE interface. The main editor displays an R script with the following code:

```

1 data <- read.csv('Final4.csv')
2
3 getLabel <- function(x,s) x[x$Age == s,]
4 deaths <- function(x) as.numeric(lapply(x$Death_yn, function(s) if(s=="Yes") 1 else 0))
5 death_rate <- function(L) sum(L[L == 1])/length(L)
6 process <- function(x,s) death_rate(deaths(getLabel(x,s)))
7
8 d3039 <- process(data, "30 - 39 Years")
9 d1019 <- process(data, "10 - 19 Years")
10 d4049 <- process(data, "40 - 49 Years")
11 d7079 <- process(data, "50 - 59 Years")
12 d2029 <- process(data, "20 - 29 Years")
13 d09 <- process(data, "0 - 9 Years")
14 d6069 <- process(data, "60 - 69 Years")
15 d80p <- process(data, "80+ Years")
16
17 max(d3039, d1019, d4049, d7079, d2029, d09, d6069, d80p)
18 print('The age group with the highest death rate is 80+')
19
20 death_rates <- data[, c("Age", "Death_yn")]
21 death_rates$Death_yn <- as.numeric(lapply(death_rates$Death_yn, mod2))
22 model <- aov(Death_yn ~ Age, data=death_rates)
23
24 summary(model)

```

The Environment pane on the right shows the following functions:

Object	Class
death_rate	function (L)
deaths	function (x)
getLabel	function (x, s)
mod2	function (s)
mod2sex	function (s)
notna	function (x)
process	function (x, s)

The Console pane at the bottom shows the output of the R script:

```

> 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

```

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.

- 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.

Activities RStudio Apr 15 16:50 96%

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins Project: (None)

p1.r x p2.r x p3.r x p4.r x p5.r x practice.r x

Source on Save Run Source

```

1 data <- read.csv("Final4.csv")
2 ser <- data[, c("Cholesterol", "Ser")]
3 nser <- ser[ser$Ser>0,]
4 str(nser)
5 chisq.test(nser)
6
7 print("Since the p-value is less than 0.05, we
8     reject the null hypothesis that there is
9     no association between cholesterol and seriability")
10

```

10:1 (Top Level) R Script

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 there is \n no association bet
ween cholesterol and seriability"
> |

```

Environment History Connections Tutorial

R Global Environment

Object	Class	Size
ser	data.frame	3600 obs. of 2 variables
sex	data.frame	1700 obs. of 2 variables
types_of_dd	list	List of 4
white	data.frame	1946 obs. of 16 variables
x3039	data.frame	558 obs. of 16 variables

Values

Object	Value
adr	0.0411764705882353
andr	0.0714285714285714

Files Plots Packages Help Viewer

Home Find in Topic Refresh Help Topic

R Resources

- Learning R Online
- CRAN Task Views
- R on StackOverflow
- Getting Help with R

RStudio

- RStudio IDE Support
- RStudio Community Forum
- RStudio Cheat Sheets
- RStudio Tip of the Day
- RStudio Packages
- RStudio Products

Manuals

- An Introduction to R
- Writing R Extensions
- R Data Import/Export
- The R Language Definition
- R Installation and Administration
- R Internals

Reference