

SEIS631 Extra Credit

Ethan Wang

5/21/2019

Purpose of Study

The purpose of this study is to gain a better understanding of the overdose deaths in the United States. This data has been filtered utilizing the ICD-10 Version:2016 code X42

ICD-10 CODE: X42 Accidental poisoning by and exposure to narcotics and psychodysleptics [hallucinogens], not elsewhere classified

Incl.: cannabis (derivatives) cocaine codeine heroin lysergide [LSD] mescaline methadone morphine opium (alkaloids)

CDC 2015 Data set

I chose the CDC data set from Kaggle to pull sample data. Specifically the 2015 year dataset. To view the data please find the link here: https://www.kaggle.com/cdc/mortality#2015_data.csv.

Since this contained all the deaths for 2015, I randomly sampled 100 female and 100 male deaths from the population. These two samples will be what I will use for my tests.

Questions 1

I want to test whether or not males and females overdose at different ages. To conduct this test I will be using a **2-sample T-Test**.

Hypothesis 1

Null Hypothesis: H_0 : average male overdose age = average female overdose age $H_0: m_1 = m_2$

Alternative Hypothesis H_A : average male overdose age \neq average female overdose age $H_A: m_1 \neq m_2$

Two-Sample T-Test

```
load(file = "samplefemalegroup.RData")
load(file = "samplemalegroup.RData")
result <- t.test(samplefemalegroup$detail_age,samplemalegroup$detail_age, var.equal = TRUE)

result

##
## Two Sample t-test
##
## data: samplefemalegroup$detail_age and samplemalegroup$detail_age
## t = 2.2397, df = 198, p-value = 0.02622
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.4935618 7.7664382
## sample estimates:
## mean of x mean of y
##    43.89    39.76
```

Test Result

Test Statistic = 2.2397

Degree of Freedom = 198

P-value = 0.02622

95% confidence interval is 0.4935618 7.7664382

Interpretation of Data

Since the P-value = 0.02622 is less than the alpha of 0.05 we can reject the Null Hypothesis. Thus we do have evidence that men and woman have different average age for drug overdoses. We can be 95% confident that the difference in population means is between 0.4935618 and 7.7664382.

Questions 2

I want to test and see average age of overdoses differ among different education levels. ages. To conduct this test I will be using **ANOVA**.

Hypothesis 2

Null Hypothesis: H_0 : 8th grade or less = 9-12th, no diploma = high school graduate or GED completed = some college credit, but no degree = Associate degree = Bachelor's Degree = Master's degree = Doctorate or professional degree = unknown

$H_0: m_1 = m_2 = m_3 = m_4 = m_5 = m_6 = m_7 = m_8 = m_9$

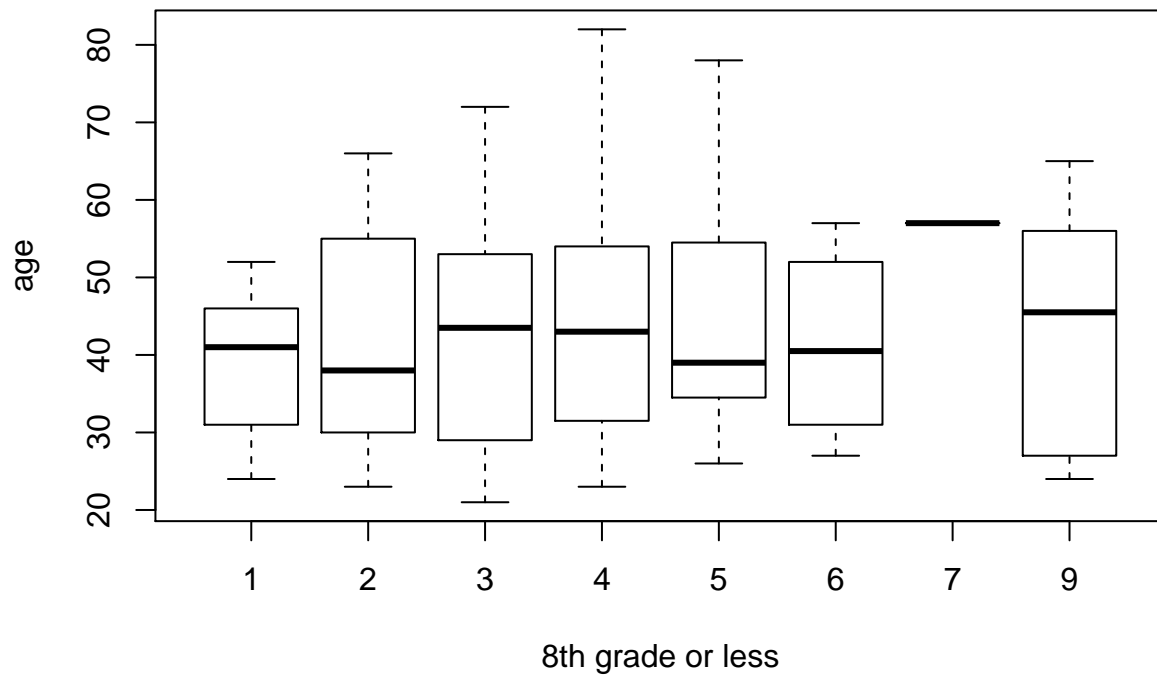
Alternative Hypothesis H_A : At least one of the means is different

ANOVA

```
combinedgroups <- bind_rows(samplemalegroup,samplefemalegroup)

boxplot(detail_age~education_2003_revision,
        data = combinedgroups,
        main = "Overdoses Among Different Education groups",
        xlab = "8th grade or less", "9-12th, no diploma",
        "high school graduate or GED completed", "some college credit, but no degree",
        "Associate degree", "Bachelor's Degree", "Master's degree",
        "Doctorate or professional degree", "unknown",
        ylab = "age")
```

Overdoses Among Different Education groups



```
edu1 <- lm(detail_age ~ education_2003_revision, data = combinedgroups)
anova(edu1)
```

```
## Analysis of Variance Table
##
## Response: detail_age
##
##              Df Sum Sq Mean Sq F value Pr(>F)
## education_2003_revision  1    159    159.02    0.916 0.3397
## Residuals              190  32985    173.60
```

Test Result

Degree of Freedom = 1 P-value = 0.3397

Interpretation of Data

Since the P-value = 0.3397 is more than the alpha of 0.05. Thus we fail to reject the Null Hypothesis. Thus we do not have evidence that different education groups have different average ages for drug overdoses. It currently appears that the average age of death for overdoses is the same among different education groups.