Project Part II

Carson Crenshaw (cgc8gdt)

An Investigation into the American Drug Epidemic

Introduction

The drug epidemic has been around for decades and continues to worsen in the United States. The country is struggling with one of its worst drug crises as overdose deaths remain a leading cause of injury-related death in the U.S. [1]. The majority of overdose deaths involve opioids: more than 1,500 people per week die from opioid related-overdoses [2].

The U.S. drug problem is a long-standing and deep-rooted disease that has not been cured. It is pertinent to complete an investigation of the drug epidemic's impact on the U.S. population in order to draw conclusions on the potential effect drug overdoses have had on the nation in recent years. In this regard, statistical hypothesis testing can help illuminate which populations are the most at risk, therefore aiding in the focus and allocation of community resources to those that are most in need.

The following question will be asked as an extension of this research goal: Is there an association between an individual's sex assigned at birth and the classification of overdose drug?

In order to generate the appropriate data set to answer the aforementioned research questions, data was collected from the National Center for Health Statistics (NCHS) and the Centers for Disease Control and Prevention (CDC) [3]. Using publicly stored data, the information is based on death certificates for U.S. residents spanning the years 1999-2021. The final data set contains both the age-adjusted rate of drug overdose deaths by race and the raw count of overdose deaths by sex. The data is further distinguished between six unique drug categories. Only the latter (raw counts by sex) will be used to answer the research question above.

Utilizing a chi-squared test for two-way tables, this report hopes to reveal a statistically significant relationship between sex assigned at birth and type of overdose drug. The result of this hypothesis test can then be used in an effort to combat the drug epidemic in the United States by highlighting demographic groups that may be disproportionately at risk.

Chi-Squared Test

The completion of a chi-squared test is derived from the conclusions within the researcher's previous exploratory data analysis of the drug overdose data set. By studying the raw counts and proportions of overdose deaths by sex in 2021, the EDA bar graphs revealed that while men experience higher numbers of overdose deaths than women in the U.S., the distribution of deaths between different types of drugs are relatively similar when considering sex. Based on these findings, one might jump to the conclusion that there is no outstanding relationship between sex and specific type of drug. The chi-squared test will therefore examine if the two variables are independent and will be able to prove if the conclusions derived from an observational analysis are statistically significant.

Non-parametric, categorical testing is chosen because it does not require the sampling distribution of the test statistic for the relevant parameter to be known or follow the normality assumption. Chi-square is also chosen over a more traditional correlation test because the characteristics of the values to be studied are categorical and not quantitative.

While the characteristics of the data above match that of a chi-squared test, the test assumptions must also be addressed. In order to use a chi-squared test, two assumptions must hold: (1) No table cells can contain zero individuals and (2) no more than 20% of the table cells can expect to have less than 5 individuals under the null hypothesis. Fortunately, both assumptions are met. Table 1 below illustrates the distribution of data from 2021 that will be used in the test.

Drug	Opioids	Heroin	Stimulants	Cocaine	Other
Female	23654	2372	15087	6858	17617
Male	56757	6801	38408	17628	33278

Table 1: Number of National Drug Overdose Deaths, by Sex (2021).

Data from 2021 was selected specifically because it was the most recent year which included no null values, satisfying the first assumption. Furthermore, 2021 holds the most amount of data because of the increase in drug use over the pandemic, satisfying the second assumption.

To examine statistically whether sex assigned at birth is related to drug overdose type, the question must be framed in terms of hypotheses. The null hypothesis is that the two variables are independent (i.e. no relationship or association) and the alternative hypothesis is that the two variables are related. In other words, depending on the sex of the victim, there will be a different distribution of proportions of drug overdose type. The specific hypotheses can be found below:

Ho: There is no association between sex and category of overdose drug

Ha: There is a association between sex and category of overdose drug

The test itself will be calculated using the R function chisq.test( ). The raw counts demonstrated in Table 1 will be compared to expected counts calculated for every cell. If there is no association, the counts in the expected table – where the same proportion of overdose

deaths between male and females were found for every drug type – should match the original table. In other words, if there was no relationship between the two variables, one would expect to see the number of drug overdoses for each sex to be evenly distributed across the drug types.

Results of Test and Conclusion

The expected table generated from the chi-squared test (when the null is true) is represented as Table 2 below. Even before an analysis of the p-value from the test, one can denote that there is an observable difference between the original and generated distribution of drug overdose death counts.

Drug	Opioids	Heroin	Stimulants	Cocaine	Other
Female	24141.7	2753.999	16060.74	7351.404	15280.15
Male	56269.3	6419.001	37434.26	17134.596	35614.85

Table 2: Expected Number of National Drug Overdose Deaths, by Sex (2021).

The chi-square statistic is then used to determine whether the difference initially observed is statistically significant. In comparing the observed values to the expected values, the test results in a p-value less than 2.2e-16 (3.705273e-157). By convention, social scientists often use a significance level of 0.05 ($p<0.05$) to determine if the observed differences are significant. Because the chi-squared p-value from this test is exceptionally close to zero, the null hypothesis is rejected and one can conclude that the differences are due to something other than random variation. In other words, the completion of a chi-squared test illustrates that there is a statistically significant association between sex assigned at birth and overdose drug category. This conclusion reverses the aforementioned assumption at the beginning of this report derived from the original exploratory data analysis.

A limitation of the chi-squared test, however, is that it can only show whether two variables are related to one another. It does not necessarily imply that one variable has any causal effect on the other. In order to establish causality, a more detailed analysis would be required in the future.

That being said, the conclusions drawn from this report provide enough evidence that there is an association between sex assigned at birth and overdose drug type. With this understanding, the U.S. should face its drug problem squarely, taking actions to deal with the domestic issue of drug abuse and protecting the American people's right to life by specifically catering to the observed difference between female and male victims of drug abuse.

Appendix

Cleaning and Merging Data

```
## Cleaning and Merging Data
dem_data <- read.csv('~/UVA Coursework/STAT3080/Project/Dem_Drug_Data.csv')
raw_data <- read.csv('~/UVA Coursework/STAT3080/Project/RawCounts_Drug_Data.csv')
drug_data <- merge(dem_data, raw_data, by="Year")
drug_data <- drug_data[1:73]
## Show the first few lines of the data set
head(drug_data)
```

```
##   Year Total_Overdose_Deaths.x TOD_Female.x TOD_Male.x TOD_White TOD_Black
## 1 1999                    6.1          3.9        8.2       6.2       7.5
## 2 2000                    6.2          4.1        8.3       6.6       7.3
## 3 2001                    6.8          4.6        9.0       7.4       7.6
## 4 2002                    8.2          5.8       10.6       9.2       8.2
## 5 2003                    8.9          6.4       11.5      10.2       8.2
## 6 2004                    9.4          6.9       11.8      11.0       8.3
##   TOD_Asian TOD_PI TOD_Hispanic TOD_AI Opioids.x O_Female.x O_Male.x O_White
## 1        NA     NA          5.4    6.0       2.9        1.4      4.3     2.8
## 2        NA     NA          4.6    5.5       3.0        1.6      4.4     3.1
## 3        NA     NA          4.5    6.9       3.3        1.9      4.8     3.7
## 4        NA     NA          5.4    8.5       4.1        2.6      5.7     4.7
## 5        NA     NA          5.6   10.8       4.5        2.8      6.1     5.2
## 6        NA     NA          5.2   12.5       4.7        3.1      6.3     5.7
##   O_Black O_Asian O_PI O_Hispanic O_AI Heroin.x H_Female.x H_Male.x H_White
## 1     3.5     0.3   NA        3.5  2.9      0.7        0.2      1.2     0.7
## 2     3.5     0.3   NA        2.7  2.7      0.7        0.2      1.1     0.6
## 3     3.3     0.3   NA        2.6  3.5      0.6        0.2      1.0     0.6
## 4     3.6     0.5   NA        3.2  4.1      0.7        0.2      1.2     0.7
## 5     3.5     0.3   NA        3.2  5.3      0.7        0.2      1.2     0.7
## 6     3.2     0.4   NA        2.9  6.2      0.6        0.2      1.1     0.7
##   H_Black H_Asian H_PI H_Hispanic H_AI Stimulants.x S_Female.x S_Male.x S_White
## 1     0.8      NA   NA        1.1   NA          1.5        0.7      2.3     1.2
## 2     0.9      NA   NA        0.9   NA          1.4        0.7      2.2     1.2
## 3     0.8      NA   NA        0.9   NA          1.5        0.8      2.3     1.3
## 4     0.9      NA   NA        1.0   NA          1.9        1.0      2.8     1.7
## 5     0.8      NA   NA        1.0   NA          2.1        1.1      3.2     2.0
## 6     0.6      NA   NA        0.7   NA          2.3        1.2      3.3     2.2
##   S_Black S_Asian S_PI S_Hispanic S_AI Cocaine.x C_Female.x C_Male.x C_White
## 1     3.7      NA   NA        1.8  1.1       1.4        0.6      2.1     1.0
## 2     3.4      NA   NA        1.4  1.2       1.3        0.6      1.9     1.0
```

```
## 3     3.6      NA   NA        1.5  1.4       1.3        0.7      2.0     1.0
## 4     4.0      NA   NA        1.8  1.6       1.6        0.8      2.4     1.3
## 5     4.2      NA   NA        1.9  2.4       1.8        0.9      2.7     1.6
## 6     4.4      NA   NA        1.8  2.7       1.9        1.0      2.8     1.7
##   C_Black C_Asian C_PI C_Hispanic C_AI Other.x Other_Female.x Other_Male.x
## 1     3.7      NA   NA        1.7  0.9     0.2            0.1          0.3
## 2     3.3      NA   NA        1.3  1.0     0.2            0.1          0.3
## 3     3.6     0.2   NA        1.3  1.0     0.2            0.1          0.3
## 4     4.0     0.2   NA        1.5  1.1     0.3            0.2          0.5
## 5     4.1     0.2   NA        1.6  1.7     0.4            0.2          0.6
## 6     4.3     0.2   NA        1.4  1.6     0.4            0.3          0.6
##   Other_White Other_Black Other_Asian Other_PI Other_Hispanic Other_AI
## 1         0.2         0.1          NA       NA            0.2       NA
## 2         0.2          NA          NA       NA            0.2       NA
## 3         0.2          NA         0.2       NA            0.2       NA
## 4         0.4         0.1         0.2       NA            0.3       NA
## 5         0.5         0.1         0.3       NA            0.4       NA
## 6         0.5         0.1         0.2       NA            0.4      1.3
##   Total_Overdose_Deaths.y TOD_Female.y TOD_Male.y Opioids.y O_Female.y O_Male.y
## 1                   16849         5591      11258      8050       2057     5993
## 2                   17415         5852      11563      8407       2264     6143
## 3                   19394         6736      12658      9496       2767     6729
## 4                   23518         8490      15028     11920       3760     8160
## 5                   25785         9386      16399     12940       4138     8802
## 6                   27424        10304      17120     13756       4643     9113
##   Heroin.y H_Female.y H_Male.y Stimulants.y S_Female.y S_Male.y Cocaine.y
## 1     1960        306     1654         4271        980     3291      3822
## 2     1842        279     1563         4017        980     3037      3544
## 3     1779        313     1466         4308       1083     3225      3833
## 4     2089        359     1730         5423       1400     4023      4599
## 5     2080        358     1722         6215       1626     4589      5199
## 6     1878        341     1537         6591       1767     4824      5443
##   C_Female.y C_Male.y Other.y Other_Female.y Other_Male.y
## 1        850     2972    3431           1504         1927
## 2        843     2701    3674           1628         2046
## 3        957     2876    4174           1775         2399
## 4       1143     3456    5333           2366         2967
## 5       1322     3877    5939           2622         3317
## 6       1405     4038    6690           3021         3669
```

Creating a Two-Way Table

```
## Create a two-way table for race demographics and overdose drug deaths for 2021
SexDataRaw <- matrix(c(23654,56757,2372,6801,15087,38408,6858,17628,17617,33278),
            nrow=2, ncol=5)
dimnames(SexDataRaw) <- list(Sex.at.Birth=c("Female","Male"),
                    Overdose.Drugs=c("Opioids","Heroin","Stimulants",
                                     "Cocaine", "Other"))
SexDataRaw
```

```
##             Overdose.Drugs
## Sex.at.Birth Opioids Heroin Stimulants Cocaine Other
##       Female   23654   2372      15087    6858 17617
##       Male     56757   6801      38408   17628 33278
```

Illustrating the Conditional Distributions (Not Referenced in Report)

```
## Conditional distribution (column totals)
sex.cond.drug <- prop.table(SexDataRaw,2)
sex.cond.drug
```

```
##             Overdose.Drugs
## Sex.at.Birth   Opioids   Heroin Stimulants   Cocaine    Other
##       Female 0.2941637 0.258585  0.2820264 0.2800784 0.346144
##       Male   0.7058363 0.741415  0.7179736 0.7199216 0.653856
```

```
## Conditional distribution (row totals)
sex.cond.drug2 <- prop.table(SexDataRaw,1)
sex.cond.drug2
```

```
##             Overdose.Drugs
## Sex.at.Birth   Opioids     Heroin Stimulants   Cocaine     Other
##       Female 0.3606452 0.03616515  0.2300268 0.1045618 0.2686010
##       Male   0.3712714 0.04448820  0.2512429 0.1153122 0.2176854
```

Chi-Squared Test

```
## Run the chi-squared test
sextest <- chisq.test(SexDataRaw, correct=FALSE)

## Print results
sextest
```

```
##
##  Pearson's Chi-squared test
##
## data:  SexDataRaw
## X-squared = 732.2, df = 4, p-value < 2.2e-16
```

```
sextest$expected
```

```
##              Overdose.Drugs
## Sex.at.Birth Opioids   Heroin Stimulants   Cocaine    Other
##       Female 24141.7 2753.999   16060.74  7351.404 15280.15
##       Male   56269.3 6419.001   37434.26 17134.596 35614.85
```

```
## When the value is too close to zero R will automatically default to 2.2e-16.
## To find the exact p-value, the code must call for it specifically.
sextest$p.value
```

```
## [1] 3.705273e-157
```

References

1. "Understanding Drug Overdoses and Deaths." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 14 Feb. 2022, https://www.cdc.gov/drugoverdose/epidemic/index.html.
2. "The U.S. Opioid Epidemic." Council on Foreign Relations, Council on Foreign Relations, https://www.cfr.org/backgrounder/us-opioid-epidemic#chapter-title-0-1.
3. "Multiple Cause of Death 2018-2021 by Single Race." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, https://wonder.cdc.gov/wonder/help/mcd-expanded.html#Frequently%20Asked%20Questions%20about%20Death%20Rates.