Data Analysis Section

Qiheng(Michael) Yan

2023-11-21

```
library(tidyverse)
library(stats)
library(ggplot2)
library(expss)
set.seed(1028)
```

Epi 3 Data Analysis

Import data

```
# Import the dataset
complete_data<-readRDS("combined_data.rds")
subset_1168_data<-readRDS("subset_1168.rds")
#write.csv(complete_data, "complete_data.csv", row.names=FALSE)
#write_labelled_csv(complete_data, "complete_data_labeled.csv", row.names=FALSE)
#write.csv(subset_1168_data, "subset_1168_data.csv", row.names=FALSE)
#write_labelled_csv(subset_1168_data, "subset_1168_data_labeled.csv", row.names=FALSE)
```

Bivariate Analysis

In bivariate analysis, we'll use Chi-squared tests for categorical variables like health insurance status, race/ethnicity, gender, and a T-test for continuous variables like age.

```
# Chi-squared test for medication adherence and family income
chisq_income <- chisq.test(table(complete_data$adherence, complete_data$income_cat))

# Chi-squared test for medication adherence and total insurance status
chisq_insurance <- chisq.test(table(complete_data$adherence, complete_data$ins_classif))
```

```
## Warning in chisq.test(table(complete_data$adherence,
## complete_data$ins_classif)): Chi-squared approximation may be incorrect
```

```
# Chi-squared test for medication adherence and race
chisq_race <- chisq.test(table(complete_data$adherence, complete_data$race_6cat))

# Chi-squared test for medication adherence and sex
chisq_gender <- chisq.test(table(complete_data$adherence, complete_data$sex))
```

```r
# T-test for medication adherence and age
t_test_age <- t.test(complete_data$age ~ complete_data$adherence)

# Print the results
print(chisq_income)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(complete_data$adherence, complete_data$income_cat)
## X-squared = 8.3414, df = 2, p-value = 0.01544
```

```r
print(chisq_insurance)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(complete_data$adherence, complete_data$ins_classif)
## X-squared = 54.283, df = 8, p-value = 6.085e-09
```

```r
print(chisq_race)
```

```
##
##  Pearson's Chi-squared test
##
## data:  table(complete_data$adherence, complete_data$race_6cat)
## X-squared = 15.478, df = 5, p-value = 0.008503
```

```r
print(chisq_gender)
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(complete_data$adherence, complete_data$sex)
## X-squared = 2.4827, df = 1, p-value = 0.1151
```

```r
print(t_test_age)
```

```
##
##  Welch Two Sample t-test
##
## data:  complete_data$age by complete_data$adherence
## t = -11.074, df = 901.08, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group FALSE and group TRUE is not equal to 
## 95 percent confidence interval:
##  -9.332160 -6.522391
## sample estimates:
## mean in group FALSE  mean in group TRUE
##            56.91137            64.83864
```

The results from the bivariate analyses provide valuable insights into the relationships between adherence to prescribed cholesterol medication and various factors such as family income, health insurance status, race/ethnicity, gender, and age.

1. Family Income vs. Adherence: Chi-squared test result:

   $$X^2 = 8.3414, df = 2, p - value = 0.01544$$

   Interpretation: There is a statistically significant association between family income category and adherence to cholesterol medication. Since the p-value is less than 0.05, we can conclude that the differences in adherence rates across different income categories are not due to random chance.

2. Health Insurance Status vs. Adherence: Chi-squared test result:

   $$X^2 = 54.283, df = 8, p - value = 6.085 \times 10^{-9}$$

   Interpretation: There is a highly statistically significant association between health insurance classification and medication adherence. The extremely low p-value indicates strong evidence against the null hypothesis of no association.

3. Race/Ethnicity vs. Adherence: Chi-squared test result:

   $$X^2 = 15.478, df = 5, p - value = 0.008503$$

   Interpretation: Race/ethnicity shows a statistically significant association with medication adherence. The p-value below 0.05 suggests that different racial/ethnic groups have different adherence rates to cholesterol medication.

4. Gender vs. Adherence: Chi-squared test result:

   $$X^2 = 2.4827, df = 1, p - value = 0.1151$$

   Interpretation: There is no statistically significant association between gender and medication adherence. The p-value is greater than 0.05, indicating that any observed differences in adherence between genders could be due to chance.

5. Age vs. Adherence: T-test result:

   $$t = -11.074, df = 901.08, p - value = 2.2 \times 10^{-16}$$

   Interpretation: There is a highly statistically significant difference in the average age between those who adhere to their medication and those who do not. The negative t-value indicates that the mean age of the group adhering to the medication (mean = 64.84 years) is higher than that of the non-adhering group (mean = 56.91 years). The extremely low p-value provides strong evidence against the null hypothesis of no difference in means.

These results suggest that socioeconomic factors (like income and insurance status), as well as demographic characteristics (like race/ethnicity and age), are associated with adherence to cholesterol medication. Gender, however, does not seem to show a significant association in this context. These findings can inform further multivariate analysis to understand the independent effect of each factor while controlling for others.

Multivariate Analysis