R_Sai_Omkar_K_PS8

Sai Omkar Kandukuri

08/12/2021

```
rm(list=ls())

library(haven)
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --

## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.4     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(boot)
library(copula)
```

```
## Warning: package 'copula' was built under R version 4.1.2
```

```
data <- read_dta("E:/Autumn'21/Advanced_Stats/ProblemSets/8/homework_8.dta")
```

```
data_treated <- data %>% filter(data$treated == 1)
data_untreated <- data %>% filter(data$treated == 0)

data_treated_work <- c(data_treated$work)
data_untreated_work <- c(data_untreated$work)
```

```
suppressMessages(library(dplyr))


#TTest two sided

t.test(data_treated_work, data_untreated_work, alternative = "two.sided", var.equal = FALSE, conf.level
```

1A. Use the t-test command;

```
##
##  Welch Two Sample t-test
##
## data:  data_treated_work and data_untreated_work
## t = 2.3621, df = 672.37, p-value = 0.01846
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.01315123 0.14272914
## sample estimates:
## mean of x mean of y
## 0.7744108 0.6964706
```

```
#Method 2
t.test(work~treated, alternative="two.sided", data=data)
```

```
##
##  Welch Two Sample t-test
##
## data:  work by treated
## t = -2.3621, df = 672.37, p-value = 0.01846
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -0.14272914 -0.01315123
## sample estimates:
## mean in group 0 mean in group 1
##       0.6964706       0.7744108
```

```
#TTest one sided
t.test(data_treated_work, data_untreated_work, alternative = "greater", var.equal = FALSE, conf.level =
```

```
##
##  Welch Two Sample t-test
##
## data:  data_treated_work and data_untreated_work
## t = 2.3621, df = 672.37, p-value = 0.009229
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.02359057        Inf
## sample estimates:
## mean of x mean of y
## 0.7744108 0.6964706
```

```
#Observed p-value for the t-test with a 95% confidence interval is 0.01846. This value is less than 0.0
```

```
#ChiSquare Test Question 1B
data_work_treat_type <- data %>%
  group_by(treated, work) %>%
  count(treated, work) %>%
  group_by(treated) %>%
  mutate(treat_count = sum(n), perc_freq = round(n / sum(n), 3)*100) %>%
  filter(work == 1)
data_work_treat_type
```

```
## # A tibble: 2 x 5
## # Groups:   treated [2]
##   treated  work     n treat_count perc_freq
##     <dbl> <dbl> <int>       <int>     <dbl>
## 1       0     1   296         425      69.6
## 2       1     1   230         297      77.4
```

```
chisq_data <- matrix(c(230,67,296,129), nrow=2)
chisq_test <- chisq.test(chisq_data)
chisq_test
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  chisq_data
## X-squared = 4.983, df = 1, p-value = 0.0256
```

```
chisq_test$residuals
```

```
##             [,1]       [,2]
## [1,]  0.926333 -0.7743742
## [2,] -1.517511  1.2685737
```

```
#Method 2
chisq.test(table(data$work, data$treated))
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(data$work, data$treated)
## X-squared = 4.983, df = 1, p-value = 0.0256
```

```
#Observed p-value here is 0.0256 for this test. As the observed p-value is less than 0.05, we can rejec
```

```
#Fischer Test
fisher.test(chisq_data, alternative = "two.sided", conf.level = 0.95)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  chisq_data
## p-value = 0.02176
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.049737 2.142052
## sample estimates:
## odds ratio
##   1.495233
```

```r
#Method 2
fisher.test(table(data$work, data$treated))
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  table(data$work, data$treated)
## p-value = 0.02176
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  1.049737 2.142052
## sample estimates:
## odds ratio
##   1.495233
```

```r
#Observed p-value is 0.02176 for the two-tail fisher exact test. As this p-value is less than 0.05, we
```

```r
##Why do the p-values on the two-sided test differ? Which should you believe?

#We observed the following p values t-test : 0.0185 , Chi-square test : 0.02, Fisher exact test : 0.021
```

###Question 2

```r
y_uni = lm(work ~ treated, data)

summary(y_uni)
```

```
##
## Call:
## lm(formula = work ~ treated, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7744 -0.6965  0.2256  0.3035  0.3035
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.69647    0.02152  32.362   <2e-16 ***
## treated      0.07794    0.03356   2.323   0.0205 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4437 on 720 degrees of freedom
## Multiple R-squared:  0.007437,   Adjusted R-squared:  0.006059
## F-statistic: 5.395 on 1 and 720 DF,  p-value: 0.02047
```

```r
#Observed p-value is 0.02047. Less than 0.05 . So, we can reject the null hypothesis and not the Altern

y_multivariate = lm(work ~ treated + age + educ + black + hisp + married, data)
summary(y_multivariate)
```

```
##
## Call:
## lm(formula = work ~ treated + age + educ + black + hisp + married,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.9664 -0.6295  0.2477  0.3246  0.4679
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.896116   0.131574   6.811 2.06e-11 ***
## treated      0.077544   0.033221   2.334  0.01986 *
## age         -0.003375   0.002547  -1.325  0.18565
## educ         0.002398   0.009730   0.246  0.80544
## black       -0.185623   0.056309  -3.297  0.00103 **
## hisp        -0.012145   0.074278  -0.164  0.87016
## married      0.052648   0.045494   1.157  0.24756
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4383 on 715 degrees of freedom
## Multiple R-squared:  0.03803,    Adjusted R-squared:  0.02996
## F-statistic: 4.711 on 6 and 715 DF,  p-value: 0.0001024
```

```
###We observed a higher adjusted R square value. This means that the additional input variables are add

#Univariate  R-squared:  0.007437,  Adjusted R-squared:  0.006059 F-statistic: 5.395 on 1 and 720 DF,  p
#Multivariate R-squared:  0.03803,  Adjusted R-squared:  0.02996 F-statistic: 4.711 on 6 and 715 DF,  p

#As seen above, the treatment indicator Coeff/estimate decreases slightly, but however it can still be

#p-value t-test : 0.0185 , Chi-square test : 0.02, Fisher exact test : 0.0218
#p-value for univariate model : 0.0205 and 0.0001 for multivariate. Comparing the p-values , regression
```

```
f_treat <- data %>%  filter(treated == 1) %>% count(work)
f_treat %>%  mutate (prob = n/sum(n))
```

```
## # A tibble: 2 x 3
##    work     n  prob
##   <dbl> <int> <dbl>
## 1     0    67 0.226
## 2     1   230 0.774
```

```
f_control <- data %>%  filter(treated == 0) %>% count(work)
f_control %>%  mutate (prob = n/sum(n))
```

```
## # A tibble: 2 x 3
##    work     n  prob
##   <dbl> <int> <dbl>
## 1     0   129 0.304
## 2     1   296 0.696
```

```r
#People who benefots from treatment : Bounds of joint distribution are [0.07, 0.30]
#People who loses from the treatment : Bounds of joint distribution are [0.00, 0.23]
```

```r
#From given information
#             Not affected       Affected
# Moderna       14329              269
# Pfizer        21569              100

#Fisher
Fisher_data <- matrix(c(14329, 269, 21569, 100), nrow=2)
fisher.test(Fisher_data, alternative = "two.sided", conf.level = 0.95)
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  Fisher_data
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.1941302 0.3121725
## sample estimates:
## odds ratio
##  0.2469742
```

```r
#Observed p-value < 2.2e-16 two-tail fisher exact test is less than 0.05. So we can reject the null hyp




#B These are the actually numbers from the treated observations of the Moderna and Pfizer clinical tria

#Two arguments here. One is we donot have data of a person in two cases where in one case a vaccine is
```