

Birth Weight

Team C

11/03/2021

Needed Packages

```
if(!require(FSA)){install.packages("FSA")}
if(!require(ggplot2)){install.packages("ggplot2")}
if (!require("mosaic")) install.packages("mosaic", dep=FALSE)
if (!require("nortest")) install.packages("nortest", dep=TRUE)
if (!require("epitools")) install.packages("epitools", dep=TRUE)
if (!require("prettyR")) install.packages("prettyR", dep=TRUE)
if (!require("rms")) install.packages("rms", dep=TRUE)
library(tidyverse)
# add other as needed
```

Problem – Infant Birth Weight Data

Read data into R

```
# this data came from SASHELP.BWEIGHT
bw <- read.csv('bwgt.csv')
bw <- data.frame(bw)
#summary(bw)
bw = transform(bw, AgeGroup.f = as.factor(AgeGroup))
bw = transform(bw, Race.f = as.factor(Race))
bw = transform(bw, Drinking.f = as.factor(Drinking))
bw = transform(bw, Death.f = as.factor(Death))
bw = transform(bw, Smoking.f = as.factor(Smoking))
bw = transform(bw, SomeCollege.f = as.factor(SomeCollege))
bw = transform(bw, LowBirthWgt.f = as.factor(bw$LowBirthWgt))
```

Code 1 - Data Summary

```
summary(bw)
```

```
## LowBirthWgt      Married      AgeGroup      Race
## Length:3089      Length:3089      Min. :1.000      Length:3089
## Class :character  Class :character  1st Qu.:2.000      Class :character
## Mode :character  Mode :character  Median :2.000      Mode :character
##                                     Mean :2.021
##                                     3rd Qu.:2.000
##                                     Max. :3.000
##      Drinking      Death      Smoking      SomeCollege
## Length:3089      Length:3089      Length:3089      Length:3089
## Class :character  Class :character  Class :character  Class :character
## Mode :character  Mode :character  Mode :character  Mode :character
##
##
##
## AgeGroup.f      Race.f      Drinking.f  Death.f      Smoking.f  SomeCollege.f
## 1: 357      Asian : 145      : 197      No :2483      : 197      : 245
## 2:2310      Black : 518      No :2493      Yes: 606      No :2191      No :1526
## 3: 422      Hispanic: 624      Yes: 399      Yes: 701      Yes:1318
##      Native : 32
##      White :1770
##
## LowBirthWgt.f
## No :2476
## Yes: 613
##
##
##
##
```

```
summary(bw$LowBirthWgt)
```

```
##      Length      Class      Mode
##      3089 character character
```

```
summary(bw$Smoking)
```

```
##      Length      Class      Mode
##      3089 character character
```

```
mytab <- tally(~ Race.f | AgeGroup.f, data=bw)
addmargins(mytab)
```

```
##      AgeGroup.f
## Race.f      1      2      3      Sum
## Asian      8     101    36    145
## Black     91    375    52    518
## Hispanic  83    475    66    624
## Native     6     22     4     32
## White    169   1337   264   1770
## Sum      357   2310   422   3089
```

```
mytab1 <- tally(~ Married | AgeGroup.f, data=bw)
addmargins(mytab1)
```

```
##           AgeGroup.f
## Married      1      2      3 Sum
##      No      61 1533  351 1945
##      Yes     296  777   71 1144
##      Sum     357 2310  422 3089
```

```
mytab2 <- tally(~ LowBirthWgt.f | Race.f, data=bw)
addmargins(mytab2)
```

```
##           Race.f
## LowBirthWgt.f Asian Black Hispanic Native White Sum
##           No      117   342      523      26 1468 2476
##           Yes       28   176      101       6  302  613
##           Sum      145   518      624      32 1770 3089
```

Code 2

```
bw <- read.csv('bwgt.csv')
# Create a new smaller data set of size 2500
set.seed(1568)
ssize <- sample(1:nrow(bw),round(nrow(bw) * .8092))
new_bwgt <- bw %>% slice(ssize)
new_bwgt <-new_bwgt %>% na_if("") %>% na.omit

# Make the following columns
new_bwgt <- new_bwgt %>% mutate( AgeGroup.f = factor(AgeGroup), Race.f = factor(Race),
Drinking.f = factor(Drinking), Death.f = factor(Death),Smoking.f = factor(Smoking),
SomeCollege.f = factor(SomeCollege), LowBirthWgt.f = factor(LowBirthWgt))
```

Test for Association between Low Birth Weight and smoking

```
#Risk ratio
RR1<-riskratio(x=new_bwgt$Smoking.f, y=new_bwgt$LowBirthWgt.f)
RR1$measure
```

```
##           risk ratio with 95% C.I.
## Predictor estimate      lower      upper
##      No      1.00000          NA        NA
##      Yes     1.13645 0.9433483 1.369078
```

```
RR1$p.value
```

```
##           two-sided
## Predictor midp.exact fisher.exact chi.square
##      No      NA          NA          NA
##      Yes     0.1838204    0.1963432 0.1814622
```

Since 95% C.I. does contain the value of 1, which is the value of no smoking. So, there is insufficient evidence to conclude that the two groups are statistically significantly different. Also, the p.values can confirm our found.

```
#chisq test
chisq.test(new_bwgt$Smoking.f,new_bwgt$LowBirthWgt.f)
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  new_bwgt$Smoking.f and new_bwgt$LowBirthWgt.f
## X-squared = 1.625, df = 1, p-value = 0.2024
```

The chi-squared result shows that there is not enough evidence to say the low birth weight is causing by the smoking.

```
#oddsratio
oddsratio(new_bwgt$Smoking.f, new_bwgt$LowBirthWgt.f,rev=c("b"),method="wald")$measure
```

```
##           odds ratio with 95% C.I.
## Predictor estimate      lower      upper
##      Yes 1.000000          NA        NA
##      No  1.174036 0.9276977 1.485787
```

Even though the odd ratio of the non-smoking group is greater than one, the value is just 1.17. Also, the 95% C.I. does include 1, so our conclusion will be the same as above. We cannot say when women is smoking, it is more likely to deliver a low birth weight baby.

```
# Fisher's Exact Test
tab22<- table(x=new_bwgt$Smoking.f , y=new_bwgt$LowBirthWgt.f)
tab22<- as.matrix(tab22)
```

```
# Two-sided alternative hypothesis
fisher.test(tab22)
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  tab22
## p-value = 0.1963
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.9190588 1.4931839
## sample estimates:
## odds ratio
##  1.173928
```

```
# One-sided less than alternative hypothesis
fisher.test(tab22,alternative="less")
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  tab22
## p-value = 0.918
## alternative hypothesis: true odds ratio is less than 1
## 95 percent confidence interval:
##  0.0000000 1.438835
## sample estimates:
## odds ratio
##  1.173928
```

```
# One-sided greater than alternative hypothesis
fisher.test(tab22,alternative="greater")
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  tab22
## p-value = 0.1019
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  0.955358      Inf
## sample estimates:
## odds ratio
##  1.173928
```

All these three will lead to the same conclusion as we mentioned above, that is, there is no significant association at $\alpha = 0.5$ between smoking and low birth weight.

```
# MH test:
mantelhaen.test(new_bwgt$Smoking.f, new_bwgt$LowBirthWgt.f, new_bwgt$Drinking.f, correct=T)

##
## Mantel-Haenszel chi-squared test with continuity correction
##
## data: new_bwgt$Smoking.f and new_bwgt$LowBirthWgt.f and new_bwgt$Drinking.f
## Mantel-Haenszel X-squared = 5.9382, df = 1, p-value = 0.01482
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
## 1.089772 2.005459
## sample estimates:
## common odds ratio
## 1.478341
```

Test for Association between Low Birth Weight and drinking

```
#Risk ratio
RR1<-riskratio(x=new_bwgt$Drinking.f, y=new_bwgt$LowBirthWgt.f)
RR1$measure
```

```
## risk ratio with 95% C.I.
## Predictor estimate lower upper
## No 1.0000000 NA NA
## Yes 0.9057919 0.7032788 1.16662
```

```
RR1$p.value
```

```
## two-sided
## Predictor midp.exact fisher.exact chi.square
## No NA NA NA
## Yes 0.4434666 0.4896086 0.4393742
```

```
#chisq test
chisq.test(new_bwgt$Drinking.f, new_bwgt$LowBirthWgt.f)
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: new_bwgt$Drinking.f and new_bwgt$LowBirthWgt.f
## X-squared = 0.48508, df = 1, p-value = 0.4861
```

```
#oddsratio
oddsratio(new_bwgt$Drinking.f, new_bwgt$LowBirthWgt.f, rev=c("b"), method="wald")$measure
```

```
## odds ratio with 95% C.I.
## Predictor estimate lower upper
```

```

##      Yes 1.000000      NA      NA
##      No  0.885103 0.6494903 1.206188

# Fisher's Exact Test
tab23<- table(x=new_bwgt$Drinking.f, y=new_bwgt$LowBirthWgt.f)
tab22<- as.matrix(tab23)

# Two-sided alternative hypothesis
fisher.test(tab22)

##
## Fisher's Exact Test for Count Data
##
## data:  tab22
## p-value = 0.4896
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.6372546 1.2133958
## sample estimates:
## odds ratio
##  0.885151

# One-sided less than alternative hypothesis
fisher.test(tab22,alternative="less")

##
## Fisher's Exact Test for Count Data
##
## data:  tab22
## p-value = 0.2449
## alternative hypothesis: true odds ratio is less than 1
## 95 percent confidence interval:
##  0.0000000 1.157093
## sample estimates:
## odds ratio
##  0.885151

# One-sided greater than alternative hypothesis
fisher.test(tab22,alternative="greater")

##
## Fisher's Exact Test for Count Data
##
## data:  tab22
## p-value = 0.8015
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  0.6714868      Inf
## sample estimates:
## odds ratio
##  0.885151

# MH test:
mantelhaen.test(new_bwgt$Drinking.f, new_bwgt$LowBirthWgt.f, new_bwgt$Smoking.f,correct=T)

##
## Mantel-Haenszel chi-squared test with continuity correction

```

```
##
## data:  new_bwgt$Drinking.f and new_bwgt$LowBirthWgt.f and new_bwgt$Smoking.f
## Mantel-Haenszel X-squared = 4.4378, df = 1, p-value = 0.03515
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.4277696 0.9516248
## sample estimates:
## common odds ratio
##          0.6380252
```

All 5 tests' outputs give the same result, at $\alpha = 0.5$, there is no strong evidence to conclude the relation between having a low birth weight baby and drinking is significant.

Code 3

Test for Association between Low Birth Weight and Smoking controlling for Death

```
new_bwgt1<-table(new_bwgt)
tab1<-margin.table(new_bwgt1, c(15,13,12))
d_yes<-tab1[, , Death.f = "No"]
d_no<-tab1[, , Death.f = "Yes"]

#Risk ratio for Death= Yes
RR1<-riskratio(d_yes)
RR1$measure

##                risk ratio with 95% C.I.
## LowBirthWgt.f estimate      lower      upper
##           No  1.000000         NA         NA
##           Yes 1.162103 0.8812414 1.532478

RR1$p.value

##                two-sided
## LowBirthWgt.f midp.exact fisher.exact chi.square
##           No         NA         NA         NA
##           Yes 0.2991542 0.3191694 0.2964308

#Risk ratio for Death= No
RR1<-riskratio(d_no)
RR1$measure

##                risk ratio with 95% C.I.
## LowBirthWgt.f estimate      lower      upper
##           No  1.000000         NA         NA
##           Yes 0.8456945 0.6186096 1.15614

RR1$p.value

##                two-sided
## LowBirthWgt.f midp.exact fisher.exact chi.square
##           No         NA         NA         NA
##           Yes 0.3015446 0.3031373 0.2982256

#chisq test for Death= Yes
chisq.test(d_yes)

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  d_yes
## X-squared = 0.89219, df = 1, p-value = 0.3449

#chisq test for Death= No
chisq.test(d_no)

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  d_no
## X-squared = 0.8562, df = 1, p-value = 0.3548
```

```
#oddsratio for Death= Yes
oddsratio(d_yes,rev=c("b"),method="wald")$measure
```

```
##           odds ratio with 95% C.I.
## LowBirthWgt.f estimate      lower      upper
##           Yes 1.000000          NA          NA
##           No  1.220403 0.8392721 1.774614
```

```
#oddsratio for Death= No
oddsratio(d_no,rev=c("b"),method="wald")$measure
```

```
##           odds ratio with 95% C.I.
## LowBirthWgt.f estimate      lower      upper
##           Yes 1.000000          NA          NA
##           No  0.7904843 0.5073498 1.231626
```

```
# Fisher's Exact Test for Death=Yes
# Two-sided alternative hypothesis
fisher.test(d_yes)
```

```
##
## Fisher's Exact Test for Count Data
##
## data: d_yes
## p-value = 0.3192
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.8174221 1.7929076
## sample estimates:
## odds ratio
##  1.220293
```

```
# One-sided less than alternative hypothesis
fisher.test(d_yes,alternative="less")
```

```
##
## Fisher's Exact Test for Count Data
##
## data: d_yes
## p-value = 0.8728
## alternative hypothesis: true odds ratio is less than 1
## 95 percent confidence interval:
##  0.00000 1.69251
## sample estimates:
## odds ratio
##  1.220293
```

```
# One-sided greater than alternative hypothesis
fisher.test(d_yes,alternative="greater")
```

```
##
## Fisher's Exact Test for Count Data
##
## data: d_yes
## p-value = 0.172
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
```

```

## 0.8708259      Inf
## sample estimates:
## odds ratio
## 1.220293

# Fisher's Exact Test for Death=No
# Two-sided alternative hypothesis
fisher.test(d_no)

##
## Fisher's Exact Test for Count Data
##
## data: d_no
## p-value = 0.3031
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.4968447 1.2674330
## sample estimates:
## odds ratio
## 0.7909397

# One-sided less than alternative hypothesis
fisher.test(d_no,alternative="less")

##
## Fisher's Exact Test for Count Data
##
## data: d_no
## p-value = 0.1772
## alternative hypothesis: true odds ratio is less than 1
## 95 percent confidence interval:
## 0.0000000 1.178722
## sample estimates:
## odds ratio
## 0.7909397

# One-sided greater than alternative hypothesis
fisher.test(d_no,alternative="greater")

##
## Fisher's Exact Test for Count Data
##
## data: d_no
## p-value = 0.8756
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
## 0.532999      Inf
## sample estimates:
## odds ratio
## 0.7909397

# MH test:
mantelhaen.test(tab1,correct=T)

##
## Mantel-Haenszel chi-squared test with continuity correction
##

```

```
## data:  tab1
## Mantel-Haenszel X-squared = 0.0010773, df = 1, p-value = 0.9738
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.7604251 1.3570002
## sample estimates:
## common odds ratio
##          1.015823
```

Test for Association between Low Birth Weight and Drinking controlling for Death

```
tab2<-margin.table(new_bwgt1, c(15,11,12))
d_yes1<-tab2[, , Death.f = "No"]
d_no1<-tab2[, , Death.f = "Yes"]
```

#Risk ratio for Death= Yes

```
RR1<-riskratio(d_yes1)
RR1$measure
```

```
##          risk ratio with 95% C.I.
## LowBirthWgt.f estimate      lower      upper
##          No  1.000000          NA          NA
##          Yes 1.121268 0.7618986 1.650145
```

```
RR1$p.value
```

```
##          two-sided
## LowBirthWgt.f midp.exact fisher.exact chi.square
##          No          NA          NA          NA
##          Yes 0.5566217    0.5458626 0.5644661
```

#Risk ratio for Death= No

```
RR1<-riskratio(d_no1)
RR1$measure
```

```
##          risk ratio with 95% C.I.
## LowBirthWgt.f estimate      lower      upper
##          No  1.000000          NA          NA
##          Yes 0.7852063 0.4619528 1.334658
```

```
RR1$p.value
```

```
##          two-sided
## LowBirthWgt.f midp.exact fisher.exact chi.square
##          No          NA          NA          NA
##          Yes 0.377272    0.4237677 0.3729685
```

#chisq test for Death= Yes

```
chisq.test(d_yes1)
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  d_yes1
```

```
## X-squared = 0.20717, df = 1, p-value = 0.649
#chisq test for Death= No
chisq.test(d_no1)

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: d_no1
## X-squared = 0.53419, df = 1, p-value = 0.4648
#oddsratio for Death= Yes
oddsratio(d_yes1,rev=c("b"),method="wald")$measure

##          odds ratio with 95% C.I.
## LowBirthWgt.f estimate      lower      upper
##      Yes 1.000000          NA          NA
##      No  1.143486 0.7244916 1.804796
#oddsratio for Death= No
oddsratio(d_no1,rev=c("b"),method="wald")$measure

##          odds ratio with 95% C.I.
## LowBirthWgt.f estimate      lower      upper
##      Yes 1.000000          NA          NA
##      No  0.7591707 0.4135274 1.393717
# Fisher's Exact Test for Death=Yes
# Two-sided alternative hypothesis
fisher.test(d_yes1)

##
## Fisher's Exact Test for Count Data
##
## data: d_yes1
## p-value = 0.5459
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.6922058 1.8217709
## sample estimates:
## odds ratio
##  1.143438
# One-sided less than alternative hypothesis
fisher.test(d_yes1,alternative="less")

##
## Fisher's Exact Test for Count Data
##
## data: d_yes1
## p-value = 0.7611
## alternative hypothesis: true odds ratio is less than 1
## 95 percent confidence interval:
##  0.000000 1.702521
## sample estimates:
## odds ratio
##  1.143438
```

```
# One-sided greater than alternative hypothesis
fisher.test(d_yes1,alternative="greater")
```

```
##
## Fisher's Exact Test for Count Data
##
## data: d_yes1
## p-value = 0.3177
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  0.7501199      Inf
## sample estimates:
## odds ratio
##  1.143438
```

```
# Fisher's Exact Test for Death=No
# Two-sided alternative hypothesis
fisher.test(d_no1)
```

```
##
## Fisher's Exact Test for Count Data
##
## data: d_no1
## p-value = 0.4238
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.398984 1.479963
## sample estimates:
## odds ratio
##  0.759641
```

```
# One-sided less than alternative hypothesis
fisher.test(d_no1,alternative="less")
```

```
##
## Fisher's Exact Test for Count Data
##
## data: d_no1
## p-value = 0.2304
## alternative hypothesis: true odds ratio is less than 1
## 95 percent confidence interval:
##  0.00000 1.33571
## sample estimates:
## odds ratio
##  0.759641
```

```
# One-sided greater than alternative hypothesis
fisher.test(d_no1,alternative="greater")
```

```
##
## Fisher's Exact Test for Count Data
##
## data: d_no1
## p-value = 0.8531
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
```

```
## 0.4386242      Inf
## sample estimates:
## odds ratio
## 0.759641
```

```
# MH test:
mantelhaen.test(tab2,correct=T)
```

```
##
## Mantel-Haenszel chi-squared test without continuity correction
##
## data:  tab2
## Mantel-Haenszel X-squared = 0.0060663, df = 1, p-value = 0.9379
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
## 0.6800377 1.4286344
## sample estimates:
## common odds ratio
## 0.9856598
```

According to the tests results shows above, we conclude the association between low birth weight and drinking controlling for death is not statistical significant.

Code 4

Test for Association between Low Birth Weight and Death

```
#Risk ratio
RR1<-riskratio(x=new_bwgt$Death.f, y=new_bwgt$LowBirthWgt.f)
RR1$measure
```

```
##           risk ratio with 95% C.I.
## Predictor estimate    lower    upper
##      No  1.000000      NA      NA
##      Yes 8.201933 6.962315 9.662259
```

```
RR1$p.value
```

```
##           two-sided
## Predictor midp.exact  fisher.exact    chi.square
##      No      NA      NA      NA
##      Yes      0 2.342608e-146 4.397064e-176
```

According to the risk ratio, since the 95% confidence interval does not contain 1, the association between the low birth weight and death is statistical significant.

```
#chisq test
chisq.test(new_bwgt$Death.f,new_bwgt$LowBirthWgt.f)
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  new_bwgt$Death.f and new_bwgt$LowBirthWgt.f
## X-squared = 796.62, df = 1, p-value < 2.2e-16
```

```
#oddsratio
oddsratio(new_bwgt$Death.f, new_bwgt$LowBirthWgt.f,rev=c("b"),method="wald")$measure
```

```
##           odds ratio with 95% C.I.
## Predictor estimate    lower    upper
##      Yes  1.00000      NA      NA
##      No  23.64956 18.22641 30.68632
```

```
# Fisher's Exact Test
tab24<- table(x=new_bwgt$Death.f , y=new_bwgt$LowBirthWgt.f)
tab22<- as.matrix(tab24)
```

```
# Two-sided alternative hypothesis
fisher.test(tab24)
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  tab24
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  18.06601 30.93969
## sample estimates:
## odds ratio
```



```
##      23.58755
# One-sided less than alternative hypothesis
fisher.test(tab24,alternative="less")

##
## Fisher's Exact Test for Count Data
##
## data:  tab24
## p-value = 1
## alternative hypothesis: true odds ratio is less than 1
## 95 percent confidence interval:
##   0.00000 29.65712
## sample estimates:
## odds ratio
##      23.58755
# One-sided greater than alternative hypothesis
fisher.test(tab24,alternative="greater")

##
## Fisher's Exact Test for Count Data
##
## data:  tab24
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is greater than 1
## 95 percent confidence interval:
##  18.82176      Inf
## sample estimates:
## odds ratio
##      23.58755
# MH test:
mantelhaen.test(new_bwgt$Death.f, new_bwgt$LowBirthWgt.f, new_bwgt$Drinking.f,correct=T)

##
## Mantel-Haenszel chi-squared test with continuity correction
##
## data:  new_bwgt$Death.f and new_bwgt$LowBirthWgt.f and new_bwgt$Drinking.f
## Mantel-Haenszel X-squared = 794.75, df = 1, p-value < 2.2e-16
## alternative hypothesis: true common odds ratio is not equal to 1
## 95 percent confidence interval:
##  18.20822 30.65360
## sample estimates:
## common odds ratio
##      23.62515
```