Unit 7 – Logistic Regression
Practice Problems (1 of 2)

Before you begin. Download from the course website
depress.xlsx or
depress.Rdata

Introduction to depress.xlsx

Source: Afifi A., Clark VA and May S. Computer Aided Multivariate Analysis, Fourth Edition. Boca Raton: Chapman and Hall, 2004.

The data are a study of depression and was a longitudinal study. The purpose of the study was to obtain estimates of the prevalence and incidence of depression and to explore its risk factors. The study variables were of several types – demographics, life events, stressors, physical health, health services utilization, medication use, lifestyle, and social support.

In this homework, we will be using the following three variables.

Variable	Codings	Label
drink	1 = yes 2 = no	Regular Drinker
sex	1 = male 2 = female	
cases	0 = Normal 1 = Case of Depression	Depressed is cesd ≥ 16

R

```
# load data
library(tidyverse)
load(file="depress.Rdata")                          # load(file="name.Rdata") to load an R dataset

temp <- depress %>%
    select(sex,drink,cases)                         # select( ) to keep variables of interest
glimpse(temp)                                       # glimpse() in {tidyverse} to inspect dataset structure

## Observations: 294
## Variables: 3
## $ sex   <fct> 2. female, 1. male, 2. female, 2. female, 2. female, 1. ma…
## $ drink <fct> 2. no, 1. yes, 1. yes, 2. no, 1. yes, 1. yes, 2. no, 2. no…
## $ cases <fct> 0. normal, 0. normal, 0. normal, 0. normal, 0. normal, 0. …
```

#1.

By any means you like, complete the following table by supplying the cell frequencies.

	Sex at Birth		
Regular Drinker	Male	Female	Total
Yes	95	139	234
No	16	44	60
Total	111	183	294

What are the odds that a male is a regular drinker? **95 / 16 = 5.9132**

What are the odds that a female is a regular drinker? **139 / 44 = 3.159**

What is the odds ratio? That is, what is the odds that a male, relative to a female, is a regular drinker?

OR = [odds for male] / [odds for female] = 5.9132/3.159 = 1.88

```
library(summarytools)                                          # ctable() in {summarytools}
library(DescTools)                                             # OddsRatio() in {DescTools}
ctable(temp$sex,temp$drink,prop='n',total=TRUE)                # ctable(FACTOR,FACTOR) in {summarytools}
```

```
## Cross-Tabulation
## sex * drink
## Data Frame: temp
##
## ----------- ------- -------- ------- -------
##               drink   1. Yes   2. No   Total
##         sex
##     1. Male               95      16     111
##   2. Female              139      44     183
##       Total              234      60     294
## ----------- ------- -------- ------- -------

OddsRatio(temp$drink,temp$sex,method="wald",conf.level=.95)

## odds ratio     lwr.ci     upr.ci
##   1.879496   1.002142   3.524956
```

#2.

Next, repeat the tabulation that you produced for exercise #1 two times, one for persons who are depressed and the other for persons who are not depressed.

Among Persons Who are Depressed

	Sex at Birth		
Regular Drinker	Male	Female	Total
Yes	8	33	41
No	2	7	9
Total	10	40	50

What are the odds that a male is a regular drinker? **8 / 2 = 4**

What are the odds that a female is a regular drinker? **33 / 7 = 4.714**

What is the odds ratio? That is, what is the odds that a male, relative to a female is a regular drinker?

OR = [odds for male] / [odds for female] = 4/4.714 = 0.85

Among Persons Who are *NOT* Depressed

	Sex		
Regular Drinker	Male	Female	Total
Yes	87	106	193
No	14	37	51
Total	101	143	244

What are the odds that a male is a regular drinker? **87 / 14 = 6.214**

What are the odds that a female is a regular drinker? **106 /37 = 2.865**

What is the odds ratio? That is, what is the odds that a male, relative to a female, is a regular drinker?

OR = [odds for male] / [odds for female] = 6.214/2.865 = 2.17

```
library(summarytools)
library(DescTools)

subset_depressed <- subset(temp, cases=="1. depressed")     # subset(SOURCE, CONDITION THAT MUST BE TRUE)
ctable(subset_depressed$sex,subset_depressed$drink,prop='n',total=TRUE)

## Cross-Tabulation
## sex * drink
## Data Frame: subset_depressed
##
## ----------- ------- -------- ------- -------
##               drink   1. yes   2. no   Total
##         sex
##     1. male               8       2      10
##   2. female              33       7      40
##       Total              41       9      50
## ----------- ------- -------- ------- -------

OddsRatio(subset_depressed$drink,subset_depressed$sex,method="wald",conf.level=.95)

## odds ratio     lwr.ci     upr.ci
##  0.8484848  0.1472930  4.8877157
```

OR = [odds for male] / [odds for female] = (8/2)/(33/7) = 0.85

```
subset_not <- subset(temp, cases=="0. normal")
ctable(subset_not$sex,subset_not$drink,prop='n',total=TRUE)

## Cross-Tabulation
## sex * drink
## Data Frame: subset_not
##
## ----------- ------- -------- ------- -------
##               drink   1. yes   2. no   Total
##         sex
##     1. male              87      14     101
##   2. female             106      37     143
##       Total             193      51     244
## ----------- ------- -------- ------- -------

OddsRatio(subset_not$drink,subset_not$sex,method="wald",conf.level=.95)

## odds ratio     lwr.ci     upr.ci
##   2.169137   1.101997   4.269666
```

OR = [odds for male] / [odds for female] = (87/14)/(106/37) = 2.17

#3.

By any means you like, create the following 3 variables:

drink01 = 0/1 indicator of **drink=1**
female01 = 0/1 indicator of **sex=2**
fem_case = 0/1 indicator of **female01 x cases**

```
temp$drink01 <- as.numeric(temp$drink == "1. yes", na.rm=TRUE)
temp$female01 <- as.numeric(temp$sex == "2. female", na.rm=TRUE)
temp$case01 <- as.numeric(temp$cases=="1. depressed", na.rm=TRUE)
temp$fem_case <- temp$female01*temp$case01


glimpse(temp)
## Observations: 294
## Variables: 7
## $ sex      <fct> 2. female, 1. male, 2. female, 2. female, 2. female, 1.…
## $ drink    <fct> 2. no, 1. yes, 1. yes, 2. no, 1. yes, 1. yes, 2. no, 2.…
## $ cases    <fct> 0. normal, 0. normal, 0. normal, 0. normal, 0. normal, …
## $ drink01  <dbl> 0, 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 0, 0…
## $ female01 <dbl> 1, 0, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1…
## $ case01   <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0…
## $ fem_case <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 0…
```

#4.

Next, fit a logistic regression model with **drink01** as the dependent variable and **cases, female01**, and **fem_case** as predictor variables Is the interaction term **fem_case** in your model significant?

```
m3 <- glm(drink01 ~ case01 + female01 +fem_case,data=temp, family=binomial)
summary(m3)

##
## Call:
## glm(formula = drink01 ~ case01 + female01 + fem_case, family = binomial,
##     data = temp)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9880   0.5463   0.5833   0.7738   0.7738
##
## Coefficients:
##             Estimate Std. Error z value       Pr(>|z|)
## (Intercept)   1.8269     0.2880   6.344 0.000000000224 ***
## case01       -0.4406     0.8414  -0.524          0.601
## female01     -0.7743     0.3455  -2.241          0.025 *
## fem_case      0.9386     0.9579   0.980          0.327       This is the Wald test of NULL: Zero interaction
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 297.53  on 293  degrees of freedom
## Residual deviance: 291.92  on 290  degrees of freedom
```

```
## AIC: 299.92
##
## Number of Fisher Scoring iterations: 4

exp(cbind(OR=coef(m3),confint(m3)))

##                    OR     2.5 %     97.5 %
## (Intercept) 6.2142857 3.6573468 11.4086594
## case01      0.6436782 0.1423289  4.5542681
## female01    0.4610127 0.2277725  0.8901499
## fem_case    2.5564834 0.3082721 15.2551393        NOTE:  The 95% CI for the OR for interaction null value of 1
```

R

```
library(lmtest)                                          # lrtest() in {lmtest} for Likelihood-Ratio Test
reduced <- glm(drink01 ~ case01 + female01,data=temp, family=binomial)
full <- glm(drink01 ~ case01 + female01 +fem_case,data=temp, family=binomial)
lrtest(reduced,full)

## Likelihood ratio test
##
## Model 1: drink01 ~ case01 + female01
## Model 2: drink01 ~ case01 + female01 + fem_case
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   3 -146.40
## 2   4 -145.96  1 0.8759     0.3493                LR Test of NULL: zero interaction is NOT signficant (p=.35)
```

The interaction term fem_case is *not* statistically significant.

$\hat{\beta}_3 = 0.9386 \quad \hat{SE}(\hat{\beta}_3) = 0.96$ and p-value = .35

In this sample, there is no statistically significant evidence that the odds regular drinking is associated with an interaction of gender and case of depression (LR test, df=1, p-value = .3493). This is close to the Wald statistic p-value = .327 that can be seen in the coefficients table.