**Statistics 141SL**
**Exercise Four**
**Professor Esfandiari**
**Winter 2014**

**PART ONE**

**Contingency table for two categorical variables**
**Endorsing women in public office = outcome**
**Level of education = Predictor**

***>table1<-table(femaleoffice,edu)***
***> table1 # print table***

| | edu | | | |
|---|---|---|---|---|
| femaleoffice | Four year | Graduate | HS or less | Two year college |
| Yes | 322 | 260 | 398 | 378 |
| No | 62 | 49 | 217 | 141 |

***> margin.table(table1,1)***
femaleoffice

| Yes | No |
|---|---|
| 1358 | 469 |

***> margin.table(table1,2)***
edu

| Four year college | Graduate | HS or less | two year college |
|---|---|---|---|
| 384 | 309 | 615 | 519 |

| Endorsing women for public office | Four year college | Graduate | High school or less | Two year college | Column total |
|---|---|---|---|---|---|
| Yes | 322 | 260 | 398 | 378 | 1358 |
| No | 62 | 49 | 217 | 141 | 469 |
| Column total | 384 | 309 | 615 | 519 | 1827 |

**PART Two**
**Calculation and interpretation of odds ratios resulting from the contingency table given in part one.**

***The odds of participants with a four-year degree to endorse women in public office***

> (322/384)/(62/384)
[1] 5.193548 = odds
***> log(5.193548)***
***[1] 1.647417= log of the odds***

(This is the intercept; four-year college was coded as the reference group).

***The odds of endorsing women for public office for participants with a graduate degree compared to those with a four-year college degree.***

> (260/49)/(322/62)
**[1] 1.021676 = odds**
> log(1.021767)
[1] 0.02153348 = log of odds
This is not statistically significant

***The odds of endorsing women for public office with a high school degree compared to those with a graduate degree.***

> (398/217)/(322/62)
**[1] 0.35315 = odds**
> log(0.35315)
[1] -1.040862 = log of odds

***The odds of endorsing women for public office participants with a two-year college degree compared to those with a graduate degree.***

> (378/141)/(322/62)
**[1] 0.5161887 = odds**
log(0.5161877)
[1] -0.6612848 = log of odds

**PART THREE**

```
> m1<-glm(femaleoffice~edu, family="binomial")
> summary(m1)

Call:
glm(formula = femaleoffice ~ edu, family = "binomial")

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.9191  -1.4434   0.5935   0.7963   0.9329

Coefficients:
                          Estimate   Std. Error  z value  Pr(>|z|)
(Intercept)               1.64742    0.13869     11.879   < 2e-16 ***
edugraduate               0.02144    0.20854     0.103    0.918098
eduHS or less             -1.04086   0.16234     -6.411   1.44e-10 ***
edutwo year college       -0.66128   0.17021     -3.885   0.000102 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2081.3  on 1826  degrees of freedom
Residual deviance: 2015.4  on 1823  degrees of freedom
  (423 observations deleted due to missingness)
AIC: 2023.4

> exp(coef(m1))
          (Intercept)           edugraduate         eduHS or less edutwo year college
            5.1935484             1.0216758             0.3531500             0.5161887
```

**NOTICE THAT…**

- **THE ESTIMATES FROM THE REGRESSION MODEL ARE THE LOG OF THE ODDS CALCULATED BASED ON THE TABLE OF FREQUENCIES.**
- **WHEN THE LOG OF THE ODDS ARE EXPONENTIATED, WE GET THE ODDS RATIOS CALCULATED FROM THE TABLE OF FREQUENCIES.**
- **AS YOU NOTICE TWO OUT OF THREE ODDS RATIOS ARE LESS THAN ONE. IN THE FOLLOWING, WE WILL DISCUSS THE BEST WAY TO INTERPRET THEM.**

# PART FOUR

## Interpreting odds ratios less than one

**Method one**: When you are interpreting an odds ratio (or any ratio for that matter), it is often helpful to look at how much it deviates from 1. So, for example, an odds ratio of 0.75 means that in one group the outcome is 25% less likely. An odds ratio of 1.33 means that in one group the outcome is 33% more likely."

**Method two**: When we encounter an odds ratio that is less than one, it can only be interpreted as follows: "The odds of the first group experiencing an event is less than the odds of the second group experiencing it". In practice, when dealing with the odds ratio less than 1, we can calculate and interpret the inverse. For example, suppose we find the odds ratio to be 0.25. This implies that the odds of the first group experiencing an event compared to the odds of the second group experiencing it is 0.25. The inverse is equal to 1/0.25 = 4. We could then say the odds for the second group to experience the event compared to the first group is four.

**These are the results we got**

**Outcome variable**: Endorsing vs. not endorsing women for public office
**Predictor**: Level of education (graduate, four-year college, two-year college, and high school).

| | Odds of endorsing women for public office | Interpretation one | Interpretation two | **Interpretation three Recommended over method two** |
|---|---|---|---|---|
| Four-year college | 5.1935484 P=0.000 | For participants, with a four-year college degree, the odds of endorsing women for public office vs. not endorsing them is 5.2 | | |
| Graduate vs. four-year college | 1.021676 (P=0.92) | Participants with four-year college degree and graduate degree are equally likely to endorse having women in public office | | |
| Two-year vs. four-year college | 0.5161887 P=0.000 | The odds of endorsing women in public office is less for participants with two-year than those with four-year college graduates. | The participants with a two-year college degree are 48% less likely to endorse women in public office than those with a four-year college degree. | 1/0.5161887= 1.937276 The odds of endorsing women for public office is 1.93 times higher for participants with four-year college degree compared to those with two-year college degrees. **(322/62)/(378/141) [1] 1.937276** |
| High school vs. four-year college | 0.3531500 P=0.000 | The odds of endorsing women in public office is less for participants with high school degree than those with four-year college degrees. | The participants with a high school degree are 65% less likely to endorse women in public office than those with a four-year college degree. | 1/0.3531500 2.831658 The odds of endorsing women for public office is 2.83 times higher for participants with four-year college degree compared to those with high school degree. **> (322/62)/(398/217) [1] 2.831658** |

## PART FIVE

Another potential and interesting variable for recommending women for public office could be discrimination against women. We would expect the participants who do not endorse having women in public office to have a higher score on discrimination. We can examine this hypothesis by conducting a two-sample test of the mean.

```
> t.test(discrimination~femaleoffice)

data:  discrimination by femaleoffice
t = 8.2695, df = 726.507, p-value = 6.401e-16
alternative hypothesis: true difference in means is not equal
to 0
95 percent confidence interval:
  6.560748 10.645695

sample estimates:
Mean for discrimination against women in
group who do not endorse women in public
office = 52.12646
Mean for discrimination against women in
group who do not endorse women in public
office = 43.52323
```

**Conclusion: On average participants who do not endorse women for public office score higher on discrimination against women.**

```
> m2<-glm(femaleoffice~discrimination, family="binomial")
> summary(m2)
```

Coefficients:

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | 2.293826 | 0.174330 | 13.158 | < 2e-16 *** |
| **discrimination** | **-0.026676** | **0.003305** | **-8.071** | **6.97e-16 ***** |

---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1856.9 on 1603 degrees of freedom
Residual deviance: 1786.8 on 1602 degrees of freedom
(646 observations deleted due to missingness)
AIC: 1790.8

```
> exp(-0.026676)
[1] 0.9736767
> exp(-0.026676*10)
[1] 0.7658569
```

**Interpretation**. If we increase the score on discrimination against women ten points, the odds of endorsing women in public office decreases by 24%. (1-0.76).

```
> m3<-glm(femaleoffice~edu+discrimination,family="binomial")
> summary(m3)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-2.3368  -1.1502   0.6287   0.8241   1.3925

Coefficients:
                    Estimate   Std. Error   z value   Pr(>|z|)
(Intercept)         2.728137   0.219578     12.424    < 2e-16 ***
edugraduate         -0.065180  0.217954     -0.299    0.764898
eduHS or less       -0.936560  0.173512     -5.398    6.75e-08 ***
edutwo year college -0.680482  0.179918     -3.782    0.000155 ***
discrimination      -0.024270  0.003347     -7.251    4.13e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1852.3  on 1599  degrees of freedom
Residual deviance: 1738.6  on 1595  degrees of freedom
  (650 observations deleted due to missingness)
AIC: 1748.6

> exp(coef(m3))
 (Intercept)       edugraduate      eduHS or less  edutwo    year college    discrimination
 15.3043442         0.9368986         0.3919740          0.5063726                0.9760220 It
```

## Interpretation of the coefficients:

If we partial out the effect of discrimination against women..,

- The odds of recommending women for public office is similar for participants with four-year college degree and those with graduate degree .
- 1/0.39 = 2.56:The odds of recommending women for public office is 2.56 times larger for participants with a four-year college degree than those with a high school education or less.
- 1/0.51= 1.96: The odds of recommending women for public office is 1.96 times larger for participants with a four-year college degree than those with a two-year college degree.

If we partial out the effect of level of education, and if we increase discrimination against women ten points, the odds of recommending women for public office decreases by 22%.

**exp(-0.024270*10)**
**[1] 0.784506**

## PART SIX

Given the following outputs, answer the following questions.

a) What are the dependent and independent variables and how were they measured (see codebook – pages 12- 16)
b) Using output one below, show how you can calculate odds of endorsing women to deal with social issues for urban compared to rural communities.
c) What does output two show?
d) What does output three show?
e) What does output four show?
f) Interpret the odds ratios given for the different predictors in output four WITHIN CONTEXT.

## OUTPUT ONE

***> table(communitytype)***

communitytype

| R | S | U |
|---|---|---|
| 445 | 986 | 814 |

- table(representpublicr)

representpublicr

| men | women |
|---|---|
| 384 | 621 |

> table(representpublicr,communitytype)

| | communitytype | | |
|---|---|---|---|
| representpublicr | R | S | U |
| men | 102 | 148 | 133 |
| women | 108 | 273 | 239 |

## OUTPUT TWO

***> chisq.test(representpublicr,communitytype)***

Pearson's Chi-squared test

data: representpublicr and communitytype
X-squared = 12.1681, df = 2, p-value = 0.002279

**OUTPUT THREE**

***> t.test(leadership~representpublicr)***

Welch Two Sample t-test

data: leadership by representpublicr
t = -4.1925, df = 385.392, p-value = 3.425e-05
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-11.250453 -4.067064

sample estimates:
mean in group men mean in group women
39.47562 47.13438

**OUTPUT FOUR**

***> m6<-glm(representpublicr~leadership+communitytype,family="binomial")***
***> summary(m6)***

Deviance Residuals:
Min 1Q Median 3Q Max
-1.9616 -1.2524 0.7890 0.9997 1.4261

Coefficients:

| | Estimate | Std. Error | z value | Pr(>\|z\|) |
|---|---|---|---|---|
| (Intercept) | -1.067735 | 0.339659 | -3.144 | 0.00167 ** |
| leadership | 0.023320 | 0.005819 | 4.007 | 6.14e-05 *** |
| communitytypeS | 0.575974 | 0.268545 | 2.145 | 0.03197 * |
| communitytypeU | 0.668372 | 0.279160 | 2.394 | 0.01666 * |

---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 573.26 on 426 degrees of freedom
Residual deviance: 549.87 on 423 degrees of freedom
(1823 observations deleted due to missingness)
AIC: 557.87

**> exp(coef(m6))**
**(Intercept) leadership communitytypeS communitytypeU**
**0.3437863 1.0235942 1.7788630 1.9510593**
**> exp( 0.023320 *10)**
**[1] 1.262634**

**Exercise to be done for week eight.**

**Do this as a group and present the results to the class in week eight.**

Using the subset of sex discrimination data set and the codebook posted on week seven, conduct a logistic regression with two qualitative and one quantitative variable. Interpret the coefficients of the resulting model within context. If you like, you can also try interaction between the two qualitative variables as well.

**Code book for the subset of sex discrimination data**

**edu**

Level of education

a) Graduate
b) Four-year college
c) Two-year college
d) High school or less

**race**

a) White
b) Black
c) Others

**citizen**

a) US citizen
b) Not US citizen

**USborn**

a) US born
b) Not US born

**boyorgirl**

If you could have only one child what would you want?

a) Boy
b) Girl
c) Either

**son**

Do you have a son?

a) Yes
b) No

**daughter**

Do you have a daughter?

a) Yes
b) No

**leadership**

High scores indicate endorsing women in leadership positions

**lifesatisfacion**

High scores indicate satisfaction with life

**Discrimination**
High scores means participant is pro discrimination toward women

**Publicoffice**
High scores indicate respondent endorses having women in public office

**happywithlife**
Score on life satisfaction is categorized above and below median

**femaleoffice**
Score on endorsing having women in public office is categorized above and below median

**Femaledisc**
Score on discrimination against women is categorized above and below median

**femaleleader**
Score on endorsing having women in the position of leadership is categorized into below and above median

**Census region**
a) Midwest
b) South
c) North east
d) West

**Community type**
a) Suburban
b) Rural
c) Urban

**Language**
a) English
b) Spanish

**gender**
a) Female
b) Male

**age**

**Party**
a) Republican
b) Democrat

**Partyln**
As of today, do you lean more toward the Republican or the Democratic party?
a) Republican
b) Democratic
c) Do not know

**ideology**
a) Conservative
b) Very conservative

**popdensity**
(population density)
a) High
b) Low

**Standingup**
Who is more likely to stand up?
a) Men
b) Women
c) No difference

**keepgovhonest**
Who is more likely to keep government honest?
a) Men
b) Women
c) No difference

**standingup**
Who is more likely to stand up?
a) Men
b) Women
c) No difference

**compromise**
Who is more likely to compromise?
a) Men
b) Women
c) No difference

**publicsaftey**
Who is more likely to keep public safety?
a) Men
b) Women
c) No difference

**dealsocialissue**
Who is more likely to deal with social issues?
a) Men
b) Women
c) No difference

**representpublic**
Who can represent the public better?
a) Men
b) Women
c) No difference

**Nationalsecurity**
Who is better at dealing with issues related to national security?
a) Men
b) Women
c) No difference

**The above variables were recoded and no difference was set equal to (NA)**

**Standingupr**
Who is more likely to stand up?
a) Men
b) Women

**keepgovhonest**
Who is more likely to keep government honest?
a) Men
b) Women

**Standingup1**
Who is more likely to stand up?
a) Men
b) Women

**compromiser**
Who is more likely to compromise?
a) Men
b) Women

**publicsafteyr**
Who is more likely to keep public safety?
a) Men
b) Women

**dealsocialissuer**
Who is more likely to deal with social issues?

a) Men
b) Women

**representpublicr**
Who can represent the public better?

a) Men
b) Women

**Nationalsecurityr**
Who is better at dealing with issues related to national security?

a) Men
b) Women

It can be shown that…

$$\log\left(\frac{\theta(x)}{1-\theta(x)}\right) = \beta_0 + \beta_1 x$$

$$\log\left(\frac{\theta(x)}{1-\theta(x)}\right)$$

$$\Downarrow$$

**Is known as logit**

$$\left(\frac{\theta(x)}{1-\theta(x)}\right)$$

$$\Downarrow$$

**Is known as odds**

Odds of success or odds in favor of success

$$\Downarrow$$

$$\left(\frac{\theta}{1-\theta}\right)=\left(\frac{P(success}{1-P(success}\right)$$

Odds of failure or odds against success

$$\Downarrow$$

$$\left(\frac{1-\theta}{\theta}\right)=\left(\frac{1-P(success}{P(success}\right)$$

$$\log\left(\frac{\hat{\theta}_x}{(1-\hat{\theta}_x)}\right)=\hat{\beta}_0+\hat{\beta}_1 x$$