# Unit2 Assignments

*Sri Seshadri*

*10/29/2017*

## Chapter 3 Question 1

```
Male_Yes <- 10
Male_No <- 30
Female_Yes <- 6
Female_No <- 34

# Probabilities of use among male and females
p.male.use <- Male_Yes/(Male_Yes + Male_No)
p.female.use <- Female_Yes/(Female_Yes + Female_No)
# Odds among male and females
odds_Male_use <- Male_Yes/Male_No
odds_Female_use <- Female_Yes/Female_No

# odds ratio Male vs Female

OR.Male.to.Female <- odds_Male_use/odds_Female_use

Q1df <- data.frame(Prob.Male.use = p.male.use, Prob.Female.use = p.female.use,
    Odd.Male.use = odds_Male_use, Odds.Female.use = odds_Female_use, `OddsRatio m/f` = OR.Male.to.Femal

knitr::kable(round(Q1df, 2))
```

| Prob.Male.use | Prob.Female.use | Odd.Male.use | Odds.Female.use | OddsRatio.m.f |
|---|---|---|---|---|
| 0.25 | 0.15 | 0.33 | 0.18 | 1.89 |

## Chapter 3 Question 2

```
religion <- read.csv(file = "religion.csv")
# sanitycheck <- do.call(rbind,dfapply(religion,favstats, select =
# is.numeric)) knitr::kable(round(sanitycheck,2), caption = 'Summary
# statistics')

# note there is missing data
attending.respondents <- sum(religion$RELSCHOL)
Total.respondents <- nrow(religion)

odds_relschol <- attending.respondents/(Total.respondents - attending.respondents)
prob_relschol <- attending.respondents/(Total.respondents)

ct <- descr::CrossTable(religion$RELSCHOL, religion$RACE, prop.r = F, prop.c = F,
    prop.chisq = F, prop.t = F)
```

```
ct
```

```
##    Cell Contents
## |-------------------------|
## |                       N |
## |-------------------------|
##
## ======================================
##                     religion$RACE
## religion$RELSCHOL      0     1   Total
## --------------------------------------
## 0                     76   470     546
## --------------------------------------
## 1                     26    54      80
## --------------------------------------
## Total                102   524     626
## ======================================
```

```
crosstable <- table(religion$RELSCHOL, religion$RACE)
Prob.nonwhite.relschool.attend <- crosstable[2, 1]/sum(crosstable[, 1])
Prob.white.relschool.attend <- crosstable[2, 2]/sum(crosstable[, 2])

odds.nonwhite.relschool.attend <- crosstable[2, 1]/crosstable[1, 1]
odds.white.relschool.attend <- crosstable[2, 2]/crosstable[1, 2]
OR <- odds.white.relschool.attend/odds.nonwhite.relschool.attend
```

a. The odds of attending religious school is 0.1465201 and the probability is 0.1277955

b. Probability of non-white attending religious school = 0.254902 Probability of white attending religious school = 0.1030534

   odds of non-white attending religious school = 0.3421053 odds of white attending religious school = 0.1148936 Odds ratio of WHITE over NON-WHITE = 0.3358429

## Chapter 3 Question 3

```
model1 <- glm(RELSCHOL ~ RACE, family = binomial, data = religion)
model2 <- glm(RELSCHOL ~ RACE + ATTEND + INCOME, family = binomial, data = religion)

# summary(model1)
coef1 <- coefficients(model1)[2]
odds.ratio <- exp(coef1)

AIC_model1 <- AIC(model1)
BIC_model1 <- BIC(model1)

AIC_model2 <- AIC(model2)
BIC_model2 <- BIC(model2)

q3bdf <- data.frame(Model = c(1, 2), AIC = c(AIC_model1, AIC_model2), BIC = c(BIC_model1,
    BIC_model2))
```

a. The odds ratio of white and non-white attendance based on logistic regression model is 0.335842883535181. It agrees with the question 2b above.

b.

```
knitr::kable(q3bdf, caption = "AIC & BIC stats")
```

Table 2: AIC & BIC stats

| Model | AIC | BIC |
|---|---|---|
| 1 | 467.4662 | 476.3449 |
| 2 | 424.7930 | 442.3135 |

based on the numbers above model 2 would be chosen. The lower the A/BICs the better the model.

c.

```
source("getequation.R")
eqn <- getequation("RELSCHOL", model2)
RACE <- c(1, 0)
ATTEND <- c(5, 5)
INCOME <- c(4, 4)

logoddsRELSHOL <- -3.58 - 1.29 * RACE + 0.33 * ATTEND + 0.2 * INCOME
odds_RELSCHOL <- round(exp(logoddsRELSHOL), 2)

paste("The odds of attendance for white and non white are ", round(odds_RELSCHOL[1],
    2), "and", round(odds_RELSCHOL[2], 2), "respectively")
```

```
## [1] "The odds of attendance for white and non white are  0.09 and 0.32 respectively"
```

d.

```
# summary(model1) summary(model2)

adj.odds.Ratio <- exp(coefficients(model2)[2])

# paste('The adjusted odds ratio for race is ', adj.odds.Ratio)
paste(" When effects of income and attend are controlled, change in race has an multiplicative effect o
    adj.odds.Ratio)
```

```
## [1] " When effects of income and attend are controlled, change in race has an multiplicative effect
```

The adjusted odds ratio for race is 0.27546 When effects of income and attend are controlled, change in race has an multiplicative effect of `r adj.odds.Ratio`

# Chapter 3 Question 4

```
model1.probit <- glm(RELSCHOL ~ RACE, family = binomial(link = "probit"), data = religion)
model2.probit <- glm(RELSCHOL ~ RACE + ATTEND + INCOME, family = binomial(link = "probit"),
    data = religion)

prob.White.attendance <- pnorm(sum(coefficients(model1.probit)))
prob.NonWhite.attendance <- pnorm(coefficients(model1.probit)[1])
```

a. The prob of white and non white attendances based on probit are 0.1030534 and 0.254902 respectively. They compare well to 2b.

b.

```
AIC_model1.probit <- AIC(model1.probit)
BIC_model1.probit <- BIC(model1.probit)

AIC_model2.probit <- AIC(model2.probit)
BIC_model2.probit <- BIC(model2.probit)

q4bdf <- data.frame(Model = c(1, 2), AIC = c(AIC_model1.probit, AIC_model2.probit),
    BIC = c(BIC_model1.probit, BIC_model2.probit))

knitr::kable(q4bdf, caption = "AIC and BICs for probit")
```

Table 3: AIC and BICs for probit

| Model | AIC | BIC |
|---|---|---|
| 1 | 467.4662 | 476.3449 |
| 2 | 423.0652 | 440.5857 |

Model 2 is still the winner, lower the A/BIcs the better the model.

c.

```
eqn2 <- getequation("ATTEND", model2.probit)
RACE <- c(1, 0)
ATTEND <- c(5, 5)
INCOME <- c(4, 4)
predicted.probs <- pnorm(-2.07 - 0.73 * RACE + 0.19 * ATTEND + 0.12 * INCOME)
```

The predicted probability for white is 0.09 and for non white 0.26

```
RACE <- c(0, 0)
ATTEND <- c(4, 4)
INCOME <- c(4, 10)
predicted.probs <- pnorm(-2.07 - 0.73 * RACE + 0.19 * ATTEND + 0.12 * INCOME)
change.prob <- predicted.probs[2] - predicted.probs[1]
```

d. The discrete change in predicted probability is 0.2529353

# Chapter 3 Question 5

Non-whites have lower probability of attending religious school, while income and service attendance increases the probability of religious attendance.