Homework 3, Question 3a

Rachel M. Smith

05 December 2016

———————————————

The data for the remaing sets of analyses are from a national telephone interview study of batterer intervention program (BIP) standards advisory and compliance monitoring committees. Respondents were asked a series of questions varying in structure from *open-ended* to simple *yes-or-no*. These analyses will concern the discrete data collected in response to the following interview questions:

> How many members currently serve on your standards committee?
> Does your organization have methods for assessing programs' feedback about the standards?
> Do the standards apply to programs designed for all genders?

The primary interest for the below analyses relate to an overarching effort to implement and sustain effective and appropriate anti-violence intervention strategies among female-identified perpetrators of same-sex violence. The above listed questions provide a mix of continuous numeric and dichotomous (0 = No; 1 = Yes) indicators of responding states' current organizational and ideological capacities for such intervention strategies.

Data Descriptives

R ────────────────────────────────────

```
dat <- read.csv("data/states_new.csv")
# datmap <- read.csv("data/usmap.csv")

dat <- dat[, c("Q3", "Q7", "Q8")]
dat <- na.omit(dat)
id <- seq(1:nrow(dat))
dat <- cbind(id, dat)
```

R ────────────────────────────────────

```
dat$Q8 <- ifelse(dat$Q8 == 2, 1, 0)
dat$Q7 <- ifelse(dat$Q7 == 2, 1, 0)
dat <- within(dat, {## could also do this with an "apply()" function ##
    Assessments <- factor(Q7, levels = c(0, 1), labels = c("No", "Yes"))
    Gender.Inclusive <- factor(Q8, levels = c(0, 1), labels = c("No", "Yes"))
})

# dat <- within(dat, {## could also do this with an "apply()" function ##
    # Q7 <- Q7 - mean(Q7)
    # Q3 <- Q3 - mean(Q3)
# })
```

Multiple Logistic Regression Model

R

```
fit <- glm(Q8 ~ Q3 + Q7 + Q3*Q7,
           data = dat,
           family = binomial(link = "logit"))
# lgm.dx <- LogisticDx::dx(fit, byCov = FALSE)
# x <- seq(1:nrow(lgm.dx))
# lgm.dx <- cbind(x, lgm.dx)
# dx1 <- lgm.dx[order(lgm.dx$n), c("x", "sPr","P", "dChisq", "dDev", "dBhat")] %>% round(5)
# kable(dx1)
# write.csv(dx1, "data/hw3-q3a.csv", row.names = FALSE)
```
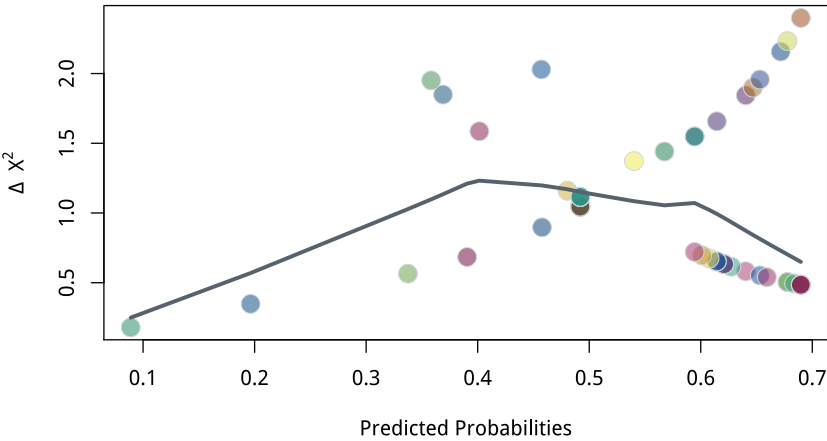
Model Fit Diagnostics

R

```
car::outlierTest(fit)
```

No Studentized residuals with Bonferonni p < 0.05
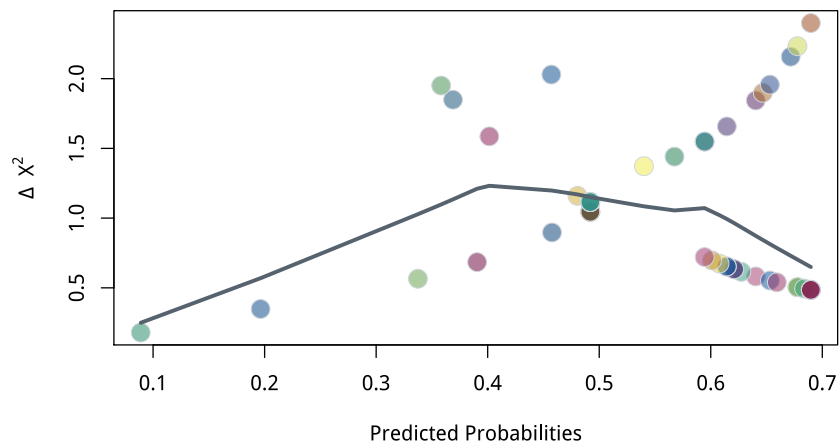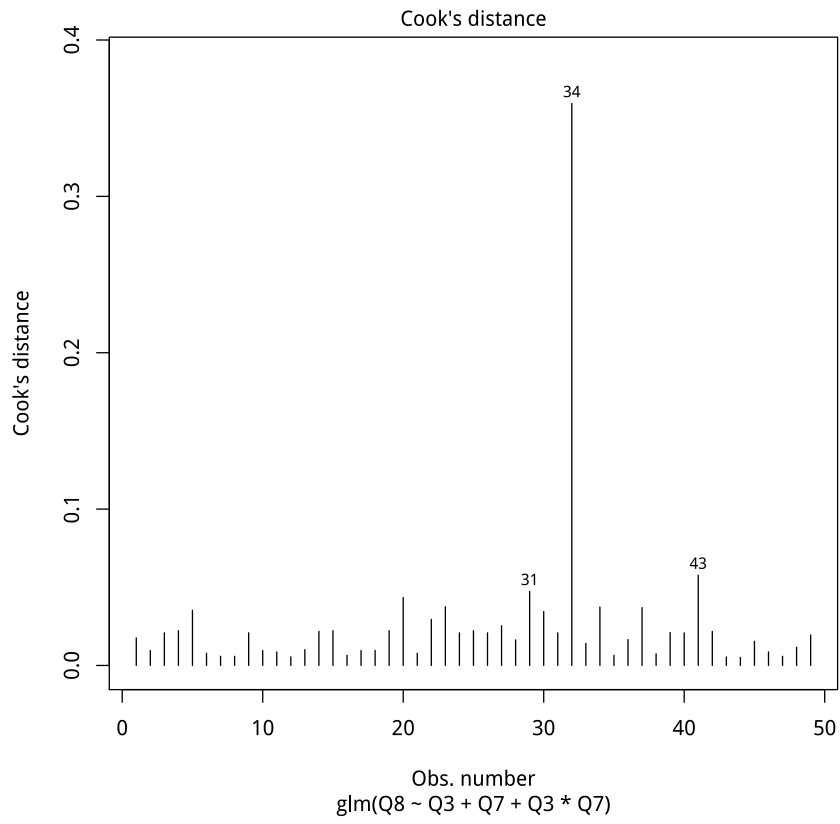Largest |rstudent|:

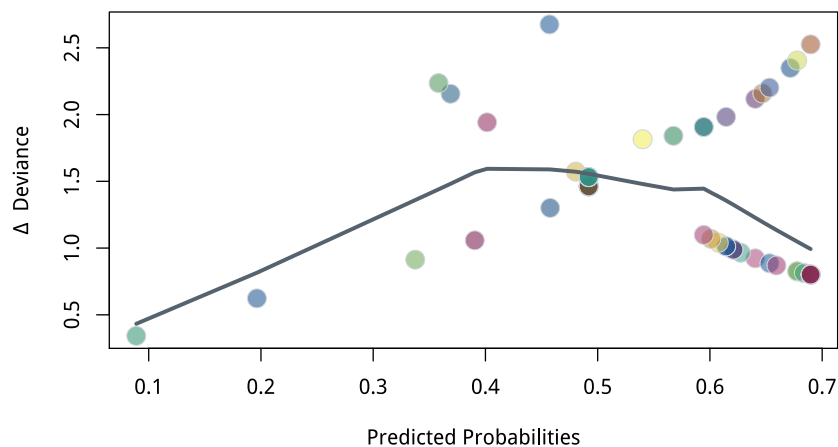|	rstudent	unadjusted p-value	Bonferonni p
31	-1.586	0.1127	NA

R
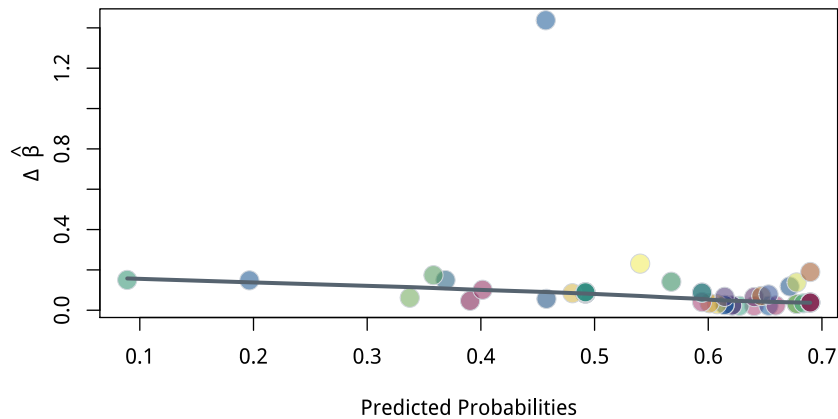
```
library(car)
cutoff <- 4/((nrow(dat) - length(fit$coefficients) - 2))
plot(fit, which = 4, cook.levels = cutoff)
```

Cook's distance

34

43

31

Obs. number
glm(Q8 ~ Q3 + Q7 + Q3 * Q7)

ΔX^2

Predicted Probabilities

MULTIPLE LOGISTIC REGRESSION DIAGNOSTICS SUMMARY. The Cook's distance outlier tests above did not seem to provide reliable information regarding potential outliers, when compared against the plotted diagnostics data. In particular, while cases 31 and 35 are identified as potential outliers according to cook's distance, neighter of these cases are associated with any remarkable deviance statistics. Rather, the visably outlying case in the $\Delta\beta$ plot above, which I identified as Case number 49, did concern me as a potential outlier; however, I found no other remarkable data for this in terms of the data used in testing the model as well as the deviance statistics obtained for the tested model. This overall suggests to me that no significant outliers exist in the data used in testing the above multiple logistic regression model.

Table 2: Descriptive Statistics for Residual Diagnostics

	M	SD	Min	Max
$\Delta\chi^2$	1.07	0.58	0.18	2.40
$\Delta Deviance$	1.42	0.58	0.34	2.68
$\Delta\hat{\beta}$	0.10	0.20	0.02	1.44