

Homework 3, Question 1b

Rachel M. Smith

05 December 2016

Data Descriptives

R

```
dat <- R.rspss("data/child2.sav", vlabels = FALSE)
dat <- na.omit(dat)
R.msmm(dat[, -5])
```

	M	SD	Min	Max	NAs
abuse	0.07	0.25	0	1	0
boyfriend	0.17	0.37	0	1	0
program	0.51	0.5	0	1	0
welfare	0.1	0.52	0	8	0
neglect	0.17	0.61	0	9	0

R

```
fit <- glm(abuse ~ program + boyfriend + white + welfare,
  data = dat,
  family = binomial(link = "logit"))
fit
```

Table 2: Fitting generalized (binomial/logit) linear model: abuse
 \sim program + boyfriend + white + welfare

	Estimate	Std. Error	z value	Pr(> z)
program	-0.2454	0.2369	-1.035	0.3004
boyfriend	-0.8127	0.3954	-2.056	0.03981
white	0.8045	0.2428	3.313	0.0009237
welfare	0.8665	0.1468	5.901	0.000000003615
(Intercept)	-2.983	0.2222	-13.42	4.32e-41

```
anova(fit, test = "LRT")
```

Table 3: Analysis of Deviance Table

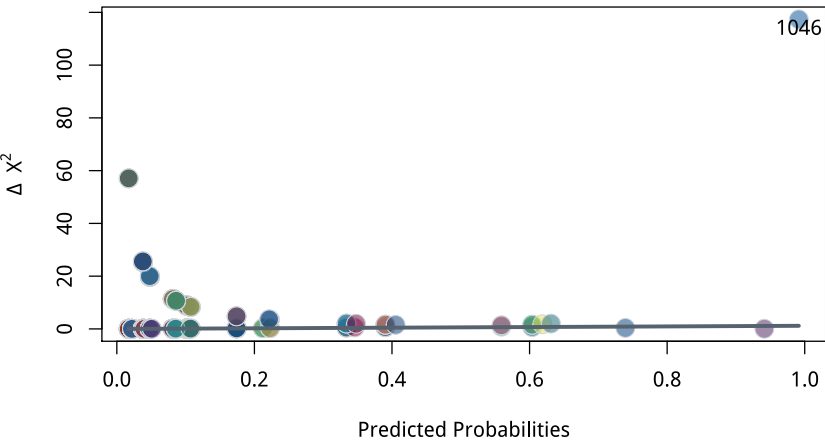
	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			1212	610.6	
program	1	1.046	1211	609.6	0.3064
boyfriend	1	3.836	1210	605.7	0.05016
white	1	13.42	1209	592.3	0.0002496
welfare	1	36.54	1208	555.8	0.000000001497

Model Fit Diagnostics

```
R
```

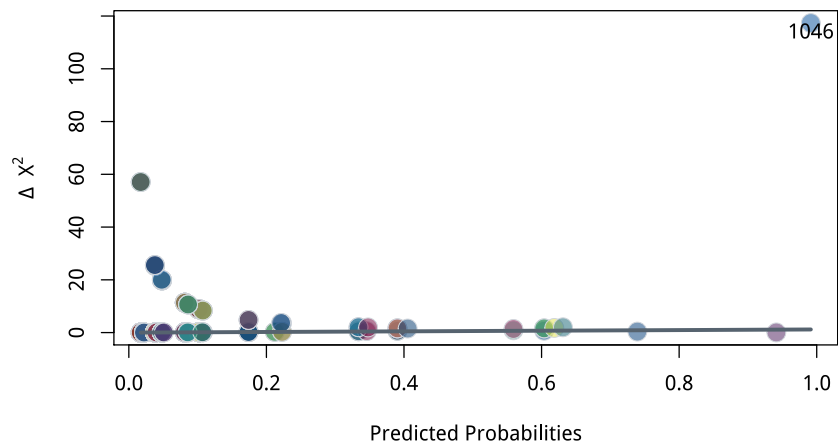
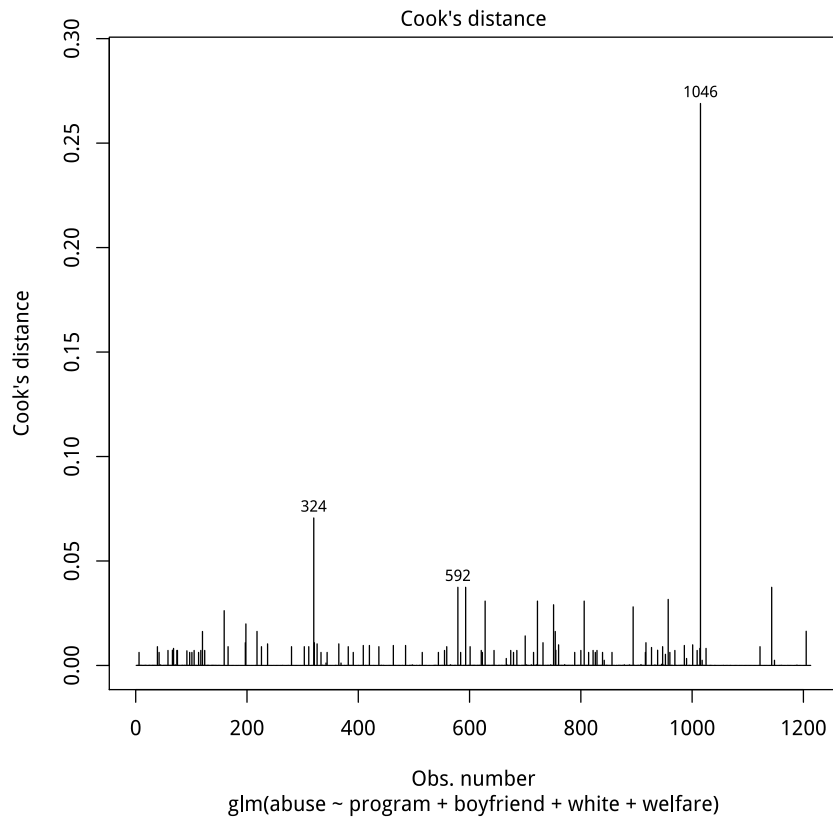
```
car::outlierTest(fit, cutoff = Inf)
```

	rstudent	unadjusted p-value	Bonferonni p
1046	-3.295	0.0009859	NA
592	2.882	0.0039569	NA
607	2.882	0.0039569	NA
1180	2.882	0.0039569	NA
648	2.589	0.0096232	NA
750	2.589	0.0096232	NA
834	2.589	0.0096232	NA
238	2.566	0.0102780	NA
330	2.566	0.0102780	NA
369	2.566	0.0102780	NA



```
R
```

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



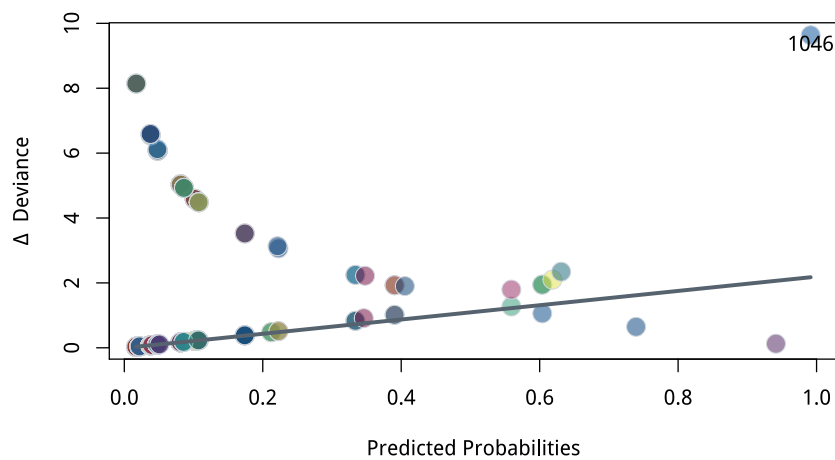
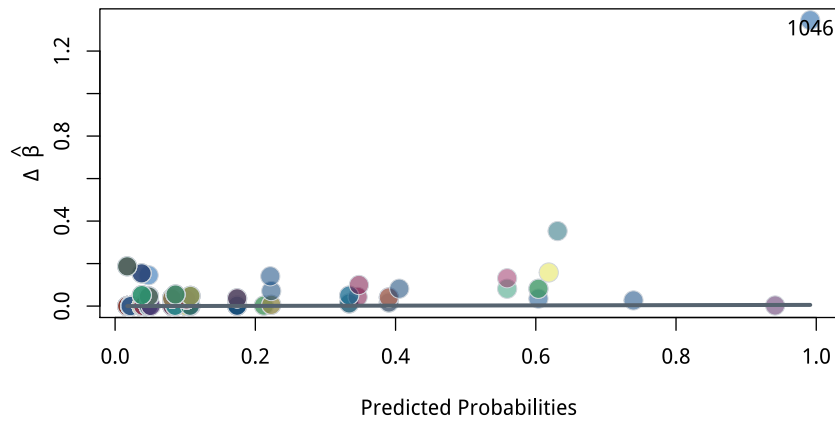


Table 5: Residual Diagnostic Statistics for Case No. 1046

	Standardized Pearson Residual	Predicted Probability	$\Delta\chi^2$	$\Delta Deviance$	$\Delta\hat{\beta}$
1213	-10.83	0.9914	117.3	9.633	1.345

MULTIPLE LOGISTIC REGRESSION DIAGNOSTICS SUMMARY. In the second residual plot above, with standardized pearson residuals on the Y-axis and predicted values on the X-axis, *Case #1046* is identified as an outlier. Examining the $\Delta\chi^2$ and $\Delta\hat{\beta}$ for this case (see above) against the aggregated descriptives for the full set of observations included in the model (see below), it is clear that this case is an outlier. This conclusion is supported by the studentized residual outlier test provided above. However, the residual data visualizations collectively suggest that this one observation (i.e., *Case #1046*) is not necessarily heavily influential on the fitted model's coefficients. For example, in the first diagnostic plot provided above, the solid line represents the fitted *loess model* for the tested model's predicted values

against the model's residuals. The fitted loess line's slope appears to correspond appropriately with the data with little influence from the outlier case¹. The same behavior is observed across subsequent visualizations three plots respectively showing the $\Delta\chi^2$, $\Delta\beta$, and $\Delta Deviance(D)$ plotted against the tested model's predicted probabilities, where the solid gray line in each plot represents the best fitting (loess) curve for each diagnostic statistic against the predicted probabilities. In all of the above-described visualizations, the best fitting line appears most heavily influenced by the data clustered toward the lower ends of each diagnostic statistic's range, rather than the labeled outlying data point in each plot. However, the difference between *Case #1046*'s predicted probability ($P = 0.9914$ in the table above) against the mean predicted probability for the full set of observations included in the model ($M = 0.07$ in the table below), suggests that this particular data point's predicted value could be influential on the tested model's outcome (*Abuse*). This influence could increase the risk of Type I error regarding the model's predictors relations with the outcome. In particular, *Case #1046*'s relatively high score on the *Welfare* predictor ($Welfare_{1046} = 8$, whereas $\mu_{Welfare} = 0.0973$) could influence the regression coefficient obtained for this predictor ($\beta_{Welfare} = 0.8665$).

¹ Located and labeled in the bottom right corner of the plot

Table 6: Descriptive Statistics for Residual Diagnostics

	M	SD	Min	Max
Standardized Pearson Residual	-0.02	1.02	-10.83	7.55
Predicted Probability	0.07	0.07	0.02	0.99
$\Delta\chi^2$	1.05	5.44	0.02	117.33
$\Delta Deviance$	0.46	1.27	0.03	9.63
$\Delta\hat{\beta}$	0.01	0.04	0.00	1.34