HW-2 R Notebook

Rachel M. Smith

14 November 2016

Question 1

	M	SD	Min	Max	NAs
id	295	170.2	1	589	0
ped	3.98	1.64	1	5	0
offense	1.63	0.66	0	2	0

Pedestrian Race	Frequency
African American	120
Asian	12
Hispanic	37
Native American	10
White	410

Offense	Frequency
major	99
minor	431

Table 4: Contingency Table of Pedestrians Stopped by Portland Police: Pedestrian Race (2 Levels) x Severity of Offense (2 Levels).

	NA	major	minor
African American	0	21	99
Asian	12	0	0
Hispanic	37	0	0
Native American	10	0	0
White	0	78	332

	Major	Minor	π_{+j}
African American	21	78	99
White	99	332	431
π_{i+}	120	410	530

1.*a*. *Loglinear Model*

$$\log(\mu_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}$$

Loglinear Analysis Summary. A loglinear model was used to test the association between pedestrian race and whether individuals were stopped by Portland Police (*PPB*) for major versus minor offenses. 77.36% of pedestrians stopped by PPB were White, and 22.64% were African American ($N_{Total} = 530$). Among the 120 African American pedestrians, 3.96% (21) were stopped for major offenses, while 18.68% (99) were stopped for minor offenses. Among the 410 White pedestrians, 14.72% (78) were stopped for major offenses, while 62.64% (332) were stopped for minor offenses. Results from the likelihood ratio test of the $2x2$ contingency table (see above) indicated no significant differences regarding these factors, $G^2(1) = 0.14, p = 0.70$.

Call:
`loglm(formula = ~ped.f + offense.f, data = tbl)`
Statistics:

	X^2	df	P(> X^2)
Likelihood Ratio	0.1437	1	0.7046
Pearson	0.142	1	0.7063

1.*b*. *Logistic Regression Model*

$$\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta X$$

$$\pi = \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$$

Logistic Regression Analysis Summary. A logistic regression analysis was conducted to test whether pedestrian race[1] predicted whether individuals were stopped by Portland Police for *major*, versus *minor*, offenses. Results for the model indicated that the severity of offenses was not predicted according to the two levels of pedestrians' race, likelihood ratio $\chi^2(1) = 0.1437$ (n.s.). The *Nagelkerke pseudo-*R^2 indicated that $< 1\%$ of the variance in offense severity was accounted for by the race of pedestrians stopped by PPB[2]. The model estimates and fit indices are summarized below.

[1] *African American* or *White*

[2] *pseudo-*$R_2 = 0.04\%$

Table 7: Likelihood Ratio χ^2

Log Likelihood	*df*	χ^2	*p*
-255.2	1	0.1437	0.7046

Table 8: Logistic Regression Model Summary

	Estimate	SE	Z	p.value
(Intercept)	1.5761	0.3020	5.2198	0.0000
ped	-0.0255	0.0678	-0.3767	0.7064

Table 9: Logistic Regression Coefficients (β) & Coresponding Confidence Intervals (*CI*)

		CI_β	
	β	2.5 %	97.5 %
(Intercept)	1.5761	1.0111	2.2009
ped	-0.0255	-0.1634	0.1035

Table 10: Logistic Regression Odds Ratios (Φ) & Coresponding Confidence Intervals (*CI*) [1]

		CI_{Φ}	
	Φ	2.5 %	97.5 %
(Intercept)	4.8363	2.7485	9.0333
ped	0.9748	0.8492	1.1091

Note:
[1] Confidence intervals are based on the logistic regression model's profiled log-likelihood function, rather than the standard errors

Table 11: Logistic Regression Model Fit Statistics

	Estimate	Degrees of Freedom
Null Deviance	510.4	529
Residual Deviance	510.3	528

1.*c*. *Findings Comparisons*

Findings for the loglinear and logistic regression models above support findings from the χ^2 analysis conducted previously (*HW-1*). In particular the likelihood ratio χ^2 found in *Homework-1, Question 3.c.* is exactly the same as that found from testing the loglinear model in *Question 1.a* here ($\chi^2 = 0.1437, p = 0.70$).

Original Contingency Table & Group Comparisons via Pearson χ^2.

	X^2	df	P(> X^2)
Likelihood Ratio	0.1437	1	0.70463
Pearson	0.14201	1	0.70629

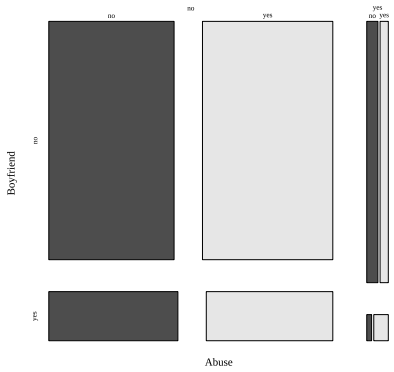
Phi-Coefficient : 0.016
Contingency Coeff.: 0.016
Cramer's V : 0.016

Question-2

Q-2 Data: `'child.sav'` & `'child2.sav'`

2.*a*. *Three-Way Loglinear Model*

$$\log(\mu_{ijk}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_k^z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}$$



Three-Way Loglinear Analysis Summary. A three-way loglinear model was conducted examing the three-way contingency table (abuse-x-boyfriend-program) to investigate whether program participation impacted the association between the boyfriend variable and any report of abuse (joint frequencies are provided in table below). Results from the likelihood ratio test (see below) indicated no significant differences regarding these factors, $G^2(1) = 3.318, p = 0.07$.

Call:

```
loglm(formula = ~Abuse:Boyfriend + Abuse:Program + Boyfriend:Program,
data = tbl, digits = 4)
```

Statistics:

	X^2	df	P(> X^2)
Likelihood Ratio	3.318	1	0.06853
Pearson	3.245	1	0.07163

2.*b*. *Logistic Regression with an Interaction Term*

$$\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta X$$

$$\pi = \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}$$

```
lgm <- glm(abuse ~ boyfriend + program + boyfriend*program, data = dat, family = "binomial")
```

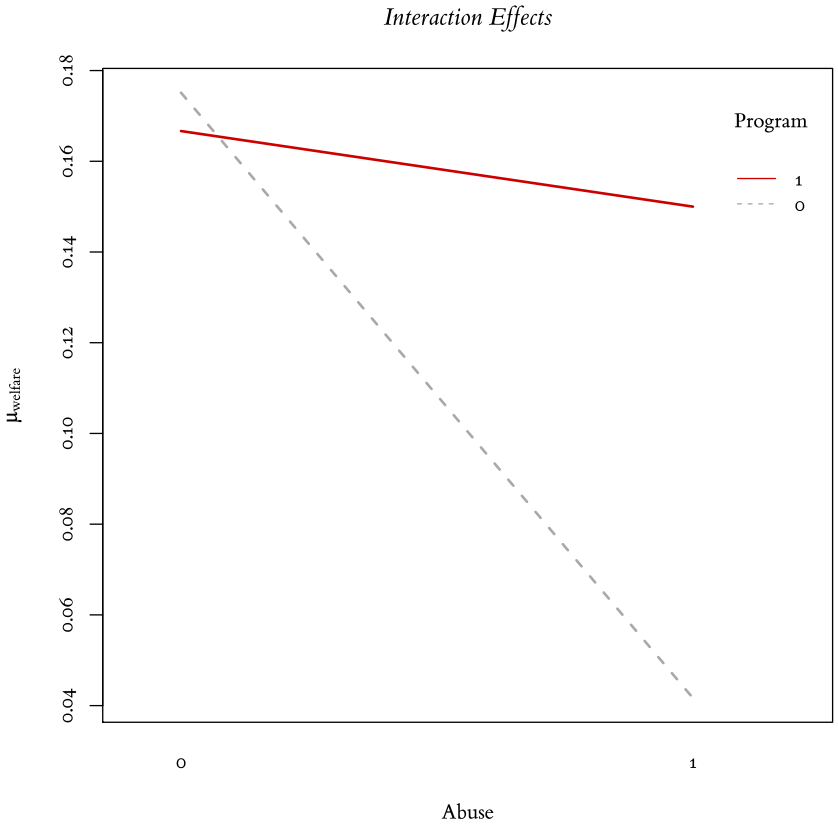
Multiple Logistic Regression Analysis Summary. A multiple logistic regression analysis was conducted predicting abuse by program participation, whether the mother had a boyfriend, and the program-x-boyfriend interaction. The interaction was only marginally significant, $\beta = 1.461, SE = 0.8615, p = 0.09$. The model's results are summarized below.

Table 14: Fitting generalized (binomial/logit) linear model: abuse ~ boyfriend + program + boyfriend * program

	Estimate	Std. Error	z value	Pr(>|z|)
boyfriend	-1.586	0.7306	-2.171	0.02997
program	-0.3418	0.2352	-1.453	0.1461
boyfriend:program	1.461	0.8615	1.696	0.08997
(Intercept)	-2.326	0.1545	-15.06	3.016e-51

Table 15: Analysis of Deviance Table

	Df	Deviance	Resid. Df	Resid. Dev
NULL			1246	636.2
boyfriend	1	4.348	1245	631.9
program	1	0.9113	1244	631
boyfriend:program	1	3.318	1243	627.7


Interaction Effects
Program
1
0
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18
μwelfare
0
1
Abuse

2.*c Findings Comparison*

Results from the original analysis (HW-1) showed a χ^2 statistic similar to that found in the loglinear analysis (*Q*2.*a*.) here ($\chi^2_1 = 4, \chi^2_2 = 3.3$). However, the signifance found in the original analysis was not replicated in the analysis conducted here, such that the association between the boyfriend variable and abuse, as tested in the analyses above, is not contingent upon the program participation condition.

2.*d% Multiple Logistic Regression*

Data Summary

	abuse	boyfriend	program	white
no	1129	1010	599	667
yes	84	203	614	546

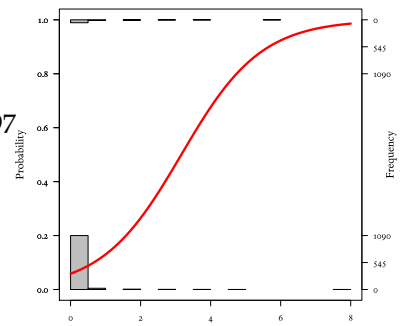
Multiple Logistic Regression Analysis Summary. A multiple logistic regression analysis was conducted predicting abuse by program participation, whether the mother had a boyfriend, the mother's race, and the frequency of welfare encounters. All predictors in the model except for the program participation term were significant. The model estimates and fit indices are summarized below.

Table 17: Fitting generalized (binomial/logit) linear model: abuse ~ 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

Table 18: 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



Confidence Intervals (CI) & Odds Ratios (OR)

Table 19: Logistic Regression Coefficients (β) & Coresponding Confidence Intervals (CI)

		CI_β	
	β	2.5 %	97.5 %
(Intercept)	-2.983	-3.4407	-2.5676
program	-0.2454	-0.7143	0.2177
boyfriend	-0.8127	-1.6664	-0.0977
white	0.8045	0.3347	1.2902
welfare	0.8665	0.5839	1.1638

Table 20: Logistic Regression Odds Ratios (Φ) & Coresponding Confidence Intervals (CI) [1]

		CI_Φ	
	Φ	2.5 %	97.5 %
(Intercept)	0.0506	0.032	0.0767
program	0.7824	0.4895	1.2432
boyfriend	0.4437	0.1889	0.9069
white	2.2356	1.3975	3.6335
welfare	2.3786	1.7931	3.2019

Note:
[1] Confidence intervals are based on the logistic regression model's profiled log-likelihood function,
rather than the standard errors

Table 21: Logistic Regression Model Fit Statistics

	Estimate	Degrees of Freedom
Null Deviance	610.61	1212
Residual Deviance	555.77	1208
AIC	565.77	

2.*e. Moderation Analysis*

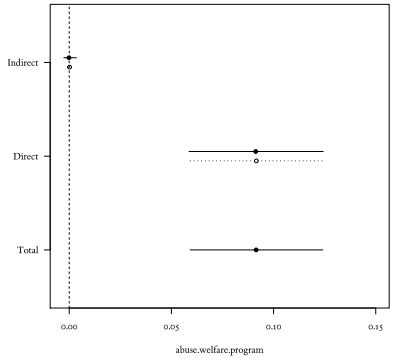
MEDIATION ANALYSIS SUMMARY. A regression model was tested to investigate whether participation in the Early Head Start program moderated the assocation between the number of welfare encounters and the probability of abuse, with the mother's race included in the model as a covariate. Except for the program participation term, all model coefficients were significant, including the welfare-x-program interaction term (see Table 13 below).

Table 22: Fitting generalized (binomial/probit) linear model: f2

	Estimate	Std. Error	z value	Pr(>|z|)
welfare	0.5103	0.08834	5.777	0.000000007623
program	-0.164	0.117	-1.402	0.1609
white	0.371	0.1165	3.185	0.001448
welfare:program	0.3526	0.1761	2.002	0.0453
(Intercept)	-1.565	0.06014	-26.03	2.225e-149

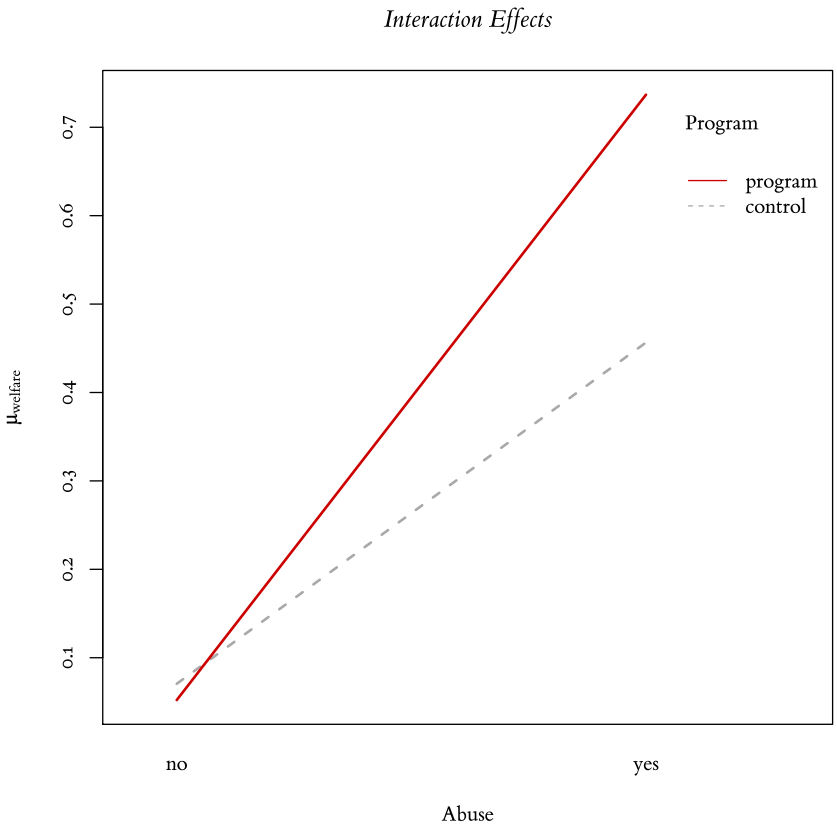
Table 23: Fitting generalized (gaussian/identity) linear model: f1

	Estimate	Std. Error	t value	Pr(>|t|)
welfare	-0.005106	0.02783	-0.1835	0.8544
white	-0.004296	0.02893	-0.1485	0.882
(Intercept)	3.098e-17	0.01437	2.155e-15	1




```
## "interaction.plot()" from the {stats} package ##
with(dat2, {
    interaction.plot(
    abuse,
    program,
    welfare,
    col = c("darkgray", "#cd0000"),
    lwd = 2,
    main = "Interaction Effects",
    ylab = expression(mu[welfare]),
```

```
    xlab = "Abuse",
    trace.label = "Program"
    )
})
```



Question-3

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:

> **Q3:** How many members currently serve on your standards committee?
> **Q7:** Does your organization have methods for assessing programs' feedback about the standards?
> **Q8:** 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.

Multiple Logistic Regression Analysis Summary. A multiple logistic regression analysis was conducted predicting whether states' batterer intervention program (BIP) standards were gender inclusive (*Q8*) by the size of state standards' committees (*Q3*), whether proccesses were in place for assessing BIPs' feedback about the standards (*Q7*), and the interaction of these two predictors (*Q3-x-Q7*). None of these predictors were significant. The model estimates and fit indices are summarized below.

Table 24: Fitting generalized (binomial/logit) linear model: Q8 ~ Q3 + Q7 + Q3 * Q7

	Estimate	Std. Error	z value	Pr(>|z|)
Q3	-0.00000002919	3077	-9.487e-12	1
Q7	49.13	96165	0.0005109	0.9996
(Intercept)	2.233	39501	0.00005654	1

Table 25: Analysis of Deviance Table

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
NULL			10	15.16	
Q3	1	9.751	9	5.407	0.001792
Q7	1	5.407	8	4.715e-10	0.02006
Q3:Q7	0	0	8	4.715e-10	

Table 26: Likelihood Ratio χ^2

Log Likelihood	*df*	χ^2	*p*
-7.579	1	15.16	0.0005

Table 27: Logistic Regression Model Fit Statistics

	Estimate	Degrees of Freedom
Null Deviance	15.16	10
Residual Deviance	0.00	8