Simple Logistic Regression Examples

Dichotomous Predictor

The YouGov survey data[1] is reanalyzed with simple logistic with a binary predictor. Compare these results to the results from the contingency table analyses in the handout “2 × 2 Contingency Chi-Square.”

```
logistic regression vars=voted with youth
   /print=summary ci(95) iter(1).  *the CI option should have a value between 1 and 99 (no decimal)
```

The first section of the output includes results for Block 0. Block 0 provides information about the model before any predictors have been entered. That is, is the intercept significant and what would be the effects if the list of predictors was entered? This Block 0 step is rarely of any interest except as a comparison for fit (-2 log likelihood, which will be discussed later), so most researchers will skip to Block 1, which I do here. I have also omitted the fit information which we will discuss in connection with multiple regression (see the subsequent handout “Multiple Logistic Regression and Model Fit”).

SPSS

Block 1: Method = Enter

Iteration History [a,b,c,d]

Iteration		-2 Log likelihood	Coefficients	
			Constant	youth
Step 1	1	1078.247	1.273	-.920
	2	1069.851	1.490	-1.133
	3	1069.817	1.505	-1.148
	4	1069.817	1.505	-1.148

a. Method: Enter
b. Constant is included in the model.
c. Initial -2 Log Likelihood: 1091.769
d. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B)	
								Lower	Upper
Step 1 [a]	agegrp	1.148	.235	23.854	1	<.001	3.152	1.988	4.996
	Constant	.357	.220	2.619	1	.106	1.429		

a. Variable(s) entered on step 1: agegrp.

R

```
> logmod <- glm(voted ~ agegrp, data = d, family = "binomial")
> summary(logmod)

Call:
glm(formula = voted ~ agegrp, family = "binomial", data = d)

Coefficients:
            Estimate Std. Error z value   Pr(>|z|)
(Intercept)   0.3567     0.2204   1.618      0.106
agegrp        1.1480     0.2351   4.884 0.00000104 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1091.8  on 1091  degrees of freedom
Residual deviance: 1069.8  on 1090  degrees of freedom
AIC: 1073.8

Number of Fisher Scoring iterations: 4
```

[1] These results use a random sample taken from a YouGov/Harvard/MIT survey called the Cooperation Election Study (CES) conducted by Stephen Ansolabehere & Brian Schaffner, 2022, "CES Common Content, 2021", https://doi.org/10.7910/DVN/OPQOCU

```
> #easy way to get odds ratios
> exp(cbind(OR=coef(logmod), confint(logmod)))
Waiting for profiling to be done...
                  OR     2.5 %   97.5 %
(Intercept) 1.428571 0.9310293 2.216340
agegrp      3.151913 1.9765107 4.980886
```

Notice that the odds ratio, given as Exp(B) in SPSS and OR in R, obtained from these analyses is 3.15, and this number is exactly same that we would obtain if we used the frequencies from the 2 × 2 Contingency Chi-square handout.

$$OR = e^{B_1} = \frac{ad}{cb} = \frac{(35)(824)}{(183)(50)} = \frac{28,840}{9,150} = 3.15$$

Sample Write-up

A logistic regression model was tested with voter age category as a predictor of whether the respondent voted. Results indicated a significant association between age category and voting, B = 1.148, SE = .24, p < .001, OR = 3.15, 95%CI[1.99,5.00]. The odds ratio indicated that older voters were approximately three times more likely to vote than the youngest age group. *[note that the chi-square model fit information and pseudo-R^2 might also be provided but we will discuss that information in detail when we discuss multiple logistic regression]*

Continuous Predictor

To illustrate a logistic regression with a continuous predictor, I used data from the Late Life Study of Social Exchanges (LLSSE; Sorkin & Rook, 2004) to predict self-reported heart disease (yes = 1) with depression. Depression symptomatology was measured with the the brief 9-item version (Santor & Coyne, 1997) of the Center for Epidemiologic Studies-Depression scale (Radloff, 1977).

```
GET FILE='c:\jason\spsswin\cdaclass\heart.sav'.

logistic regression vars=w1hheart with w1cesd9
   /print=summary ci(95) iter(1).
```

Block 1: Method = Enter

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients: Constant	Coefficients: 9-item CES-D
Step 1	1	656.638	-1.425	.019
	2	647.850	-1.737	.030
	3	647.782	-1.770	.032
	4	647.782	-1.771	.032

a. Method: Enter
b. Constant is included in the model.
c. Initial -2 Log Likelihood: 650.262
d. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	95% C.I.for EXP(B) Upper
Step 1[a]	9-item CES-D	.032	.020	2.592	1	.107	1.032	.993	1.073
	Constant	-1.771	.142	155.434	1	<.001	.170		

a. Variable(s) entered on step 1: 9-item CES-D.

R

```
> #simple logistic with continuous predictor
> logmod <- glm(w1hheart ~ w1cesd9, data = d, family = "binomial")
> summary(logmod)


Call:
glm(formula = w1hheart ~ w1cesd9, family = "binomial", data = d)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.7990  -0.6119  -0.5774  -0.5607   1.9636

Coefficients:
            Estimate Std. Error z value            Pr(>|z|)
(Intercept) -1.77061    0.14202  -12.47 <0.0000000000000002 ***
w1cesd9      0.03170    0.01969    1.61               0.107
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 650.26  on 723  degrees of freedom
Residual deviance: 647.78  on 722  degrees of freedom
AIC: 651.78


Number of Fisher Scoring iterations: 4

> #get odds ratio
> exp(cbind(OR=coef(logmod), confint(logmod)))

                  OR     2.5 %    97.5 %
(Intercept) 0.170230 0.1280805 0.2236314
w1cesd9     1.032205 0.9920395 1.0719640
```

Sample Write-up

A logistic regression analysis was conducted to examine whether there was a risk of heart disease for those with greater depression symptomatology. Results indicated a nonsignificant relationship, B = .03, SE = .20, p = .11, OR = 1.03, 95%CI[.99,1.07], with just a slight (approximately 3%) increase in risk of heart disease for each unit increase in depression. *[note that the chi-square model fit information and pseudo-R^2 might also be provided but we will discuss that information in detail when we discuss multiple logistic regression]*

References

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied psychological measurement*, *1*(3), 385-401.

Santor, D. A., & Coyne, J. C. (1997). Shortening the CES–D to improve its ability to detect cases of depression. *Psychological assessment*, *9*(3), 233-243.

Sorkin, D. H., & Rook, K. S. (2004). Interpersonal control strivings and vulnerability to negative social exchanges in later life. *Psychology and Aging*, *19*(4), 555-564.