

Multiple Logistic Example

To illustrate multiple logistic regression, I again used data from the Late Life Study of Social Exchanges (LLSSE; Sorkin & Rook, 2004) to predict self-reported heart disease. Predictors included sex (*wlsex*; men=0, women=1), vigorous physical activity (*wlactiv*), depression symptomatology from the brief 9-item version (Santor & Coyne, 1997) of the Center for Epidemiologic Studies-Depression scale (Radloff, 1977), and a measure of negative social exchanges (*wlneg*; Newsom, Rook, Nishishiba, Sorkin, & Mahan, 2005), which assesses the frequency of interpersonal conflicts.

SPSS¹

```
logistic regression vars=wlhheart with wlsex wlactiv wlcesd9 wlneg
/print=summary ci(95) iter(1).
```

Exerpts from output:

Block 0: Beginning Block

Iteration History^{a,b,c}

Iteration		-2 Log likelihood	Coefficients Constant
Step 0	1	633.358	-1.329
	2	625.763	-1.581
	3	625.718	-1.602
	4	625.718	-1.603

a. Constant is included in the model.

b. Initial -2 Log Likelihood: 625.718

c. Estimation terminated at iteration number 4 because parameter estimates changed by less than .001.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	23.238	4	<.001
	Block	23.238	4	<.001
	Model	23.238	4	<.001

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	602.480 ^a	.033	.055

a. Estimation terminated at iteration number 5 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	A1-sex of R	-.978	.214	20.824	1	<.001	.376	.247	.572
	N11-how often vigorous phy activities (Health)	-.041	.048	.721	1	.396	.960	.874	1.055
	9-item CES-D	.035	.022	2.550	1	.110	1.036	.992	1.082
	negative exchanges-total (mean)	.068	.186	.132	1	.716	1.070	.743	1.542
	Constant	-1.199	.206	33.922	1	<.001	.301		

a. Variable(s) entered on step 1: A1-sex of R, N11-how often vigorous phy activities (Health), 9-item CES-D, negative exchanges-total (mean).

¹ Note that the ci(95) keyword uses whole numbers to refer to percents not decimals. Also, adding the word *goodfit* to the /print subcommand will produce the Hosmer-Lemeshow test if desired (not recommended unless N > 400).

R

The `glm` procedure in R does not provide the likelihood ratio ("chi-square" test), so I used the `lmttest` package to compute that. The chi-square is the difference between the Block 0 (or null) model and the full model, so each has to be tested separately and then the `lrtest` function computes the difference between the -2 log likelihoods. To test the null model, the `glm` function is used but with a 1 instead of the predictors. Note that the N must be the same, so I use a listwise deletion routine for the variables in the full model before testing either model (always be sure to check descriptives to make sure it was done correctly before proceeding). The `modEVA` package is used to obtain pseudo- R^2 values.

```
> #listwise deletion so Ns the same for nested test
> d = d[complete.cases(d[,c("wlhheart", "wlsex", "wlactiv", "wlcesd9", "wlneg")]),]

> #test null/block 0 model--a 1 is used in place of variables--for chi-square
> rm(logmodn)
> logmodn <- glm(wlhheart ~ 1, data = d, family = "binomial")
> summary(logmodn)

Call:
glm(formula = wlhheart ~ 1, family = "binomial", data = d)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.6025      0.1018  -15.75  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 625.72  on 691  degrees of freedom
Residual deviance: 625.72  on 691  degrees of freedom
AIC: 627.72

Number of Fisher Scoring iterations: 3

> rm(logmod2)
> logmod2 <- glm(wlhheart ~ wlsex + wlactiv + wlcesd9 + wlneg, data = d, family = "binomial")
> summary(logmod2)

Call:
glm(formula = wlhheart ~ wlsex + wlactiv + wlcesd9 + wlneg, family = "binomial",
    data = d)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.19911      0.20588  -5.824 5.74e-09 ***
wlsex        -0.97841      0.21440  -4.563 5.03e-06 ***
wlactiv      -0.04065      0.04788  -0.849  0.396
wlcesd9       0.03539      0.02216   1.597  0.110
wlneg        0.06780      0.18643   0.364  0.716
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 625.72  on 691  degrees of freedom
Residual deviance: 602.48  on 687  degrees of freedom
AIC: 612.48

Number of Fisher Scoring iterations: 4

> #easy way to get odds ratios
> exp(cbind(OR=coef(logmod2), confint(logmod2)))
waiting for profiling to be done...
              OR      2.5 %      97.5 %
(Intercept) 0.3014610 0.1998632 0.4485558
wlsex       0.3759076 0.2458771 0.5707173
wlactiv     0.9601616 0.8724972 1.0530681
wlcesd9     1.0360220 0.9909311 1.0812665
wlneg       1.0701472 0.7324376 1.5282734

> #get chi-square (LR) test comparing logmodn and logmod2
> library(lmttest)
> lrtest(logmodn, logmod2)
Likelihood ratio test

Model 1: wlhheart ~ 1
Model 2: wlhheart ~ wlsex + wlactiv + wlcesd9 + wlneg
#Df LogLik Df Chisq Pr(>Chisq)
1 1 -312.86
2 5 -301.24 4 23.238 0.0001135 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> #obtain psuedo-R-sq values with modEVA package
> library(modEVA)
> RsqGLM(model=logmod2) #model on right of equal sign is name I gave my model above
$CoxSnell
[1] 0.03302352

$Nagelkerke
[1] 0.05548855

$McFadden
[1] 0.03713834

$Tjur
[1] 0.03423869

$sqPearson
[1] 0.03373417
```

Sample Write-Up

To identify factors that predict self-reported heart disease in a sample of older adults, a multiple logistic regression analysis was conducted, simultaneously entering gender, self-reported physical activity, depression scores, and negative social exchanges into the model. The results indicated that, together, the predictors accounted for a significant amount of variance in success, likelihood ratio $\chi^2(4) = 23.238$, $p < .001$. The Nagelkerke pseudo- R^2 indicated approximately 6% of the variance in heart disease was accounted for by the predictors overall. Out of all of the predictors in the model, only gender was a significant independent predictor of heart disease, $B = -.978$, $SE = .214$, $p < .001$, with women more than two and a half times less likely to report heart disease, $OR = .376$, 95% CI[.247,.572] (where the odds for men vs. women = $1/.376 = 2.660$) after controlling for activity level, $B = -.041$, $SE = .048$, $p = .396$, $OR = .960$, 95% CI[.847,1.1055], depression, $B = .035$, $SE = .022$, $p = .110$, $OR = 1.036$, 95% CI[.992,1.082], and negative social exchanges, $B = .068$, $SE = .186$, $p = .716$, $OR = 1.070$, 95% CI[.743,.1.542].²

Nested Test Example

Any two nested models in which the same N is used and one model has a subset of predictors can be compared with a likelihood ratio test. This is akin to the F -test for the change in R -square, although we are only getting significance test, not the increment in R -square (I see no reason you could not also report the difference in pseudo R -square as long as you refer to it as an “approximate” increment). I’ll illustrate a comparison of the model with just `wlcesd9` and the full model. In SPSS (omitted here), you can add `/method=enter` subcommands to the logistic regression command as in OLS regression, and, in R (below), you can conduct the significance test with `lrtest` function in the same manner we used above for comparing to the null model. This analysis answers the question as to whether adding the three additional variables (`wlsex`, `wlactiv`, and `wlneg`) accounted for a significant amount of additional variance in the outcome. I omitted some of the output.

```
> #listwise deletion so Ns the same for nested test
> d = d[complete.cases(d[,c("wlhheart", "wlsex", "wlactiv", "wlcesd9", "wlneg")]),]

> rm(logmod1)
> #simple logistic with continuous predictor
> logmod1 <- glm(wlhheart ~ wlcesd9, data = d, family = "binomial")
> summary(logmod1)
Null deviance: 625.72 on 691 degrees of freedom
Residual deviance: 623.72 on 690 degrees of freedom

> rm(logmod2)
> logmod2 <- glm(wlhheart ~ wlsex + wlactiv + wlcesd9 + wlneg, data = d, family = "binomial")
> summary(logmod2)
Null deviance: 625.72 on 691 degrees of freedom
Residual deviance: 602.48 on 687 degrees of freedom
AIC: 612.48

> #get LR chi-square nested test
> library(lmtest)
> lrtest(logmod2, logmod1)
Likelihood ratio test
Model 1: wlhheart ~ wlsex + wlactiv + wlcesd9 + wlneg
Model 2: wlhheart ~ wlcesd9
```

² Had negative social exchanges been significant, we might say that the odds of heart disease increased by about 7% for each unit increase on the scale, $OR = 1.070$. Depending on the number of predictors and whether there is a table used to present results, non-significant coefficients might or might not be reported in practice. (Please present nonsignificant results for the HW, however).

```
#Df LogLik Df Chisq Pr(>Chisq)
1 5 -301.24
2 2 -311.86 -3 21.239 9.391e-05 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The likelihood ratio test is simply the difference between the two -2 log likelihood values (R calls them “residual deviance”): $\chi^2 = 623.72 - 602.48 = 21.239$. The result is a chi-square value tested with $df = 690 - 687 = 3$. Note that the `lmtest` package takes the difference in loglikelihoods (rather than $-2 \times$ the likelihood), which is half the -2 log likelihood (residual deviance) values, and then it multiplies the difference by 2 at the end to arrive at the same result.

References

- Newsom, J. T., Rook, K. S., Nishishiba, M., Sorkin, D. H., & Mahan, T. L. (2005). Understanding the relative importance of positive and negative social exchanges: Examining specific domains and appraisals. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 60(6), P304-P312.
- 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.