Logistic Regression Examples

Simple Logistic Examples
The Quinnipiac polling 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 “Analysis of Contingency Tables.”

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

SPSS
Block 0: Beginning Block

Block 1: Method = Enter

CIs in the table below have been corrected (prior result was incorrect—CIs did not include 1 and should have). Error was in syntas using ci(.95) which is incorrect instead ci(95) which is correct.

R
> logmod <- glm(response ~ ind, data = mydata, family = "binomial")
> summary(logmod)

Coefficients:

| Estimate | Std. Error | z value | Pr(>|z|) |
|----------|------------|---------|----------|
| (Intercept) | 0.07136 | 0.07559 | 0.944 | 0.345 |
| ind | 0.15019 | 0.14186 | 1.059 | 0.290 |

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1358.1 on 981 degrees of freedom
Residual deviance: 1357.0 on 980 degrees of freedom

AIC: 1361

Number of Fisher Scoring iterations: 3

Number of Fisher Scoring iterations: 4

> #easy way to get odds ratios
> exp(coef(logmod))

       OR     2.5 %   97.5 %
(Intercept)  1.073964 0.9261414 1.245707
ind         1.162050 0.8804390 1.535866

> #obtain psuedo-R-sq values with modEvA package
> install.packages("modEvA") # install at first use
> library(modEvA)
> RsqGLM(model=logmod) #model on right of equal sign is name I gave my model above

$CoxSnell
[1] 0.001143395

$Nagelkerke
[1] 0.001526185

$McFadden
[1] 0.0008271988

$Tjur
[1] 0.001142238

$sqPearson
[1] 0.001142238
Multiple Logistic Examples

To illustrate multiple logistic regression, I used data from the Late Life Study of Social Exchanges (LLSSE; Sorkin & Rook, 2004) to predict self-reported heart disease. Predictors include gender (\texttt{w1sex; men=0, women=1}), vigorous physical activity (\texttt{w1activ}), 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 (\texttt{w1neg}; Newsom, Rook, Nishishiba, Sorkin, & Mahan), which assesses the frequency of interpersonal conflicts. To save space, I only include one output.

**SPSS**

```spss
LOGISTIC REGRESSION /VARIABLES=w1hheart WITH w1sex w1activ w1cesd9 w1neg /PRINT=SUMMARY CI(95) GOODFIT ITER(1).
```

**R**

```r
#need to convert data types in order to avoid dummy coding
mydata <-Transform(w1activ = as.numeric(w1activ), data=mydata)

logmod <- glm(w1hheart ~ w1sex + w1activ + w1cesd9 + w1neg, data = mydata, family = "binomial")
summary(logmod)

#obtain pseudo-R-sq values with modEvA package
library(modEvA)
RsqGLM(model=logmod)  #model on right of equal sign is name I gave my model above
```

**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, \( B = -.041, SE = .721, p = .396, 95\% CI [.960,1.074] \), depression, \( B = .035, SE = .110, p = .100, 95\% CI [.952,1.062] \), and negative social exchanges, \( B = .068, SE = .716, p = .976, 95\% CI [.374,1.542] \).

Note that the ci(95) keyword uses whole numbers to refer to percents not decimals.

Also, make sure that your dependent variable is coded as 0,1. I needed to recode mine, using:

```r
library(memisc)
mydata$w1hheart = recode(mydata$w1hheart, 0 <- 1, 1 <- 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 there is a table, non-significant coefficients might be reported in the text.