Example: Multinomial Regression

A retrospective cohort study conducted at the University of Massachusetts sought to evaluate the attitudes and characteristics associated with mammography. A total of $n = 412$ women were classified into one of the following categories:

- *never* had a mammogram ($Y_i = 0$)
- had a mammogram *within the past year* ($Y_i = 1$)
- had a mammogram, but *more than one year ago* ($Y_i = 2$).

Covariates of of interest were based on responses to a questionnaire and included:

SYMP REQ: 1=SA, 2=A, 3=D, 4=SD

"*You do not need a mammogram unless you develop symptoms*"

PERC BEN: 5, . . . , 20

"*Score the potential benefit of a mammogram*"

FAM HIST: 0=No, 1=Yes

"*Do you have a mother or sister with breast cancer?*"

TAUGHT BSE: 0=No, 1=Yes

"*Were you taught how to do a breast self-exam?*"

MAMM DET: 1=NL, 2=SL, 3=VL

"*How likely would a mammogram detect a new breast tumor?*"

(a) Recode the response variate such that response value could be treated as ordered.

- Recode:
  - *never* had a mammogram ($Y_i = 0$)
  - had a mammogram *within the past year* ($Y_i = 2$)
  - had a mammogram, but *more than one year ago* ($Y_i = 1$).

(b) To begin, we focus on family history. Print out a cross-classification of mammography experience and family history.

- See the SAS code

(c) Write down a generalized logit model pertaining to the 2×3 table.

$$\log\left(\frac{\pi_{ij}}{\pi_{i0}}\right) = \beta_{0j} + \beta_{1j} X_i \quad j = 1, 2$$

X_i: Family history (1=YES, 0=NO)

$\pi_{ij} = P(Y_i = j | X_i)$

(d) Fit the generalized logit model to the table, by hand.

$$\widehat{\beta}_{01} = log(63/220) = -1.25$$

$$\widehat{\beta}_{11} = log(11/14) - log(63/220) = 1.01$$

$$\widehat{\beta}_{02} = log(85/220) = -0.95$$

$$\widehat{\beta}_{12} = log(19/14) - log(85/220) = 1.26$$

- OR: $\exp(\widehat{\beta}_{11}) = 2.74 \quad \exp(\widehat{\beta}_{12}) = 3.51$

(e) Re-estimate the family history effect parameters, this time using 2×2 table calculations.

- Table $Y = 0$ vs $Y = 1$

$$OR = (220 * 11)/(63 * 14) = 2.744$$

- Table $Y = 0$ vs $Y = 2$

$$OR = (220 * 19)/(85 * 14) = 3.51$$

- OR estimates are identical.

(f) Estimate the standard errors of $\widehat{\beta}_{11}$ and $\widehat{\beta}_{12}$.
NOTE: standard error of logOR in 2×2 table:
$SE = \sqrt{1/n_{11} + 1/n_{10} + 1/n_{01} + 1/n_{00}}$

$\widehat{SE}(\beta_{11}) = \sqrt{1/220 + 1/63 + 1/14 + 1/11} =$
0.4275

$\widehat{SE}(\beta_{12}) = \sqrt{1/220 + 1/85 + 1/19 + 1/11} =$
0.3747

(g) Re-fit the GL model, this time using PROC LOGISTIC. Compare your parameter estimates and SEs to those obtained by hand.

- See the SAS code (Results are identical)

(h) Interpret $\exp\{\widehat{\beta}_{11}\}$.

Odds ratio of having a mammogram comparing with and without family history, given that the women did not have a mammogram within the past year.

(i) Carry out a Wald test of $H_0 : \beta_{11} = \beta_{12} = 0$.

Test statistic: $X_w^2 = 12.01$

P-value: 0.0025

Reject H0.

(j) Carry out a likelihood ratio test of the hypothesis listed in (i).

Full model -2 log L: 792.340

Reduced model -2 log L: 805.198

Test statistic: $X_L^2 = 805.198 - 792.340 = 12.858$

$12.858 > 5.99 = \chi^2_{2,0.95}$

Reject H0.

(k) Write out an appropriate proportional odds model, again focusing only on family history.

$$\log\{\frac{P(Y_i \le j|X_i)}{P(Y_i > j|X_i)}\} = \gamma_{0j} + \gamma_1 X_i \quad j = 0, 1$$

(ℓ) Fit the PO model using PROC LOGISTIC.

See the SAS code

(m) Interpret $\exp\{\widehat{\gamma}_1\}$ based on the PO model.

- $\exp\{\widehat{\gamma}_1\} = 0.355$

  Odds ratio of being in lower mammogram categories comparing with and without family history is 0.355

(n) Estimate the probability that a woman with a family history of breast cancer had a mammogram more than one year ago.

$$P(Y_i \leq 0|X_i = 1) = \frac{\exp(\widehat{\gamma_{00}} + \widehat{\gamma_1})}{1+\exp(\widehat{\gamma_{00}} + \widehat{\gamma_1})} = 0.344$$

$$P(Y_i \leq 1|X_i = 1) = \frac{\exp(\widehat{\gamma_{01}} + \widehat{\gamma_1})}{1+\exp(\widehat{\gamma_{01}} + \widehat{\gamma_1})} = 0.547$$

$P(Y_i = 1|X_i = 1) = P(Y_i \leq 1|X_i = 1) - P(Y_i \leq 0|X_i = 1) = 0.203$

(o) Carry out a score test of the proportionality assumption.

$X_s^2 = 0.5951 < 3.84 = \chi^2_{1,0.95}$

Cannot reject H0

There is no strong evidence that the proportionality assumption is not satisfied.

(p) List the two models being compared in the test for proportionality.

- H0 (3 parameters)

$$\log\{\frac{P(Y_i \leq j|X_i)}{P(Y_i > j|X_i)}\} = \gamma_{0j} + \gamma_1 X_i \quad j = 0, 1$$

- H1 (4 parameters)

$$\log\{\frac{P(Y_i \leq j|X_i)}{P(Y_i > j|X_i)}\} = \gamma_{0j} + \gamma_{1j} X_i \quad j = 0, 1$$