1 Collapsibility and generalized linear model

References

- Harvard Schan School BIO233 note page 233
- Greenland et al. Confounding and collapsibility in causal inference. Statistical Sicence 1999;14:29-46.
-

Demonstrates if the effect measure of interest is collapsible over variable Z, which is may be a predictor, but is independent of X ($Z \coprod X$).

$$
\begin{aligned}
&\text{Conditional model conditioning on Z}\\
g(E[Y|X,Z]) &= \beta_0 + \beta_1 X + \beta_2 Z\\
E[Y|X,Z] &= h(\beta_0 + \beta_1 X + \beta_2 Z)\\
\\
&\text{Marginal model}\\
g(E[Y|X]) &= \beta_0^* + \beta_1^* X\\
E[Y|X] &= h(\beta_0^* + \beta_1^* X)\\
\\
&\text{where}\\
g(\cdot)&\text{Link function}\\
h(\cdot)&\text{Inverse link function}
\end{aligned}
$$

1.1 Identity link

$$
\begin{aligned}
& \text{Weighted mean to marginalize as to } Z \\
E[Y|X] &= \int_z E[Y|Z,X] f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Inverse of identity is identity} \\
&= \int_z (\beta_0 + \beta_1 X + \beta_2 Z) f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Weighted mean of a constant is a constant} \\
&= \beta_0 + \int_z (\beta_1 X + \beta_2 Z) f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Conditional on } X,\ X \text{ and } \beta_1 X \text{ is a constant} \\
&= \beta_0 + \beta_1 X + \int_z (\beta_2 Z) f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Last term is expectation of } Z \text{ conditional on } X \\
&= \beta_0 + \beta_1 X + \beta_2 E[Z|X] \\
& \text{If } Z \coprod X, \text{ conditional expectation = marginal} \\
&= \beta_0 + \beta_1 X + \beta_2 E[Z] \\
& \text{Collect constants} \\
&= (\beta_0 + \beta_2 E[Z]) + \beta_1 X \\
& \text{This expression has same form as marginal model} \\
&= \beta_0^* + \beta_1^* X \\
& \text{Thus,} \\
& \begin{cases} \beta_0^* = \beta_0 + \beta_2 E[Z] \\ \beta_1^* = \beta_1 \end{cases}
\end{aligned}
$$

1.2 Identity link (with effect modification by binary Z)

$$
\begin{aligned}
& \text{Weighted mean to marginalize as to } Z \\
E[Y|X] &= \int_z E[Y|Z,X] f_{Z|X}(Z=z|X)\mathrm{d}z \\
& \text{Inverse of identity is identity} \\
&= \int_z (\beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ) f_{Z|X}(Z=z|X)\mathrm{d}z \\
& \text{Weighted mean of a constant is a constant} \\
&= \beta_0 + \int_z (\beta_1 X + \beta_2 Z + \beta_3 XZ) f_{Z|X}(Z=z|X)\mathrm{d}z \\
& \text{Conditional on } X,\ X \text{ and } \beta_1 X \text{ is a constant} \\
&= \beta_0 + \beta_1 X + \int_z (\beta_2 Z + \beta_3 XZ) f_{Z|X}(Z=z|X)\mathrm{d}z \\
& \text{Second last term is expectation of } Z \text{ conditional on } X \\
&= \beta_0 + \beta_1 X + \beta_2 E[Z|X] + \int_z (\beta_3 XZ) f_{Z|X}(Z=z|X)\mathrm{d}z \\
& \text{If } Z \coprod X, \text{ conditional expectation} = \text{marginal} \\
&= \beta_0 + \beta_1 X + \beta_2 E[Z] + \int_z (\beta_3 XZ) f_{Z|X}(Z=z|X)\mathrm{d}z \\
& \text{By similar logic} \\
&= \beta_0 + \beta_1 X + \beta_2 E[Z] + \beta_3 X E[Z|X] \\
&= \beta_0 + \beta_1 X + \beta_2 E[Z] + \beta_3 X E[Z] \\
& \text{Collect constants and } X \text{ terms} \\
&= (\beta_0 + \beta_2 E[Z]) + (\beta_1 + \beta_3 E[Z])X \\
& \text{If } Z \text{ is binary} \\
&= (\beta_0 + \beta_2 E[Z]) + (\beta_1(1 - E[Z]) + (\beta_1 + \beta_3)E[Z])X \\
& \text{This expression has same form as marginal model} \\
&= \beta_0^* + \beta_1^* X \\
& \text{Thus,} \\
& \begin{cases} \beta_0^* = \beta_0 + \beta_2 E[Z] \\ \beta_1^* = \beta_1(1 - E[Z]) + (\beta_1 + \beta_3)E[Z] \end{cases}
\end{aligned}
$$

Therefore, the marginal effect of X, β_1^* is the weighted average of the stratum-specific effects of X, β_1 for $Z = 0$ and $\beta_1 + \beta_3$ for $Z = 1$ weighted by the prevalence of $Z = 0$ and $Z = 1$, respectively.

1.3 Log link

$$
\begin{aligned}
& \text{Weighted mean to marginalize as to } Z \\
E[Y|X] &= \int_z E[Y|Z,X] f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Inverse of log is exp} \\
&= \int_z e^{(\beta_0+\beta_1 X+\beta_2 Z)} f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Decompose into a product} \\
&= \int_z e^{\beta_0} e^{\beta_1 X} e^{\beta_2 Z} f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Weighted mean of a transformed constant is a constant} \\
&= e^{\beta_0} \int_z e^{\beta_1 X} e^{\beta_2 Z} f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Conditional on } X\text{, } X \text{ and } \beta_1 X \text{ is a constant} \\
&= e^{\beta_0} e^{\beta_1 X} \int_z e^{\beta_2 Z} f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Last term is expectation of } e^{\beta_2 Z} \text{ conditional on } X \\
&= e^{\beta_0} e^{\beta_1 X} E[e^{\beta_2 Z}|X] \\
& \text{If } Z \coprod X\text{, conditional expectation = marginal} \\
&= e^{\beta_0} e^{\beta_1 X} E[e^{\beta_2 Z}] \\
& \text{Collect constants} \\
&= (e^{\beta_0} E[e^{\beta_2 Z}]) e^{\beta_1 X} \\
& \text{This expression has same form as marginal model} \\
&= e^{(\beta_0^*+\beta_1^* X)} = e^{\beta_0^*} e^{\beta_1^* X} \\
& \text{Thus,} \\
& \begin{cases} e^{\beta_0^*} = e^{\beta_0} E[e^{\beta_2 Z}] \\ \beta_1^* = \beta_1 \end{cases}
\end{aligned}
$$

1.4 Logit link

$$
\begin{aligned}
&\text{Weighted mean to marginalize as to } Z \\
E[Y|X] &= \int_z E[Y|Z,X] f_{Z|X}(Z=z|X) \mathrm{d}z \\
&\text{Inverse of logit is expit} \\
&= \int_z expit(\beta_0 + \beta_1 X + \beta_2 Z) f_{Z|X}(Z=z|X) \mathrm{d}z \\
&= \int_z \frac{e^{\beta_0 + \beta_1 X + \beta_2 Z}}{1 + e^{\beta_0 + \beta_1 X + \beta_2 Z}} f_{Z|X}(Z=z|X) \mathrm{d}z \\
&= \int_z \frac{e^{\beta_0} e^{\beta_1 X} e^{\beta_2 Z}}{1 + e^{\beta_0} e^{\beta_1 X} e^{\beta_2 Z}} f_{Z|X}(Z=z|X) \mathrm{d}z \\
&\text{Weighted mean of a transformed constant is a constant} \\
&= e^{\beta_0} \int_z \frac{e^{\beta_1 X} e^{\beta_2 Z}}{1 + e^{\beta_0} e^{\beta_1 X} e^{\beta_2 Z}} f_{Z|X}(Z=z|X) \mathrm{d}z \\
&\text{Conditional on } X,\ X \text{ and } \beta_1 X \text{ is a constant} \\
&= e^{\beta_0} e^{\beta_1 X} \int_z \frac{e^{\beta_2 Z}}{1 + (e^{\beta_0} e^{\beta_1 X}) e^{\beta_2 Z}} f_{Z|X}(Z=z|X) \mathrm{d}z \\
&= e^{\beta_0} e^{\beta_1 X} E\left[\frac{e^{\beta_2 Z}}{1 + (e^{\beta_0} e^{\beta_1 X}) e^{\beta_2 Z}} |X \right] \\
&\text{No further transformation is possible} \\
&\text{This expression does not have same form as marginal model} \\
&\neq \frac{e^{\beta_0^*} e^{\beta_1^* X}}{1 + e^{\beta_0^*} e^{\beta_1^* X}}
\end{aligned}
$$

There is no way to separate β_1 and β_2 in the denominator because we cannot apply the expectation to the denominator only, and it is not possible to transform this to the form of the marginal model.

1.5 Log link (rate with varying observation time T)

I am not really sure about this one.

$$\begin{aligned}
& \text{Weighted mean to marginalize as to } Z \\
E[Y|X] &= \int_z E[Y|Z,X] f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Inverse of log is exp} \\
&= \int_z e^{(\beta_0 + \beta_1 X + \beta_2 Z + log(T))} f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Decompose into a product} \\
&= \int_z e^{\beta_0} e^{\beta_1 X} e^{\beta_2 Z} T f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Weighted mean of a transformed constant is a constant} \\
&= e^{\beta_0} \int_z e^{\beta_1 X} e^{\beta_2 Z} T f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Conditional on } X,\ X \text{ and } \beta_1 X \text{ is a constant} \\
&= e^{\beta_0} e^{\beta_1 X} \int_z e^{\beta_2 Z} T f_{Z|X}(Z=z|X) \mathrm{d}z \\
& \text{Last term is expectation of } e^{\beta_2 Z} T \text{ conditional on } X \\
&= e^{\beta_0} e^{\beta_1 X} E[e^{\beta_2 Z} T|X] \\
& Z \coprod X \text{ is insufficient to make } E[e^{\beta_2 Z} T|X] \text{ independent of } X \\
& \text{This expression does not have same form as marginal model} \\
&\neq e^{\beta_0^*} e^{\beta_1^* X} T
\end{aligned}$$

If the observation time T is the same for all observations (constant; essentially just count data), then T can come out of $E[e^{\beta_2 Z} T|X]$, and the expression conform to the marginal model expression.

2 Collapsibility and (generalized) linear mixed model

The marginal model for correlated data and the mixed model are compared.

References

- Applied Longitudinal Analysis. http://www.hsph.harvard.edu/fitzmaur/ala2e/

$$
\begin{aligned}
& \text{Mixed effects model conditioning on random effects } b_i \\
g(E[Y_i|X_i, b_i]) &= X_i\beta + Z_i b_i \\
E[Y_i|X_i, b_i] &= h(X_i\beta + Z_i b_i) \\
\\
& \text{Marginal model} \\
g(E[Y_i|X_i]) &= X_i\beta^* \\
E[Y_i|X_i] &= h(X_i\beta^*) \\
\\
& \text{where} \\
g(\cdot) & \text{Link function} \\
h(\cdot) & \text{Inverse link function}
\end{aligned}
$$

2.1 Identity link

$$
\begin{aligned}
& \text{Marginalizing over random effects } b_i \text{ yields marginal model} \\
E[Y_i|X_i] &= E[E[Y_i|X_i, b_i]] \\
& \text{Inverse of identity is identity} \\
&= E[X_i\beta + Z_i b_i] \\
& \text{Data and fixed parameters are constant} \\
&= X_i\beta + Z_i E[b_i] \\
& E[b_i] \text{ is defined as } 0 \\
&= X_i\beta + Z_i(0) \\
&= X_i\beta \\
& \text{This expression has same form as marginal model} \\
&= X_i\beta^* \\
& \text{Thus, conditional effect = marginal effect} \\
\beta^* &= \beta
\end{aligned}
$$

2.2 Nonlinear link

$$
\begin{aligned}
& \text{Marginalizing over random effects } b_i \text{ yields marginal model} \\
E[Y_i|X_i] &= E[E[Y_i|X_i, b_i]] \\
& \text{Inverse of } g(\cdot) \text{ is } h(\cdot) \\
&= E[h(X_i\beta + Z_i b_i)] \\
& \text{If the link is not identity, } E[\cdot] \text{ cannot go into } h(\cdot) \\
&\neq h(E[X_i\beta + Z_i b_i]) \\
& \text{Thus, we cannot obtain } E[b_i] = 0. \text{ No further transformtion}
\end{aligned}
$$

2.3 Log link special case

The log link is another special case. The conditional coefficients for variables that do not have random effects (Only in X_i, but not in Z_i) are equivalent to the corresponding coefficients in the marginal model.