Background Statistical Theory For Logistic Regression

Let $i = 1, 2, \ldots, s$ index a set of s populations from which independent simple random samples of sizes n_i are available. Let Π_i be fraction of i-th population with target attribute.

Let $(y_i,\, n_i - y_i)$ denote numbers of subjects in i-th sample with and without the target attribute. The $\{y_i\}$ have independent binomial $Bin(n_{i,} \Pi_i)$ distributions.

Variation among $\{\Pi_i\}$ is described with logistic model

$$\Pi_i = \frac{exp(\eta + \boldsymbol{x}_i'\boldsymbol{\beta})}{\{1+exp(\eta+\boldsymbol{x}_i'\boldsymbol{\beta})\}} = \frac{exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A)}{\{1+exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A)\}} \text{ where } \boldsymbol{x}_{iA}'\boldsymbol{\beta}_A = \eta + \boldsymbol{x}_i'\boldsymbol{\beta}\ ;$$

i.e., $\boldsymbol{x}_{iA}' = (1,\, \boldsymbol{x}_i')$, $\boldsymbol{\beta}_A' = (\eta,\, \boldsymbol{\beta}')$. The $\{\boldsymbol{x}_{iA}'\}$ are rows of an $(s \times t)$ matrix $\boldsymbol{X}_A$ with rank t.

The parameters $\boldsymbol{\beta}_A$ are estimated by maximum likelihood method. Consider likelihood

$$\phi = \prod_{i=1}^{s} \binom{n_i}{y_i} \Pi_i^{y_i}(1-\Pi_i)^{n_i - y_i} = \prod_{i=1}^{s} \binom{n_i}{y_i} \frac{exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A y_i)}{\{1+exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A y_i)\}^{n_i}}$$

Since $ln\phi = ln\left\{\prod_{i=1}^{s}\binom{n_i}{y_i}\right\} + \sum_{i=1}^{s}\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A y_i - \sum_{i=1}^{s} n_i ln\{1+exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A)\}$

$$\begin{aligned}
\frac{\partial ln\phi}{\partial \boldsymbol{\beta}_A'} &= \sum_{i=1}^{s}\boldsymbol{x}_{iA}' y_i - \sum_{i=1}^{s} n_i\{1+exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A)\}^{-1}\{exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A)\}\boldsymbol{x}_{iA}' \\
&= \sum_{i=1}^{s}\boldsymbol{x}_{iA}'\{y_i - n_i\Pi_i(\boldsymbol{\beta}_A)\} \text{ where } \Pi_i(\boldsymbol{\beta}_A) = \frac{exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A)}{\{1+exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A)\}} \\
&= \sum_{i=1}^{s}\boldsymbol{x}_{iA}'\{y_i - \mu_i(\boldsymbol{\beta}_A)\} \text{ where } \mu_i(\boldsymbol{\beta}_A) = n_i\Pi_i(\boldsymbol{\beta}_A) \text{, and so} \\
\frac{\partial ln\phi}{\partial \boldsymbol{\beta}_A} &= \boldsymbol{X}_A'\{\boldsymbol{y} - \widehat{\boldsymbol{\mu}}(\boldsymbol{\beta}_A)\} \text{ where } \boldsymbol{y} = (y_1,\, y_2,\, \ldots,\, y_s)'
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2 ln\phi}{\partial \boldsymbol{\beta}_A \delta \boldsymbol{\beta}_A'} &= \frac{\partial}{\partial \boldsymbol{\beta}_A'}\left\{\sum_{i=1}^{s}\boldsymbol{x}_{iA}\{y_i - n_i\Pi_i(\boldsymbol{\beta}_A)\}\right\} = \frac{\partial}{\partial \boldsymbol{\beta}_A'}\left\{\sum_{i=1}^{s} -n_i\boldsymbol{x}_{iA}\boldsymbol{\Pi_i}(\boldsymbol{\beta}_A)\right\} \\
&= \sum_{i=1}^{s} -n_i\boldsymbol{x}_{iA}\frac{\{[1+exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A)][exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A)]\boldsymbol{x}_{iA}' - [exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A)]^2\boldsymbol{x}_{iA}'}{[1+exp(\boldsymbol{x}_{iA}'\boldsymbol{\beta}_A)]^2} \\
&= \sum_{i=1}^{s} -n_i\Pi_i(\boldsymbol{\beta}_A)[1-\Pi_i(\boldsymbol{\beta}_A)]\boldsymbol{x}_{iA}\boldsymbol{x}_{iA}' = -\boldsymbol{X}_A'\boldsymbol{D}_v\boldsymbol{X}_A
\end{aligned}$$

where $\boldsymbol{D}_v = Diag\,(v_1, v_2, \ldots, v_s)$ with $v_i = Var(y_i) = n_i\Pi_i(1-\Pi_i)$

The maximum likelihood equations are given by $\frac{\partial ln\phi}{\partial \boldsymbol{\beta}_A'} = 0$ or $\sum_{i=1}^{s}\boldsymbol{x}_{iA}'\left(y_i - n_i\Pi_i\left(\widehat{\boldsymbol{\beta}}_A\right)\right) = 0.$

These equations implicitly define $\widehat{\boldsymbol{\beta}}_A = \widehat{\boldsymbol{\beta}}_A(\boldsymbol{y})$ as a function of $\boldsymbol{y}$ or equivalently as a function of $\boldsymbol{p} = \boldsymbol{D}_{\boldsymbol{n}}^{-1}\boldsymbol{y}$ where $\boldsymbol{D}_{\boldsymbol{n}} = Diag\,(n_1,\, n_2,\, \ldots.,\, n_s)$. The elements of $\boldsymbol{p}$ are proportions $p_i = (y_i/n_i)$.

The linear Taylor series approximation to $\widehat{\boldsymbol{\beta}}_A$ as function of $\boldsymbol{p}$ is

$$\widehat{\boldsymbol{\beta}}_A(\boldsymbol{p}) = \widehat{\boldsymbol{\beta}}_A(\boldsymbol{\Pi}) + \left\{ \frac{\partial\, \widehat{\boldsymbol{\beta}}_A}{\partial \boldsymbol{p}'}\bigg|_{\boldsymbol{p}=\boldsymbol{\pi}} \right\}(\boldsymbol{p}-\boldsymbol{\Pi}) + O(\tfrac{1}{n})$$

where $\boldsymbol{\Pi} = (\Pi_1, \Pi_2, \ldots, \Pi_s)'$ and $n = \sum_{i=1}^{s} n_i$.

Asymptotics are n $\rightarrow \infty$ with $\left(\frac{n_i}{n}\right) = \phi_i$ fixed.

$\widehat{\boldsymbol{\beta}}_A(\boldsymbol{\Pi}) = \boldsymbol{\beta}_A$ since $\sum_{i=1}^{s} \boldsymbol{x}'_{iA} n_i \Pi_i = \sum_{i=1}^{s} \boldsymbol{x}'_{iA} n_i \frac{exp(\boldsymbol{x}'_{iA}\boldsymbol{\beta}_A)}{\left|1+exp(\boldsymbol{x}'_{iA}\boldsymbol{\beta}_A)\right|}$ by model specification;
i.e., $\boldsymbol{\beta}_A$ is solution of maximum likelihood equations when $y_i = n_i\Pi_i = \mu_i$.

Transpose and rewrite maximum likelihood equations to

$$\sum_{i=1}^{s} n_i \boldsymbol{x}_{iA} \Pi_i\left(\widehat{\boldsymbol{\beta}}_A\right) = \sum_{i=1}^{s} n_i \boldsymbol{x}_{iA} p_i$$

Take derivatives on both sides with respect to $\boldsymbol{p}'$

$$\sum_{i=1}^{s} n_i \boldsymbol{x}_{iA} \frac{\partial \Pi_i\left(\widehat{\boldsymbol{\beta}}_A\right)}{\partial \widehat{\boldsymbol{\beta}}'_A} \frac{\partial \widehat{\boldsymbol{\beta}}_A}{\partial \boldsymbol{p}'} = \sum_{i-1}^{s} n_i \boldsymbol{x}_{iA}$$

Since $\frac{\partial \Pi_i\left(\widehat{\boldsymbol{\beta}}_A\right)}{\partial \widehat{\boldsymbol{\beta}}'_A} = \Pi_i\left(\widehat{\boldsymbol{\beta}}_A\right)\left[1-\Pi_i\left(\widehat{\boldsymbol{\beta}}_A\right)\right]\boldsymbol{x}'_{iA}$, then the derivatives on both sides at $\boldsymbol{p} = \boldsymbol{\Pi}$ satisfy

$$\sum_{L=1}^{s} n_i \boldsymbol{x}_{iA} \boldsymbol{x}'_{iA} \Pi_i(1-\Pi_i) \left[\frac{\partial\, \widehat{\boldsymbol{\beta}}_A}{\partial \boldsymbol{p}'}\bigg|_{\boldsymbol{p}=\boldsymbol{\Pi}} \right] = \sum_{L=1}^{s} n_i \boldsymbol{x}_{iA}$$

$$(\boldsymbol{X}'_A \boldsymbol{D}_{\boldsymbol{v}} \boldsymbol{X}_A) \left[\frac{\partial\, \widehat{\boldsymbol{\beta}}_A}{\partial \boldsymbol{p}'}\bigg|_{\boldsymbol{p}=\boldsymbol{\Pi}} \right] = \boldsymbol{X}'_A \boldsymbol{D}_{\boldsymbol{n}}$$

Thus, linear Taylor series expansion for $\widehat{\boldsymbol{\beta}}_A(\boldsymbol{p})$ is

$$\widehat{\boldsymbol{\beta}}_A(\boldsymbol{p}) = \boldsymbol{\beta}_A + (\boldsymbol{X}'_A \boldsymbol{D}_{\boldsymbol{v}} \boldsymbol{X}_A)^{-1} \boldsymbol{X}'_A \boldsymbol{D}_{\boldsymbol{n}} (\boldsymbol{p}-\boldsymbol{\Pi}) + O(\tfrac{1}{n})$$

This implies $\mathcal{E}\left(\widehat{\boldsymbol{\beta}}_A\right) \rightarrow \boldsymbol{\beta}_A$ as $n \rightarrow \infty$

$$\begin{aligned} Var\left\{\widehat{\boldsymbol{\beta}}_A\right\} &\rightarrow (\boldsymbol{X}'_A \boldsymbol{D}_{\boldsymbol{v}} \boldsymbol{X}_A)^{-1} \boldsymbol{X}'_A \boldsymbol{D}_n \boldsymbol{D}_n^{-1} \boldsymbol{D}_v \boldsymbol{D}_{\boldsymbol{n}}^{-1} \boldsymbol{D}_{\boldsymbol{n}} \boldsymbol{X}_A (\boldsymbol{X}'_A \boldsymbol{D}_{\boldsymbol{v}} \boldsymbol{X}_A)^{-1} \\ &= (\boldsymbol{X}'_A \boldsymbol{D}_{\boldsymbol{v}} \boldsymbol{X}_A)^{-1} = \left[-\frac{\partial^2 ln\phi}{\partial \boldsymbol{\beta}_A \partial \boldsymbol{\beta}'_A} \right]^{-1} \end{aligned}$$

To the extent that $\boldsymbol{X}'_A \boldsymbol{D}_{\mathbf{n}}(\boldsymbol{p}-\boldsymbol{\Pi})$ is approximately multivariate normal, $\widehat{\boldsymbol{\beta}}_A$ is approximately multivariate normal.

Thus, for appropriate $\boldsymbol{X}_A$, $\widehat{\boldsymbol{\beta}}_A$ is approximately $MN\left(\boldsymbol{\beta}_A,\ (\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{X}_A)^{-1}\right)$.

The model predicted values for $\boldsymbol{\Pi}$ are $\boldsymbol{\Pi}\left(\widehat{\boldsymbol{\beta}}_A\right)$ for which the linear Taylor series approximation is

$$\begin{aligned}\boldsymbol{\Pi}\left(\widehat{\boldsymbol{\beta}}_A\right) &= \boldsymbol{\Pi}(\boldsymbol{\beta}_A) + \left[\left.\frac{\partial\boldsymbol{\Pi}(\widehat{\boldsymbol{\beta}}_A)}{\partial\widehat{\boldsymbol{\beta}}'_A}\right|_{\widehat{\boldsymbol{\beta}}_A=\boldsymbol{\beta}_A}\right]\left[\left(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A\right)\right] + O(\tfrac{1}{n}) \\ &= \boldsymbol{\Pi} + \boldsymbol{D_n}^{-1}\boldsymbol{D_v}\boldsymbol{X}_A\left(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A\right) + O(\tfrac{1}{n})\end{aligned}$$

Thus, $\mathcal{E}\left\{\boldsymbol{\Pi}\left(\widehat{\boldsymbol{\beta}}_A\right)\right\} \to \boldsymbol{\Pi}$ as $n \to \infty$

$$Var\left\{\boldsymbol{\Pi}\left(\widehat{\boldsymbol{\beta}}_A\right)\right\} \to \boldsymbol{D}_n^{-1}\boldsymbol{D_v}\boldsymbol{X}_{_A}(\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{X}_A)^{-1}\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{D}_n^{-1}$$

The residuals for the fitted model are $\left(\boldsymbol{p} - \boldsymbol{\Pi}\left(\widehat{\boldsymbol{\beta}}_A\right)\right)$.

The linear Taylor series approximation for the residuals is

$$\begin{aligned}\left(\boldsymbol{p} - \boldsymbol{\Pi}\left(\widehat{\boldsymbol{\beta}}_A\right)\right) &= (\boldsymbol{p} - \boldsymbol{\Pi}) - \boldsymbol{D}_n^{-1}\boldsymbol{D_v}\boldsymbol{X}_{_A}\left(\widehat{\boldsymbol{\beta}}_A - \boldsymbol{\beta}_A\right) + O(\tfrac{1}{n}) \\ &= \left(\boldsymbol{I} - \boldsymbol{D}_n^{-1}\boldsymbol{D_v}\boldsymbol{X}_{_A}(\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{X}_A)^{-1}\boldsymbol{X}'_A\boldsymbol{D_n}\right)(\boldsymbol{p} - \boldsymbol{\Pi}) + O(\tfrac{1}{n}) \\ &= \boldsymbol{H}(\boldsymbol{p} - \boldsymbol{\Pi}) + O(\tfrac{1}{n})\end{aligned}$$

$$\mathcal{E}\left\{\left(\boldsymbol{p} - \boldsymbol{\Pi}\left(\widehat{\boldsymbol{\beta}}_A\right)\right)\right\} \to 0 \text{ as } n \to \infty$$

$$Var\left\{\left(\boldsymbol{p} - \widehat{\boldsymbol{\Pi}}\left(\widehat{\boldsymbol{\beta}}_A\right)\right)\right\} \to \boldsymbol{H}\boldsymbol{D_n}^{-1}\boldsymbol{D}_v\boldsymbol{D_n}^{-1}\boldsymbol{H}'$$

$$\begin{aligned}&= \left[\boldsymbol{D_n}^{-1}\boldsymbol{D_v}\boldsymbol{D_n}^{-1} - \boldsymbol{D_n}^{-1}\boldsymbol{D_v}\boldsymbol{X}_A(\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{X}_A)^{-1}\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{D_n}^{-1}\right] \times \\ &\quad \left[\boldsymbol{I} - \boldsymbol{D}_n^{-1}\boldsymbol{D_v}\boldsymbol{X}_{_A}\left(\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{X}_A\right)^{-1}\boldsymbol{X}'_A\boldsymbol{D_n}\right]' \\ &= \boldsymbol{D_n}^{-1}\boldsymbol{D_v}\boldsymbol{D_n}^{-1} - \boldsymbol{D_n}^{-1}\boldsymbol{D_v}\boldsymbol{X}_A(\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{X}_A)^{-1}\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{D_n}^{-1} \\ &= \boldsymbol{D_n}^{-1}\left[\boldsymbol{D_v} - \boldsymbol{D_v}\boldsymbol{X}_A(\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{X}_A)^{-1}\boldsymbol{X}'_A\boldsymbol{D_v}\right]\boldsymbol{D_n}^{-1}\end{aligned}$$

The observed minus expected residuals are $\left(\boldsymbol{y} - \boldsymbol{\mu}\left(\widehat{\boldsymbol{\beta}}_A\right)\right) = \boldsymbol{D_n}\left(\boldsymbol{p} - \widehat{\boldsymbol{\Pi}}\left(\widehat{\boldsymbol{\beta}}_A\right)\right)$.

Their covariance matrix is approximately

$$Var\left\{\left(\boldsymbol{y} - \boldsymbol{\mu}\left(\widehat{\boldsymbol{\beta}}_A\right)\right)\right\} \to \boldsymbol{D_v} - \boldsymbol{D_v}\boldsymbol{X}_A(\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{X}_A)^{-1}\boldsymbol{X}'_A\boldsymbol{D_v}$$

For linear functions $\boldsymbol{W}'\left(\boldsymbol{y} - \boldsymbol{\mu}\left(\widehat{\boldsymbol{\beta}}_A\right)\right)$, the covariance matrix is approximately

$$\boldsymbol{W}\left[\boldsymbol{D_v} - \boldsymbol{D_v}\boldsymbol{X}_A(\boldsymbol{X}'_A\boldsymbol{D_v}\boldsymbol{X}_A)^{-1}\boldsymbol{X}'_A\boldsymbol{D_v}\right]\boldsymbol{W}'$$