Fitting Regression models for binary responses

1. Likelihood equations
2. Asymptotic covariance matrix
3. Confidence intervals
4. Horseshoe crab example: revisited

Homework: Ch. 5, Sec. 5.1 – 3.
Fitting Logistic Regression Models

Notation: \( n = \# \) of subjects with binary response.

\[ \mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip}) \text{, for each } i = 1, 2, \ldots, N \text{ levels of explanatory variables} \]

Remark #1. For continuous explanatory variables, one can assume \( N = n \). For a discrete case, \( N \) is fixed (\( N = 2 \) corresponds to two groups).

Consider logistic regression model

\[
\Pi (\mathbf{x}_i) = \frac{\exp \left( \sum_{j=1}^{p} \beta_j x_{ij} \right)}{1 + \exp \left( \sum_{j=1}^{p} \beta_j x_{ij} \right)}
\]

\( Y_i = \# \) of '1's out of \( n_i \) units in level \( i \).

Assume \( Y_i \sim \text{Bin} (n_i, \Pi (\mathbf{x}_i)) \), \( i = 1, \ldots, N \)-fixed are independent and \( \sum_{i=1}^{N} n_i = n \).
\[ E(Y_i) = n_i \pi(x_i) \]

Likelihood function is proportional to

\[
\prod_{i=1}^{N} \pi(x_i)^{y_i} \left[1 - \pi(x_i)\right]^{n_i - y_i} = 
\]

\[
\prod_{i=1}^{N} \log \left( \frac{\pi(x_i)}{1 - \pi(x_i)} \right)^{y_i} \prod_{i=1}^{N} \left(1 - \pi(x_i)\right)^{n_i - y_i} = 
\]

\[
\exp \left[ \sum_{i=1}^{N} y_i \log \left( \frac{\pi(x_i)}{1 - \pi(x_i)} \right) \right] \prod_{i=1}^{N} \left(1 - \pi(x_i)\right)^{n_i - y_i}.
\]

But for our model

\[
\log \frac{\pi(x_i)}{1 - \pi(x_i)} = \sum_{j} \beta_j x_{ij}.
\]

\[
l - \pi(x_i) = \left[1 + \exp \left( \sum_{j} \beta_j x_{ij} \right) \right]^{-1}.
\]

So that the log likelihood function is

\[
L(\beta) = \sum_{i=1}^{N} y_i \sum_{j=1}^{p} \beta_j x_{ij} - 
\]

\[
- \sum_{i=1}^{N} n_i \log \left[1 + \exp \left( \sum_{j} \beta_j x_{ij} \right) \right] - 2.
\]
\[ L(\beta) = \sum_{j=1}^{p} \left( \sum_{i=1}^{N} y_i x_{ij} \right) \beta_j - \sum_{i=1}^{N} n_i \log \left[ 1 + \exp \left( \frac{\sum_{j=1}^{p} \beta_j x_{ij}}{\theta} \right) \right] \]

Sufficient statistics for \( \beta \) parameters are
\[ \sum_{i=1}^{N} y_i x_{ij}, \quad j = 1, 2, \ldots, p. \]

\( \beta_j \) can be constructed from equations:
\[ \frac{\partial L(\beta)}{\partial \beta_j} = 0, \quad j = 1, 2, \ldots, p. \]

\[ \sum_{i=1}^{N} y_i x_{ij} - \sum_{i=1}^{N} n_i x_{ij} \times \frac{\sum_{k=1}^{p} \beta_k x_{ik}}{1 + \exp \left( \theta \beta_k x_{ik} \right)} = 0 \]

for \( \beta_j \): \[ \sum_{i}^{N} y_i x_{ij} - \sum_{i}^{N} n_i x_{ij} \cdot \frac{1}{n_i} = 0 \]

\[ \frac{1}{n_i} = \exp \left( \frac{\sum_{k=1}^{p} \beta_k x_{ik}}{\theta} \right) / \left[ 1 + \exp \left( \frac{\sum_{k=1}^{p} \beta_k x_{ik}}{\theta} \right) \right] \]

is the MLE of \( \pi(x_i) \)
Let us denote the estimated expected frequencies by

\[ \hat{\mu}_j = n_j \frac{1}{\hat{\pi}_i} \quad i=1,2,\ldots,N \]

and

\[ X = (x_{ij})_{i=1}^{N} \quad N \times p \text{ matrix} \]

\[ \hat{\mu} = (\hat{\mu}_1, \ldots, \hat{\mu}_N) \]

\[ Y = (Y_1, \ldots, Y_N) \]

The likelihood equation can be rewritten as

\[ X'Y = X'\hat{\mu} \]

2. Asymptotic Covariance matrix of \( \hat{\beta} \)

\[ \text{cov}(\hat{\beta}) = \lim_{n \to \infty} X' \text{diag} \left[ n_i \hat{\pi}_i (1-\hat{\pi}_i) \right] X \]

\[ \text{cov}(\hat{\beta}) = L \]

\[ \text{diag} \left[ n_i \hat{\pi}_i (1-\hat{\pi}_i) \right] = \begin{pmatrix}
\frac{n_1 \hat{\pi}_1 (1-\hat{\pi}_1)}{n_1 \hat{\pi}_1 (1-\hat{\pi}_1)} & 0 \\
0 & \frac{n_2 \hat{\pi}_2 (1-\hat{\pi}_2)}{n_2 \hat{\pi}_2 (1-\hat{\pi}_2)} \\
\end{pmatrix} \]

is \( N \times N \) matrix

i.e.

\[ \hat{\pi}_i = n_i \frac{\hat{\pi}_i}{1-\hat{\pi}_i} \]

so that \( \hat{SE}(\hat{\beta}_i) = \) the square root of the corresponding element (on diagonal) of \( \text{cov}(\hat{\beta}) \) matrix.
CI for $\pi(x)$:

Since

$$\logit [\hat{\pi}(x)] = x' \hat{\beta},$$

the estimated variance of $\logit [\hat{\pi}(x)]$ is

$$x' \text{cov}(\hat{\beta}) x$$

so that when $n$ is large, $(1-\delta)100\%$

CI for $\logit \pi(x)$:

$$\logit [\hat{\pi}(x)] \pm z_{\alpha/2} \sqrt{x' \text{cov}(\hat{\beta}) x}$$

and

CI for $\pi(x)$ can be constructed using the transform

$$\pi = \exp(\logit) / [1 + \exp(\logit)]$$
Horseshoe Crab Example Revisited

Let us treat color as a qualitative variable with four categories:

Model: \[ \text{logit}[\pi] = \alpha + \beta_1 c_1 + \beta_2 c_2 + \beta_3 c_3 + \beta_4 x \]

with \( P(Y=1) = \pi \)

\( x = \) width in cm
\( c_1 = \begin{cases} 1 & \text{for medium-light color} \\ 0 & \text{otherwise} \end{cases} \)
\( c_2 = \begin{cases} 1 & \text{for medium color} \\ 0 & \text{otherwise} \end{cases} \)
\( c_3 = \begin{cases} 1 & \text{for medium-dark color} \\ 0 & \text{otherwise} \end{cases} \)

The crab color is dark (category 4) when \( c_1 = c_2 = c_3 = 0 \)

Fitted model:

\[ \text{logit}[\pi] = -12.715 + 1.3299 c_1 + 1.4023 c_2 + 1.1061 c_3 + .4680 x \]
For example, for dark crabs:

\[ \logit \hat{\pi} = -12.715 + 0.468 x \]

and for medium-light crabs (with \( c_1 = 1 \))

\[ \logit(\hat{\pi}) = -12.715 + 1.3299 + 0.468 x \]

\[ = -11.385 + 0.468 x \]

at the average width of 26.3 cm

for dark crabs:

\[ \hat{\pi} = \frac{e^{(-12.715 + 0.468 \times 26.3)}}{1 + e^{(-12.715 + 0.468 \times 26.3)}} = 0.399 \]

and for medium-light crabs

\[ \hat{\pi} = 0.715 \]

Remark: this model assumes a lack of interaction between color and width in their effects.
Model comparison

Let us consider two models:

\( M_0: \logit(\pi) = \alpha + \beta_4 x \)

i.e. the probability of a satellite is independent of color. 

(Simple model) and consider the full model

\( M_1: \logit(\pi) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x \)

Corresponding maximized log-likelihoods are \( L_0 \) and \( L_1 \) respectively.

Test Statistic: \( G^2(M_0 | M_1) = -2[L_0 - L_1] \)

Null Distribution

\( G^2(M_0 | M_1) \sim \chi^2 \) with d.f. = \( \dim(M_1) - \dim(M_0) = 5 - 2 = 3 \)

\( G^2(M_0 | M_1) = 7.0 \)

\( \chi^2_{3} \)

p-value = .07

Slight evidence of color effect.
Let us consider the quantitative (simple) model

\[ M_0 : \text{logit}(p) = \alpha + \beta_1 c + \beta_2 x \]

scores for color: \( c = 1, 2, 3, 4 \)

and compare it with \( M_1 \) model.

\[ G^2 (M_0 | M_1) = -2 \left[ L_0 - L_1 \right] = 1.7 \]

\[ G^2 \sim \chi^2 \text{ with d.f. } = 5 - 3 = 2 \]

![](image)

\[ \chi^2 \text{ d.f.} = 2 \]

\[ p \text{-value} = .44 \]

i.e. the Model \( M_0 \) is adequate given that \( M_1 \) holds.

Note that with the scores \( c = 1, 2, 3, 4 \)

the fitted model is

\[ \text{logit}(\hat{p}) = \hat{\alpha} + \hat{\beta}_1 c + \hat{\beta}_2 x \]

with \( \hat{\beta}_1 = -0.509 \), \( SE = 0.224 \)

\( \hat{\beta}_2 = 0.458 \), \( SE (\hat{\beta}_2) = 0.04 \)