STAT475/675 TUT05

http://www.sfu.ca/~zza115/teaching.html
http://people.stat.sfu.ca/~joanh/stat475-675web.html

Zhiyang Zhou (zhiyang_zhou@sfu.ca)

2018-02-06

Multiple logistic regression

- Binary response Y (e.g. success (1)/failure (0))
- Several explanatory variables $X_1, \ldots, X_K$
- $Y|X_1 = x_1, \ldots, X_K = x_K \sim \text{Bernoulli}(\pi(x_1, \ldots, x_K))$
- Model

$$\text{logit}(\pi(x_1, \ldots, x_K)) = \ln \frac{\pi(x_1, \ldots, x_K)}{1 - \pi(x_1, \ldots, x_K)} = \alpha + \beta_1 x_1 + \cdots + \beta_K x_K$$

  or, equivalently,

$$\pi(x_1, \ldots, x_K) = \frac{\exp(\alpha + \beta_1 x_1 + \cdots + \beta_K x_K)}{1 + \exp(\alpha + \beta_1 x_1 + \cdots + \beta_K x_K)} \in (0, 1)$$

  - $\beta_i = \text{logit}(\pi(x_1, \ldots, x_i = 1, \ldots, x_K)) - \text{logit}(\pi(x_1, \ldots, x_i = 0, \ldots, x_K))$, the effect of X_1 on the logarithm of odds of $Y = 1$, if X_i is binary, controlling the other explanatory variables

Inference for $\alpha, \beta_1, \ldots, \beta_K$

- MLE

$$(\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_K)^{\text{T}} \sim N((\alpha, \beta_1, \ldots, \beta_K)^{\text{T}}, \Sigma)$$

  with $\Sigma = (nI)^{-1}$ and Fisher information matrix I

- Approximate 95% joint CI

$$\left\{(\alpha, \beta_1, \ldots, \beta_K)^{\text{T}} : (\alpha - \hat{\alpha}, \beta_1 - \hat{\beta}_1, \ldots, \beta_K - \hat{\beta}_K)\Sigma^{-1}(\alpha - \hat{\alpha}, \beta_1 - \hat{\beta}_1, \ldots, \beta_K - \hat{\beta}_K)^{\text{T}} \le \chi^2_{.95}(K+1)\right\}$$

  - for any subset of $\{\alpha, \beta_1, \ldots, \beta_K\}$, the joint CI can be obtained by subsituting the corresponding submatrix of Σ for the entire Σ

- Hypothesis Testing, e.g, on $H_0 : \beta_1 = \beta_2 = 0$ vs $H_1 : \beta_1 = 0$ or $\beta_2 = 0$
  - Wald

$$\chi^2_{\text{obs}} = (\hat{\beta}_1, \hat{\beta}_2)\Sigma^{-1}(\hat{\beta}_1, \hat{\beta}_2)^{\text{T}} \sim \chi^2(2)$$

    approximately under H_0 when $n \gg 1$

  - lrt

$$G^2_{\text{obs}} = -2 \ln \frac{\max L(\alpha, 0, 0, \beta_3, \ldots, \beta_K)}{\max L(\alpha, \beta_1, \beta_2, \beta_3, \ldots, \beta_K)} \sim \chi^2(2)$$

    approximately under H_0 when $n \gg 1$

Inference for $\pi(x_1, \ldots, x_K)$

- MLE

$$\hat{\pi}(x_1, \ldots, x_K) = \text{logit}^{-1}(\hat{\alpha} + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_K x_K) = \frac{\exp(\hat{\alpha} + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_K x_K)}{1 + \exp(\hat{\alpha} + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_K x_K)}$$

- Approximate 95% CI for $\text{logit}(\pi(x_1, \ldots, x_K))$

$$\hat{\alpha} + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_K x_K \pm 1.96\sqrt{(1, x_1, \ldots, x_K)\Sigma(1, x_1, \ldots, x_K)^{\mathrm{T}}}$$

- Approximate 95% CI for $\pi(x_1, \ldots, x_K)$

$$\text{logit}^{-1}\left(\hat{\alpha} + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_K x_K \pm 1.96\sqrt{(1, x_1, \ldots, x_K)\Sigma(1, x_1, \ldots, x_K)^{\mathrm{T}}}\right)$$

More about predictors

- Transformation of predictors
  - Interaction
    - e.g., $x_1 x_2$
  - General function
    - e.g., $h(x_1)$
- Predictor with multiple levels
  - Qualitative: split it into multiple binary predictors
  - Quantitative: take it as a single ordinal predictor

Demo

The data included in “Arthritis” in package “vcd” is from Koch & Edwards (1988) after a double-blind clinical trial investigating a new treatment for rheumatoid arthritis.
