Multinomial regression

Motivating example: earthquake data

We have data from the 2015 Gorkha earthquake in Nepal. After the earthquake, a large scale survey was conducted to determine the amount of damage the earthquake caused for homes, businesses and other structures. Variables include:

- `Damage`: the amount of damage suffered by the building (none, moderate, severe)
- `age`: the age of the building (in years)
- `condition`: a de-identified variable recording the condition of the land surrounding the building

$$Damage_i \sim Categorical(\pi_{i(None)}, \pi_{i(Moderate)}, \pi_{i(Severe)})$$

$$\log\left(\frac{\pi_{i(Moderate)}}{\pi_{i(None)}}\right) = \beta_{(Moderate)}^T X_i$$

$$\log\left(\frac{\pi_{i(Severe)}}{\pi_{i(None)}}\right) = \beta_{(Severe)}^T X_i$$

Fisher scoring

$$\mu_i = \frac{\partial h(\theta_i)}{\partial \theta_i} \qquad \frac{\partial \mu_i}{\partial \theta_i} = V(\mu_i)$$

$$\boxed{V(\mu) = \begin{bmatrix} \pi_1(1-\pi_1) & -\pi_1\pi_2 & \cdots \\ -\pi_2\pi_1 & \pi_2(1-\pi_2) & \\ & & \ddots \end{bmatrix}}$$

$$u(\beta) = \frac{1}{\phi}\sum_{i=1}^{n} X_i^{*T}(Y_i^* - \mu_i)$$

(multivariate EDM w/ canonical link)

multinomial : $\phi = 1$, $Y_{ij}^* = \begin{cases} 1 & Y_i = j \\ 0 & Y_i \neq j \end{cases}$, $\mu_i = \begin{pmatrix} \pi_{i1} \\ \pi_{i2} \\ \vdots \\ \pi_{i,J-1} \end{pmatrix}$

$$\frac{\partial u(\beta)}{\partial \beta} = -\frac{1}{\phi}\sum_{i=1}^{n} \frac{\partial}{\partial \beta}(X_i^{*T}\mu_i)$$

$$\frac{\partial}{\partial \beta}(X_i^{*T}\mu_i) = \underbrace{\frac{\partial \mu_i}{\partial \beta}}_{\underbrace{\frac{\partial \theta_i}{\partial \beta}}_{X_i^{*T}} \underbrace{\frac{\partial \mu_i}{\partial \theta_i}}_{V(\mu_i)}} \underbrace{\frac{\partial X_i^{*T}\mu_i}{\partial \mu_i}}_{X_i^*}$$

(chain rule)

(canonical link)

$\theta_i = X_i^* \beta$

$$\Rightarrow \frac{\partial u(\beta)}{\partial \beta} = -\frac{1}{\phi}\sum_{i=1}^{n} X_i^{*T} V(\mu_i) X_i^* \quad \Rightarrow \mathcal{I}(\beta) = \frac{1}{\phi}\sum_{i=1}^{n} X_i^{*T} V(\mu_i) X_i^*$$

Wald tests

$H_0: \beta_{1(moderate)} = 0$

$H_A: \beta_{1(moderate)} \neq 0$

$z = \frac{0.375}{0.017} \approx 22.1$

p-value ≈ 0

```
...
## Coefficients:
##          (Intercept) sqrt(age)  conditiono conditiont
## moderate   0.6581163 0.3747641 -0.45376940 -0.5803708
## severe     0.1881145 0.4251732  0.04706934 -0.4623774
##
## Std. Errors:
##          (Intercept)  sqrt(age) conditiono conditiont
## moderate   0.1208913 0.01684468  0.2305975  0.1155475
## severe     0.1243799 0.01725782  0.2292533  0.1180182
...
```

> Suppose we want to know whether there is a relationship between age and the odds of moderate vs. no damage, after accounting for surface condition. What hypotheses would we test?

$vcov(model)$

$H_0: \beta_{1(moderate)} = \beta_{1(severe)}$

$H_A: \beta_{1(moderate)} \neq \beta_{1(severe)}$

$$\beta = \begin{pmatrix} \beta_{0(moderate)} \\ \vdots \\ \beta_{3(moderate)} \\ \beta_{0(severe)} \\ \vdots \\ \beta_{3(severe)} \end{pmatrix}$$

Wald tests

$a^T = (0, -1, 0, 0, 0, 1, 0, 0)$

$$Z = \frac{a^T \hat{\beta} - 0}{\sqrt{a^T \mathcal{I}^{-1}(\hat{\beta}) a}}$$

```
...
## Coefficients:
##          (Intercept) sqrt(age)  conditiono conditiont
## moderate   0.6581163 0.3747641 -0.45376940 -0.5803708
## severe     0.1881145 0.4251732  0.04706934 -0.4623774
##
## Std. Errors:
##          (Intercept)  sqrt(age) conditiono conditiont
## moderate   0.1208913 0.01684468  0.2305975  0.1155475
## severe     0.1243799 0.01725782  0.2292533  0.1180182
...
```

$var(a^T \hat{\beta}) = a^T var(\hat{\beta}) a$

$var(Y_1 - Y_2) = var(Y_1) + var(Y_2) - 2 cov(Y_1, Y_2)$

> Suppose we want to know whether the relationship between age and the odds of moderate vs. no damage is the *same* as the relationship between age and the odds of severe vs. no damage. What hypotheses would we test?

Wald tests

```
diff <- t(c(0, -1, 0, 0, 0, 1, 0, 0)) %*%
  c(t(coef(m1)))
std_err <- sqrt(t(c(0, -1, 0, 0, 0, 1, 0, 0)) %*%
                  vcov(m1) %*%
                  c(0, -1, 0, 0, 0, 1, 0, 0))
(diff - 0)/std_err
```

```
##          [,1]
## [1,] 4.95677
```

```
2*pnorm((diff - 0)/std_err, lower.tail = F)
```

```
##              [,1]
## [1,] 7.167478e-07
```

Likelihood ratio tests

$H_0: \beta_{2(moderate)} = \beta_{2(severe)} = \beta_{3(m)} = \beta_{3(s)} = 0$

H_A: at least one of $\beta_{2(m)}, \beta_{2(s)}, \beta_{3(m)}, \beta_{3(s)} \neq 0$

G = reduced deviance - full deviance $\sim \chi^2$ df = 4

```
...
## Coefficients:
##          (Intercept) sqrt(age)  conditiono conditiont
## moderate   0.6581163 0.3747641 -0.45376940 -0.5803708
## severe     0.1881145 0.4251732  0.04706934 -0.4623774
##
## Std. Errors:
##          (Intercept)  sqrt(age) conditiono conditiont
## moderate   0.1208913 0.01684468  0.2305975  0.1155475
## severe     0.1243799 0.01725782  0.2292533  0.1180182
...
```

Suppose we want to know whether there is a relationship between surface condition and damage, after accounting for building age. What hypotheses would we test?

Likelihood ratio tests

```
m1 <- multinom(Damage ~ sqrt(age) + condition,
               data = earthquake)
m2 <- multinom(Damage ~ sqrt(age),
               data = earthquake)

pchisq(m2$deviance - m1$deviance, df = 4,
       lower.tail=F)
```

```
## [1] 2.452814e-08
```

Deviance for multivariate EDM

$$f(y;\theta,\phi) = a(y,\phi)\exp\left\{\frac{y^T\theta - \kappa(\theta)}{\phi}\right\}$$

Dispersion model form: $f(y;\theta,\phi) = b(y,\phi)\exp\left\{-\frac{d(y,\mu)}{2\phi}\right\}$

$$t(y,\mu) = y^T\theta - \kappa(\theta) \qquad (\theta = g(\mu))$$

$$d(y,\mu) = 2(t(y,y) - t(y,\mu))$$

Multinomial regression:

$$\theta = \left(\log\left(\frac{\pi_1}{1-\sum_{j=1}^{J-1}\pi_j}\right), \ldots, \log\left(\frac{\pi_{J-1}}{1-\sum_{j=1}^{J-1}\pi_j}\right)\right)^T$$

$$\kappa(\theta) = -\log\left(1 - \sum_{j=1}^{J-1}\pi_j\right)$$

$t(y,y)$: $\pi_j = y_j^*$ $\Rightarrow$ $\sum_{j=1}^{J-1} y_j^* \log\left(\frac{y_j^*}{1-\sum_{k=1}^{J-1} y_k^*}\right) + \log\left(1 - \sum_{k=1}^{J-1} y_k^*\right)$

$$= \sum_{j=1}^{J} y_j^* \log(y_j^*)$$

$t(y,\mu) = \sum_{j=1}^{J} y_j^* \log(\pi_j)$ $\Rightarrow$ $d(y,\mu) = 2\sum_{j=1}^{J} y_j^* \log\left(\frac{y_j^*}{\pi_j}\right)$

$$y_j^* = \begin{cases} 1 & y = j \\ 0 & y \neq j \end{cases}$$

Class activity

https://sta712-f22.github.io/class_activities/ca_lecture_36.html