Lecture 29

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

Fisher scoring

vector vector scalar

Recap: multivariate EDM $f(y^*; \theta, \phi) = a(y, \phi) \exp\left\{ \frac{y^{*T}\theta - k(\theta)}{\phi} \right\}$

$\Rightarrow \log f(y^*; \theta, \phi) = \log a(y, \phi) + \frac{1}{\phi}(y^{*T}\theta - k(\theta))$

goes away when we find MLEs for θ

Multivariate GLM: $g(\mu_i) = x_i^{*} \beta$

$$\begin{bmatrix} x_i^T & & 0 \\ & x_i^T & \\ 0 & & \ddots \; x_i^T \end{bmatrix}$$

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

$$\tilde{I}(\beta) = \frac{1}{\phi} \sum_{i=1}^{n} x_i^{*T} V(\mu_i) x_i^*$$

Univariate EDM:

$$u(\beta) = \frac{1}{\phi} \sum_{i=1}^{n} x_i^T (y_i - \mu_i)$$

$$\tilde{I}(\beta) = \frac{1}{\phi} \sum_{i=1}^{n} V(\mu_i) x_i x_i^T$$

Wald tests

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

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

$\hat{\beta}_{1(\text{moderate})} = 0.375$

	(Intercept)	sqrt(age)	conditiono	conditiont
moderate	0.6581163	0.3747641	-0.45376940	-0.5803708
severe	0.1881145	0.4251732	0.04706934	-0.4623774

coefficients

	(Intercept)	sqrt(age)	conditiono	conditiont
moderate	0.1208913	0.01684468	0.2305975	0.1155475
severe	0.1243799	0.01725782	0.2292533	0.1180182

Standard errors

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?

$$Z = \frac{0.375 - 0}{0.017} \approx 22.1 \qquad p \approx 0$$

Wald tests

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

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

```
          (Intercept) sqrt(age)  conditiono conditiont
moderate    0.6581163 0.3747641 -0.45376940 -0.5803708
severe      0.1881145 0.4251732  0.04706934 -0.4623774

          (Intercept)  sqrt(age) conditiono conditiont
moderate    0.1208913 0.01684468  0.2305975  0.1155475
severe      0.1243799 0.01725782  0.2292533  0.1180182
```

$$a = \begin{pmatrix} 0 \\ -1 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$$

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?

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_{0(M)} \\ \hat{\beta}_{1(M)} \\ \vdots \\ \hat{\beta}_{0(S)} \\ \hat{\beta}_{1(S)} \\ \vdots \end{pmatrix}$$

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

vs. H_A: " " $\neq 0$

$a^T\hat{\beta}$

$\text{Var}(a^T\hat{\beta})$

$= a^T \underbrace{\text{Var}(\hat{\beta})}_{\hat{I}^{-1}(\beta)} a$

$$\text{var}(\hat{\beta}_{1(S)} - \hat{\beta}_{1(M)}) = \text{Var}(\hat{\beta}_{1(S)}) + \text{Var}(\hat{\beta}_{1(M)}) - 2\,\text{Cov}(\hat{\beta}_{1(S)}, \hat{\beta}_{1(M)})$$

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
```

Var($\hat{\beta}$)

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

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

Z — Statistic

Likelihood ratio tests

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

H_A: at least one of them is $\neq 0$

```
          (Intercept) sqrt(age)  conditiono conditiont
moderate    0.6581163 0.3747641 -0.45376940 -0.5803708
severe      0.1881145 0.4251732  0.04706934 -0.4623774

          (Intercept)  sqrt(age) conditiono conditiont
moderate    0.1208913 0.01684468  0.2305975  0.1155475
severe      0.1243799 0.01725782  0.2292533  0.1180182
```

full model

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?

reduced model: removing condition

$$G = 2(\log L_{full} - \log L_{red})$$

$$= \text{reduced deviance} - \text{full deviance}$$

$$\sim \chi^2_4 \quad (\text{under } H_0)$$

Likelihood ratio tests

```r
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
```

full

reduced

Class activity

https://sta712-f23.github.io/class_activities/ca_lecture_29.html