Discrete Response Model
Lecture 2

datascience@berkeley

Variance-Covariance Matrix of the Estimators

Variance-Covariance Matrix of the Estimators

The estimated variance-covariance matrix for $\hat{\beta}_0, \ldots, \hat{\beta}_p$ has the usual form:

$$\begin{bmatrix} \mathrm{Var}(\hat{\beta}_0) & \mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1) & \mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_p) \\ \mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_1) & \mathrm{Var}(\hat{\beta}_1) & \mathrm{Cov}(\hat{\beta}_1,\hat{\beta}_p) \\ & & \\ \mathrm{Cov}(\hat{\beta}_0,\hat{\beta}_p) & \mathrm{Cov}(\hat{\beta}_1,\hat{\beta}_p) & \mathrm{Var}(\hat{\beta}_p) \end{bmatrix}$$

This matrix is found through using the same likelihood-based methods discussed in Lecture 1.

Variance-Covariance Matrix and MLE

If $\hat{\theta}$ is the MLE for θ, we can say that

$$\hat{\theta} \sim N\left(\theta, Var(\hat{\theta})\right)$$

for a large sample, where

$$Var(\hat{\theta}) = -\left[E\left(\frac{\partial^2 \log[L(\theta \mid Y_1, \ldots, Y_n)]}{\partial\theta^2}\right)\right]^{-1}\Bigg|_{\theta=\hat{\theta}}$$

Hessian Matrix

Note that we need to account for not only 1 MLE, but for p + 1 MLEs.

The result is a matrix of second partial derivatives (a Hessian matrix) of the log-likelihood function evaluated at the parameter estimates, and multiplying this resulting matrix by -1.

For example, if there are only parameters β_0 and β_1, the matrix is

$$-E\left(\begin{bmatrix} \dfrac{\partial^2 \log[L(\beta_0,\beta_1 \mid y_1,\ldots,y_n)]}{\partial\beta_0^2} & \dfrac{\partial^2 \log[L(\beta_0,\beta_1 \mid y_1,\ldots,y_n)]}{\partial\beta_0\partial\beta_1} \\ \dfrac{\partial^2 \log[L(\beta_0,\beta_1 \mid y_1,\ldots,y_n)]}{\partial\beta_0\partial\beta_1} & \dfrac{\partial^2 \log[L(\beta_0,\beta_1 \mid y_1,\ldots,y_n)]}{\partial\beta_1^2} \end{bmatrix}\right)^{-1}\Bigg|_{\beta_0=\hat{\beta}_0,\beta_1=\hat{\beta}_1}$$

The Sandwich Estimators

In the end, the matrix in general ends up having the nice sandwich form of $(\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}$ where

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & & x_{1p} \\ 1 & x_{21} & x_{22} & & x_{2p} \\ & & & & \\ 1 & x_{n1} & x_{n2} & & x_{np} \end{bmatrix}$$

and $\mathbf{V} = \mathrm{Diag}\left(\hat{\pi}_i(1-\hat{\pi}_i)\right)$ is a n×n matrix with $\hat{\pi}_i(1-\hat{\pi}_i)$ on the diagonal and 0's elsewhere.

Variance-Covariance Matrix in R

After estimating a model, the vcov() function can be used to produce the estimated variance-covariance matrix associated with the MLE:

```
> vcov(mod.fit2)
             (Intercept)        change      distance
(Intercept)  0.111011379 -0.0094878323 -2.625598e-03
change      -0.009487832  0.0375091687 -1.311512e-04
distance    -0.002625598 -0.0001311512  7.129494e-05
```

Berkeley SCHOOL OF INFORMATION