STAT 3690 Lecture 19

zhiyanggeezhou.github.io

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca)

Mar 14, 2022

Multivariate influence measures

- Hat/projection matrix $\mathbf{H} = [h_{ij}]_{n\times n} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$
  - $|h_{ij}| \le 1$
- $\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$
  - the ith row of $\hat{\mathbf{Y}}$: $\hat{\mathbf{Y}}_{i\cdot} = \sum_{j=1} h_{ij}\mathbf{Y}_{j\cdot} = h_{ii}\mathbf{Y}_i + \sum_{j\neq i} h_{ij}\mathbf{Y}_{j\cdot}$
- Leverage: the influence of observation $\mathbf{Y}_{i\cdot}$ on $\hat{\mathbf{Y}}_{i\cdot}$
  - Observation $\mathbf{Y}_{i\cdot}$ is said to have a high leverage if h_{ii} is large compared to the other elements on the diagonal of $\mathbf{H}$.
- (Externally) Studentized residuals

$$T_i^2 = \frac{\hat{\mathbf{E}}_{i\cdot}^\top \mathbf{\Sigma}_{\mathrm{LS},(i)}^{-1}\hat{\mathbf{E}}_{i\cdot}}{1-h_{ii}}$$

  - $\hat{\mathbf{E}}_{i\cdot}^\top$: the ith row of $\hat{\mathbf{E}} = (\mathbf{I}-\mathbf{H})\mathbf{Y}$
  - $\hat{\mathbf{E}}_{(i)\cdot}^\top$: remaining part of $\hat{\mathbf{E}}$ with row i removed
  - $\mathbf{\Sigma}_{\mathrm{LS},(i)} = (n-q-2)^{-1}\hat{\mathbf{E}}_{(i)\cdot}^\top\hat{\mathbf{E}}_{(i)\cdot}$: LS estimator of $\mathbf{\Sigma}$ where we have removed row i from the residual matrix
  - Observation $\mathbf{Y}_i$ may be considered as a potential outlier if

$$T_i^2 > \frac{p(n-q-2)}{n-p-q-1}F_{1-\alpha,p,n-q-2}$$

    * $F_{1-\alpha,p,n-q-2}$: the $1-\alpha$ quantile of $F(p, n-q-2)$
- (Multivariate) Cook's distance

$$C_i = \frac{h_{ii}}{(1-h_{ii})^2(q+1)}\hat{\mathbf{E}}_{i\cdot}^\top\mathbf{\Sigma}_{\mathrm{LS}}^{-1}\hat{\mathbf{E}}_{i\cdot}$$

  - Observation $\mathbf{Y}_i$ may be considered as a potential outlier if

$$C_i > \text{median of } \chi^2(p(n-q-1))$$

Normality of residuals

- Apply techniques in Lecture 7 to checking the normality of residuals
- Apply Box-Cox transformation to colums of $\mathbf{Y}$
