3 December 2020

Transformation

Transformation on x

We simply use the natural logarithm transformation of X_j if the ratio of the largest observed value of X_j to the smallest observed of X_j is greater than 10.

Outlier & Influential observation

Outlier

The i^{th} observation is an outlier if $|t_i| > |t_{\frac{\alpha}{2n}}|$

Influential observation

Name in SAS output	Expression	Cutoff point
Student Residual	$r_i = \dfrac{\hat{e}_i}{\hat{\sigma}\sqrt{1-h_{ii}}}$	$\lvert r_i \rvert > 2$
Rstudent	$t_i = \dfrac{\hat{e}_i}{\hat{\sigma}_{-i}\sqrt{1-h_{ii}}}$	$\lvert t_i \rvert > t_{\alpha/(2n)}$
Hat Diag H	$h_{ii} = \boldsymbol{x_i}^T(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{x_i}$	$h_{ii} > 2p'/n$
Cook's D	$D_i = \left(\dfrac{t_i^2}{p'}\right)\left(\dfrac{h_{ii}}{1-h_{ii}}\right)$	$D_i >> 1$
Dffits	$(\text{DFFITS})_i = \dfrac{\hat{y}_i - \hat{y}_{i,-i}}{\hat{\sigma}_{-i}\sqrt{h_{ii}}} = (\text{Rstudent})_i \left(\dfrac{h_{ii}}{1-h_{ii}}\right)^{1/2}$	$> 2\sqrt{p'/n}$
$(\text{DFBETAS})_{j,i}$	$(\text{DFBETAS})_{j,i} = \dfrac{\hat{\beta}_j - \hat{\beta}_{j,-i}}{\hat{\sigma}_{-i}\sqrt{c_{jj}}} = \dfrac{r_{j,i}}{\sqrt{\boldsymbol{r}_j'\boldsymbol{r}_j}} \dfrac{(R-\text{student})_i}{\sqrt{1-h_{ii}}}$	$> 2/\sqrt{n}$
Cov Ratio	$(\text{COVRATIO})_i = \dfrac{(\hat{\sigma}_{-i})^{2p'}}{\hat{\sigma}^{2p'}}\left(\dfrac{1}{1-h_{ii}}\right)$	$> 1 + 3p'/n$ or $< 1 - 3p'/n$

Remarks

1. If the i^{th} observation is suspected to be an influential point, fit models of y on $x_1, \ldots, x_p$ with and without i^{th} observations. If the estimates of β_j for $j = 0, 1, \ldots, p$ from these two models are different significantly, then the i^{th} observation is an influential point.

2. If the i^{th} observation is an outlier and a high leverage point, it will be an influential point.

Multicollinearity

1. If VIF (variance inflation factor) for the j^{th} independent variable is (extremely) large, delete this variable.