Assignment2 DATA501

Adam g. 300619334

August 2024

Packages

```
library(knitr)
library(pander)
library(dplyr)
library(ggplot2)
library(basetheme)
```

Background

Data

Cook's distance is defined as

$$C_i = \frac{\sum_{j=1}^{n}(\hat{y}_j - \hat{y}_{j(i)})^2}{\hat{\sigma}^2(p+1)}$$

Leverage point: (Welsch and Kuh 1977)

$$h_i > \frac{2p}{n}$$

End of Assignment 2, DATA501

References

Chatterjee, S., and A. Hadi. 2012. "Regression Analysis by Example. 5th Edition." In. https://www.semanticscholar.org/paper/Regression-Analysis-by-Example.-5th-Edition.-Chatterjee-Hadi/0f507486adfb0124065ceae0893fcf9c53e7d127.

Cook, R. Dennis. 1977. "Detection of Influential Observation in Linear Regression." *Technometrics* 19 (1): 15–18. https://doi.org/10.2307/1268249.

Welsch, Roy, and Edwin Kuh. 1977. "Linear Regression Diagnostics." w0173. Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w0173.