Journal of Statistical Computation and Simulation ›
Volume 88, 2018 - Issue 12

Original Articles

Variance Inflation Factor and Condition Number in multiple linear regression

R. Salmerón, C. B. García & J. García

Pages 2365-2384 | Received 19 Apr 2017, Accepted 07 Apr 2018, Published online: 22 Apr 2018

Cite this article https://doi.org/10.1080/00949655.2018.1463376 Check for updates

Full Article Figures & data References Citations Metrics
Reprints & Permissions Read this article Share

ABSTRACT

The Variance Inflation Factor and the Condition Number are measures traditionally applied to detect the presence of collinearity in a multiple linear model. This paper presents the relation and the difference between both measures from theoretical and empirical perspectives by using Monte Carlo simulations and taking special interest in the computational techniques.

KEYWORDS:

Variance Inflation Factor | Condition Number | multicollinearity detection | data transformation

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

R. Salmerón http://orcid.org/0000-0003-2589-4058

C. B. García http://orcid.org/0000-0003-1622-3877

Notes

1 Note that the constant term disappears after the standardization of the data.

2 Note that, when data are standardized, the VIF and CN coincide with the result obtained from typified data.

3 Note that these examples are not regression models since n=p.

4 Denoting X1=1, the auxiliary regression to calculate the VIF is expressed as X2=γ1+w, where it is verified that γˆ=X¯2 and, consequently, SSR=∑i=1n(X2i−X¯2)2=SST. In this case, it is always verified that Raux2=1. The version of the previous regression with unit length data is given by X2,lu=γ1lu+w where X2,lu=X/a with a=∑i=1nX2i2 and 1lu=1/n. In this case, γˆ=n/a·X¯2 and, then, SSR=(1/a)∑i=1n(X2i−n·X¯2·1/n)2=SST. Thus, this situation will be similar to the initial one.