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

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 $X_1=1$, the auxiliary regression to calculate the VIF is expressed as $X_2=\gamma 1+w$, where it is verified that $\hat{\gamma}=\bar{X}_2$ and, consequently, $SSR=\sum_{i=1}^{n}(X_{2i}-\bar{X}_2)^2=SST$. In this case, it is always verified that $R^2_{aux}=1$. The version of the previous regression with unit length data is given by $X_{2,lu}=\gamma 1_{lu}+w$ where $X_{2,lu}=X/a$ with $a=\sum_{i=1}^{n}X_{2i}^2$ and $1_{lu}=1/n$. In this case, $\hat{\gamma}=n/a\cdot\bar{X}_2$ and, then, $SSR=(1/a)\sum_{i=1}^{n}(X_{2i}-n\cdot\bar{X}_2\cdot 1/n)^2=SST$. Thus, this situation will be similar to the initial one.