

Maciej Nasiński

Wiarygodność modeli regresji w zależności od liczebności próbki na przykładzie modelu wynagrodzeń

Przeprowadzono analizę badającą jak zwiększenie liczebności próbki wpływa na precyzję oszacowań jednorównaniowego modelu regresji liniowej. Wspomniany model to rozszerzona wersja słynnego wnioskowania Jacoba Mincera. W badaniu posłużono się symulacjami komputerowymi oraz zaawansowanymi algorytmami algebry liniowej w celu wielokrotnej estymacji błędu oszacowań, przy różnej liczebności próbek. Symulacje przeprowadzono zakładając różne cele modelowania, w tym opisowy oraz predykcyjny. Ze względu na dużą ilość obserwacji oraz symulacje komputerowe, konieczne było zastosowanie rozwiązań biblioteki algebry liniowej, Armadillo udostępniającej rozwiązania bibliotek numerycznych takich jak LAPACK, BLAS czy MKL.

Symulacje potwierdzają malejące korzyści, płynące ze zwiększania ilości obserwacji w próbce, zarówno w badaniach opisowych jak i predykcyjnych. Wyniki wskazują także na istotność próbkowania warstwowego, które w przypadku istotnych zmiennych objaśniających pozwala na poprawę precyzji oszacowań.

Słowa kluczowe: Regresja Liniowa, Symulacje, Próbkowanie, Armadillo

Bibliografia

Eddelbuettel D., Sanderson C., *RcppArmadillo: Accelerating R with high-performance C++ linear algebra*, Computational Statistics & Data Analysis Volume 71, 2014

Polachek S., *Earnings Over the Lifecycle: The Mincer Earnings Function and Its Applications*, IZA Discussion Paper No. 3181, 2007

Varian, H., *Big data: New tricks for econometrics*, Journal of Economic Perspectives 28(2), 2014

Tillé Y., *Sampling Algorithms*, Springer Science & Business Media, 2006