

Imputasi ganda dengan metode predictive mean matching pada data numerik = Multiple imputation with predictive mean matching method on numeric data

Emha Fathul Akmam, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20494894&lokasi=lokal>

Abstrak

Nilai yang hilang adalah suatu kondisi ketika ada beberapa entri yang hilang atau nilai kosong pada beberapa pengamatan dalam data. Ini dapat menyebabkan kita memiliki data yang tidak lengkap yang dapat menghambat proses analisis statistik dan dapat memberikan kesimpulan yang bias dari analisis jika tidak dapat ditangani dengan benar. Masalah ini juga dapat ditemukan dalam beberapa analisis regresi linier. Salah satu cara untuk mengatasi masalah ini adalah dengan menggunakan beberapa metode imputasi yang disebut Predictive Mean Matching (PMM). Beberapa imputasi adalah pendekatan umum untuk menangani data yang hilang yang memungkinkan ketidakpastian tentang data yang hilang dengan membuat serangkaian data imputasi yang masuk akal dan hasil analisis gabungan tepat yang diperoleh dari setiap data. Metode ini akan mencocokkan jarak rata-rata prediktif dari pengamatan tidak lengkap dengan pengamatan lengkap. Dengan demikian, pengamatan lengkap yang memiliki jarak terdekat akan menjadi nilai donor bagi yang tidak lengkap. Mean prediktif untuk pengamatan lengkap diperkirakan dengan kuadrat biasa yang paling tidak lengkap dan didekati oleh konsep Bayesian. Artikel ini akan menjelaskan tentang distribusi posterior yang memungkinkan ketidakpastian nilai yang hilang untuk memperkirakan parameter pengamatan yang tidak lengkap dan mendapatkan rata-rata prediktif mereka. Artikel ini akan memberikan hasil tentang penerapan metode PMM dalam data simulasi yang memiliki pola hilang univariat dengan mekanisme MAR dan dievaluasi dengan nilai bias dari koefisien regresi dari koefisien hasil pada dataset yang ditentukan.

<hr>

Missing value is a condition when there are some missing entries or empty values on some observations in the data. This can cause us to have incomplete data that can hinder the process of statistical analysis and can provide biased conclusions from the analysis if it cannot be handled properly. This problem can also be found in several linear regression analyzes. One way to overcome this problem is to use several imputation methods called Predictive Mean Matching (PMM). Some imputations are a general approach to handling missing data that allows uncertainty about missing data by making a series of reasonable imputation data and the precise combined analysis results obtained from each data. This method will match the predictive average distance from incomplete observations with complete observations. Thus, a complete observation that has the closest distance will be a donor value for the incomplete. The predictive means for complete observations are estimated with the most incomplete ordinary squares and are approached by the Bayesian concept. This article will explain posterior distributions that allow uncertainty of missing values to estimate incomplete observation parameters and obtain their predictive averages. This article will provide results about the application of the PMM method in simulation data which have a univariate missing pattern with the MAR mechanism and are evaluated with a bias value of the regression coefficient of the yield coefficient on the specified dataset.