

## RINGKASAN

### Perbandingan Metode Imputasi untuk Mengestimasi Data Hilang pada Data

#### Kesehatan Ibu dan Anak di Jawa Timur

Yudi Evrlyanto

Data yang baik adalah data yang dikumpulkan tepat waktu, lengkap dan benar. Tetapi untuk mendapatkan data yang baik tidaklah mudah, bahkan sering terjadi data yang dikumpulkan, baik data survei maupun rutin, adanya beberapa data hilang (*missing data*) atau data tidak lengkap (*incomplete data*).

Berdasarkan buku Profil Kesehatan Jawa Timur Tahun 2000 dan Tahun 2001, ditemukan banyak data hilang pada data Kesehatan Ibu dan Anak. Data yang hilang tersebut persentasenya berkisar antara 16 % sampai dengan 40 %. Salah satu cara untuk mengatasi data hilang adalah metode imputasi. Metode imputasi adalah mengisi data hilang dengan nilai yang diperkirakan cukup layak dan kemudian dianalisis dengan metode baku untuk data lengkap.

Tujuan dari penelitian ini adalah membandingkan metode imputasi untuk mengestimasi data hilang dengan metode *Series Mean*, *Mean of Nearby Points*, *Median of Nearby Points*, *Linier Interpolation*, *Linier Trend at Point* dan mencari metode mana yang lebih baik di antara kelima metode tersebut.

Data yang digunakan dalam penelitian ini adalah data Kesehatan Ibu dan Anak di Propinsi Jawa Timur, yang berasal dari Laporan Tahunan Dinas Kesehatan Provinsi Jawa Timur tahun 2000. Variabel-variabel yang akan diteliti adalah K1 (kunjungan ibu hamil pertama kali), K4 (kunjungan ibu hamil ke empat) dan Linakes (pertolongan persalinan oleh tenaga kesehatan).

Untuk menganalisis data hilang pada data kesehatan ibu dan anak dilakukan tahapan kegiatan sebagai berikut :

Pertama, data kesehatan ibu dan anak yang lengkap, sengaja dihilangkan secara acak dengan menggunakan soft ware statistik. Pada setiap variabel item yang akan dihilangkan mulai dari 5 %, 10 %, 15 %, 20 % dan terakhir 25 %.

Kedua, item yang hilang tersebut akan dilengkapi lagi dengan metode *Series Mean*, *Mean of Nearby Points (2)*, *Median of Nearby Points (2)*, *Linier Interpolation* dan *Linier Trend at Point* dengan menggunakan soft ware statistik.

Ketiga, untuk melihat apakah ada perbedaan antara ke lima metode dengan item yang hilang digunakan uji Anova Sama Subyek dan untuk memilih metode mana yang berbeda dalam mengestimasi item yang hilang digunakan uji t Sampel Berpasangan.

Keempat, untuk menentukan metode imputasi mana yang yang baik dalam mengestimasi item hilang dilihat nilai korelasinya ( $r$ ).

Setelah dilakukan uji statistik metode imputasi yang baik adalah metode *mean of nearby points*, *median of nearby points* dan *linier interpolation* dan yang tidak baik adalah metode *series means* dan metode *linier trend at point*. Dari 3 metode imputasi yang baik, secara keseluruhan metode imputasi yang paling baik dalam mengestimasi item hilang adalah metode *median of nearby points*, karena jumlahnya yang paling banyak muncul, yaitu 9 kali (60 %).

Bagi peneliti lain yang tertarik dengan analisis data hilang, disarankan untuk mempelajari metode imputasi yang lain dan data yang digunakan adalah data individual, bukan data kelompok (*agregat*) yang sudah merupakan rekapitulasi data.