

# Menangani *Missing Value* dengan Metode Imputasi Menggunakan *Package MICE* di R

Cindy Pramudita<sup>a\*</sup>, Naura Aiswarya Asri Wicaksono<sup>a</sup>, Anjar Kharisma Hartiyah<sup>a</sup>, Annisa Maisaroh<sup>a</sup>, Iqbal Kharisudin<sup>a</sup>

<sup>a</sup> Statistika dan Sains Data, Universitas Negeri Semarang, Gunungpati, Semarang, 50229, Indonesia

\* Alamat Surel: [cindypramudita@students.unnes.ac.id](mailto:cindypramudita@students.unnes.ac.id)

## Abstrak

Nilai yang hilang dalam data sering kali menjadi tantangan dalam analisis statistik, mengurangi validitas hasil dan berpotensi menyebabkan bias. Penelitian ini berfokus pada penggunaan metode *Multiple Imputation by Chained Equations* (MICE) dengan paket MICE di R untuk menangani masalah *missing value*. MICE dipilih karena kemampuannya dalam menangani berbagai jenis kasus pada nilai data yang hilang, seperti hilang sepenuhnya secara acak (*Missing Completely at Random*), hilang secara acak (*Missing at Random*), dan hilang tidak secara acak (*Missing Not at Random*). Kinerja metode MICE dibandingkan dengan beberapa metode imputasi sederhana, seperti *mean substitution* dan *listwise deletion*, dalam menjaga keakuratan hasil serta struktur korelasi antar variabel. Berdasarkan evaluasi, MICE terbukti memberikan estimasi yang lebih akurat dan mengurangi bias, terutama ketika menghadapi data yang hilang dalam jumlah yang signifikan. Penggunaan MICE memungkinkan analisis yang lebih valid dan hasil yang lebih andal dalam menghadapi permasalahan nilai hilang.

## Kata kunci:

*Missing Value*, Imputasi Data, *Multiple Imputation*, *Chained Equations*.

© 2025 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

## 1. Pendahuluan

Masalah nilai hilang (*missing value*) kerap muncul dalam berbagai proses pengumpulan dan analisis data. Faktor-faktor penyebabnya meliputi kesalahan manusia saat pemrosesan, kerusakan perangkat pengumpulan data, penolakan responden untuk menjawab pertanyaan tertentu, *dropout* atau berhenti berpartisipasi dalam studi longitudinal, serta penggabungan data dari berbagai sumber yang tidak terkait (Groothuis-oudshoorn, 2011). Kehadiran *missing value* dapat menimbulkan berbagai masalah serius, seperti penurunan kinerja analisis, hasil yang bias, dan interpretasi yang keliru. Oleh karena itu, penanganan nilai yang hilang dengan teknik yang tepat sangat penting agar hasil analisis tetap valid dan dapat diandalkan.

Tingkat dampak dari nilai yang hilang bergantung pada jumlah data yang hilang, pola hilangnya data (*missing completely at random*, *missing at random*, atau *not missing at random*), serta mekanisme yang mendasarinya (Azur et al., 2011). Untuk mengatasi hal ini, terdapat dua pendekatan utama: penghapusan data dan *imputation*. Penghapusan dapat menyebabkan hilangnya informasi berharga dan mengurangi ukuran sampel secara signifikan, terutama jika data tidak hilang secara acak (Little & Rubin, 2002). Di sisi lain, *imputation*, yaitu proses mengganti nilai yang hilang dengan estimasi tertentu, dianggap sebagai metode yang lebih baik karena mempertahankan integritas *dataset* (Rubin, 1987).

Beberapa metode *imputation* sederhana yang sering digunakan adalah *mean imputation*, regresi, dan *K-Nearest Neighbors* (KNN). Namun, metode konvensional ini memiliki keterbatasan, terutama dalam menangani data berukuran besar dan kompleks di era *big data*. Teknik-teknik ini tidak selalu optimal dalam menghasilkan estimasi yang akurat ketika data memiliki banyak dimensi atau pola *non-linear*. Sebagai

### To cite this article:

Pramudita, C., Wicaksono, N. A. A., & Hartiyah, A. K., Maisaroh, A., Kharisudin, I. (2025). Menangani *Missing Value* dengan Metode Imputasi Menggunakan *Package MICE* di R. *PRISMA, Prosiding Seminar Nasional Matematika* 8, 399-406

solusi, metode *Multiple Imputation by Chained Equations* (MICE) menjadi populer karena mampu melakukan *imputation* secara iteratif dan memperhitungkan ketidakpastian estimasi dengan menghasilkan beberapa dataset lengkap (Groothuis-oudshoorn, 2011). *Package* MICE di R memudahkan peneliti untuk menerapkan metode MICE secara efisien (Alruhaymi & Kim, 2021). Hal ini menjadikannya pilihan yang populer untuk menangani *missing value* dalam berbagai jenis penelitian, terutama yang melibatkan data kompleks. Studi ini bertujuan untuk mengeksplorasi penerapan metode MICE menggunakan R dalam menangani nilai yang hilang.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi praktis dan akademis dalam penanganan *missing value*, terutama dalam konteks data besar dan kompleks. Metode *imputation* yang tepat akan meningkatkan kualitas data, memungkinkan analisis yang lebih akurat, dan menghasilkan kesimpulan yang dapat diandalkan.

---

## 2. Metode

Penelitian ini menggunakan metode imputasi berganda (*Multiple Imputation*) untuk menangani nilai yang hilang pada dataset dengan memanfaatkan *package* MICE di R. Imputasi berganda merupakan teknik yang digunakan untuk mengatasi *missing value* dengan menghasilkan beberapa dataset lengkap melalui proses imputasi lalu menggabungkan hasil analisis dari dataset untuk mendapatkan prediksi yang lebih akurat.

### 2.1. Deskripsi Dataset

Dataset ini berisi informasi yang relevan untuk menganalisis pola kerusakan dan pemeliharaan kendaraan, memberikan landasan bagi pemahaman yang lebih dalam mengenai hubungan antara penggunaan kendaraan, biaya perawatan, dan faktor-faktor lainnya. Dengan total 1.624 entri, dataset ini mencakup tujuh variabel utama yang memberikan wawasan mendalam tentang performa kendaraan, durasi, dan biaya yang terkait dengan perbaikan, serta konteks lingkungan di mana kendaraan beroperasi. Struktur dataset ini dirancang untuk memberikan perspektif yang komprehensif terhadap bagaimana kendaraan mengalami kerusakan dan memerlukan pemeliharaan, menjadikannya sumber data yang penting untuk analisis yang bertujuan meningkatkan efisiensi dan manajemen kendaraan. Dataset ini sendiri mencakup tujuh variabel utama, yaitu: *vehicle*, yang berupa data numerik; *fm* (*failure month*), juga berupa data numerik; *mileage*, yang menunjukkan jarak tempuh; *lh* (*labor hours*) dan *lc* (*labor cost*), yang keduanya berupa data numerik terkait waktu dan biaya tenaga kerja; *mc* (*maintenance cost*), yang menggambarkan biaya pemeliharaan; serta *state*, yang merupakan variabel kategori. Beberapa variabel dalam dataset ini mengandung nilai yang hilang (*missing value*), sehingga perlu dilakukan penanganan khusus agar analisis data yang dilakukan menjadi lebih valid, komprehensif, dan akurat. Dengan struktur ini, dataset memberikan gambaran menyeluruh tentang hubungan antara penggunaan kendaraan, kerusakan, dan kebutuhan pemeliharaan, sehingga menjadi sumber yang kaya untuk analisis lebih lanjut.

### 2.2. Eksplorasi Data

Tahapan awal dalam analisis data dimulai dengan proses eksplorasi data, yang bertujuan untuk memahami lebih dalam struktur, karakteristik, dan distribusi data yang tersedia. Eksplorasi data merupakan langkah penting yang membantu mengidentifikasi pola, anomali, serta potensi masalah seperti *missing value*, sebelum melangkah ke analisis lebih lanjut. Dalam penelitian ini, eksplorasi data diawali dengan analisis deskriptif menggunakan fungsi *str()* dan *summary()* di R, yang memberikan gambaran umum mengenai tipe data dan statistik dasar dari setiap variabel.

Setelah itu, perhitungan persentase *missing value* untuk setiap variabel dilakukan menggunakan fungsi *apply()*, guna mengetahui sejauh mana data yang hilang dapat mempengaruhi hasil analisis. Pola *missing value* divisualisasikan lebih lanjut dengan menggunakan fungsi *md.pattern()* dari *package* MICE, yang menunjukkan kombinasi *missing value* antar variabel. Selain itu, distribusi *missing value* pada variabel spesifik seperti *Mileage* dan *lc* divisualisasikan menggunakan *marginplot()* dari *package* VIM, sehingga dapat memberikan gambaran lebih detail mengenai hubungan antara variabel-variabel tersebut dan bagaimana data yang hilang terdistribusi.

### 2.3. Pemilihan Metode Imputasi

Untuk menangani *missing value* pada dataset, penting untuk memilih metode yang mampu mempertahankan integritas data asli sekaligus meminimalkan distorsi analisis. Salah satu teknik yang

efektif digunakan adalah *Multiple Imputation by Chained Equations* (MICE). Metode ini memungkinkan kita untuk melakukan imputasi pada data yang hilang dengan mempertimbangkan pola dan hubungan antar variabel dalam dataset. MICE tidak hanya mengisi *missing value*, tetapi juga menghasilkan beberapa dataset yang terpisah, sehingga memperhitungkan ketidakpastian yang ada pada data yang diimputasi. Hal ini meningkatkan akurasi dan keandalan hasil analisis.

Dalam penelitian ini, teknik *Predictive Mean Matching* (PMM) diterapkan untuk mengisi *missing value* pada variabel numerik, seperti *Mileage*, *lh*, dan *lc*, karena PMM mampu menghasilkan nilai imputasi yang mendekati distribusi asli data numerik tersebut. Sementara itu, untuk variabel kategorikal, digunakan metode regresi logistik agar hasil imputasi tetap konsisten dengan kategori aslinya. Penelitian ini menetapkan  $m = 3$  (tiga kali imputasi), artinya tiga dataset lengkap dihasilkan dari proses ini, yang kemudian digabungkan untuk mendapatkan estimasi yang lebih akurat dan memperhitungkan variasi antar imputasi.

#### 2.4. Proses Imputasi Berganda

Proses imputasi dimulai dengan membuat *package* yang diperlukan (MICE, VIM, dan *lattice*) dan mengimputasi *missing value* pada dataset menggunakan fungsi *MICE()*. Kolom yang diimputasi adalah dari kolom ke-2 hingga kolom ke-7,

```
impute <- mice(data[,2:7], m=3, seed = 123)
```

berikut adalah langkah-langkah imputasi berganda yang dilakukan:

- menetapkan  $seed = 123$  untuk memastikan hasil imputasi yang konsisten,
- menggunakan metode PMM untuk variabel numerik dan metode regresi untuk variabel kategorikal,
- melakukan imputasi sebanyak tiga kali untuk menghasilkan tiga dataset yang lengkap.

#### 2.5. Evaluasi Hasil Imputasi

Setelah proses imputasi dilakukan, langkah selanjutnya adalah memastikan bahwa hasil imputasi sesuai dengan distribusi data asli dan tidak memperkenalkan bias baru. Untuk tujuan ini, dataset yang telah diimputasi diekstraksi menggunakan fungsi *complete()* dari *package* MICE. Evaluasi terhadap kualitas imputasi dilakukan melalui berbagai visualisasi. Salah satu teknik yang digunakan adalah *boxplot()*, yang berfungsi untuk membandingkan distribusi antara data asli dan data yang telah diimputasi. Dengan cara ini, kita dapat mengidentifikasi apakah nilai yang diimputasi berada dalam rentang yang wajar dan tidak menyimpang jauh dari data awal.

Selain itu, visualisasi *xypplot()* digunakan untuk mengevaluasi konsistensi hasil imputasi, khususnya pada variabel *lc* dan *lh*, di berbagai iterasi (imputasi ke-1, ke-2, dan ke-3). Visualisasi ini memberikan gambaran apakah hasil imputasi konsisten dan stabil di setiap iterasi, yang penting untuk memastikan bahwa tidak ada pola atau bias yang tidak diinginkan dalam data yang diimputasi. Tujuan dari evaluasi ini adalah untuk memastikan bahwa proses imputasi berjalan optimal dan tidak mengubah karakteristik asli dari dataset.

#### 2.6. Validasi dan Interpretasi

Setelah proses imputasi selesai, langkah selanjutnya adalah melakukan validasi terhadap dataset yang sudah diimputasi. Validasi ini dilakukan dengan cara membandingkan hasil imputasi dengan dataset asli melalui analisis statistik deskriptif. Analisis ini bertujuan untuk memastikan bahwa data yang telah diimputasi memiliki distribusi yang serupa dengan data asli. Selain itu, visualisasi hasil imputasi melalui berbagai grafik digunakan untuk mengecek apakah data setelah imputasi tetap konsisten dan tidak menyimpang dari pola data awal.

Jika hasil analisis menunjukkan bahwa data yang diimputasi memiliki kesesuaian yang baik dengan data asli, maka dataset tersebut dapat dianggap valid dan siap untuk digunakan dalam analisis lebih lanjut. Proses validasi ini sangat penting untuk memastikan bahwa hasil penelitian menjadi lebih andal, serta mengurangi risiko bias yang mungkin muncul akibat adanya *missing value* dalam dataset awal. Dengan demikian, langkah ini membantu meningkatkan akurasi dan kredibilitas analisis secara keseluruhan.

### 3. Hasil dan Pembahasan

Pada penelitian ini, dilakukan analisis *missing data* menggunakan metode *multiple imputation* dengan bantuan *package* MICE di R. Berdasarkan hasil eksplorasi data, didapati bahwa variabel-variabel tersebut memiliki persentase *missing value* yang bervariasi. Tabel di bawah menunjukkan hasil analisis *missing value* pada setiap variabel yang dianalisis tersebut.

**Tabel 1.** Persentase *missing value* tiap variabel

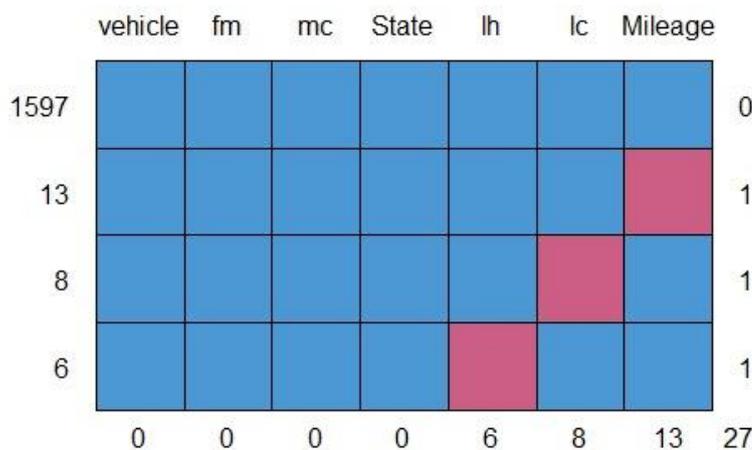
vehicle	fm	Mileage	lh	lc	mc	State
0.0000000	0.0000000	0.8004926	0.3694581	0.4926108	0.0000000	0.0000000

Jika dipersentasekan *missing value* menonjol yang ditunjukkan beberapa variabel yaitu: *Mileage* (0.80%), *lh* (0.37%), dan *lc* (0.49%). Meskipun persentase *missing value* yang cukup rendah dibawah 1% namun perlu diperhatikan. Sebagai contoh, *Mileage* dan *lc* yang merupakan variabel kunci dalam analisis bisa menyebabkan distorsi pada model jika *missing value* tidak diimputasi.

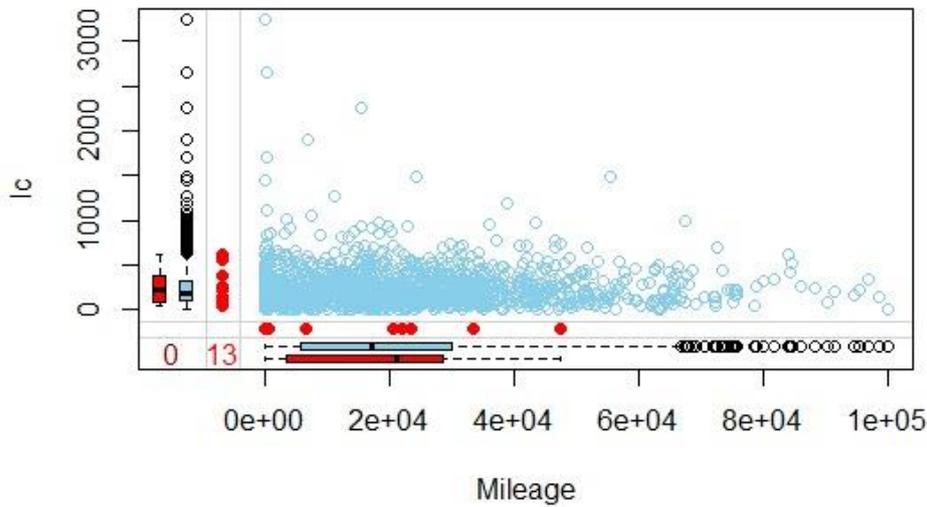
**Tabel 2.** Visualisasi *missing value* bentuk tabel

Entri	vehicle	fm	mc	State	lh	lc	Mileage	Missing values
1597	1	1	1	1	1	1	1	0
13	1	1	1	1	1	1	0	1
8	1	1	1	1	1	0	1	1
6	1	1	1	1	0	1	1	1
	0	0	0	6	8	13	15	27

Visualisasi pola *missing value* menunjukkan bahwa sebanyak 1.597 entri tidak memiliki *missing value* pada semua variabel, sementara 13 entri memiliki *missing value* pada variabel *Mileage*, 8 entri pada variabel *lc*, dan 6 entri pada variabel *lh*. Hal ini mencerminkan pola ketidakhadiran data yang lebih kompleks. Analisis ini memberikan gambaran awal mengenai distribusi *missing value* yang penting untuk pemilihan metode imputasi yang tepat. Hasil pola *missing value* menunjukkan bahwa sebagian besar data dalam dataset ini lengkap, dengan mayoritas entri (1597) tidak mengalami *missing value*. Namun, terdapat *missing value* yang tersebar pada variabel-variabel lain dengan persentase tertinggi pada variabel *Mileage*. Gambar berikut memperlihatkan pola *missing data* pada dataset.



**Gambar 1.** Visualisasi *missing value* bentuk heatmap



**Gambar 2.** Visualisasi *missing value* dalam bentuk *scatter plot*

Pada visualisasi *scatter plot* di atas ini, terlihat bahwa data *Mileage* cenderung tersebar, namun terdapat beberapa *outliers* yang signifikan pada variabel *lc*. Boxplot pada bagian kiri bawah grafik menunjukkan distribusi data yang tidak normal, dimana terdapat banyak *outliers* pada kedua variabel tersebut. Selain itu, visualisasi *scatter plot* memberikan informasi penting terkait distribusi data. Pada variabel *lc*, ditemukan banyak *outliers* yang mungkin disebabkan oleh adanya data yang tidak konsisten atau *error* pada pengumpulan data. Adanya *outliers* ini penting untuk ditangani dalam analisis lebih lanjut agar tidak mempengaruhi hasil akhir dari analisis regresi atau model lainnya yang akan digunakan pada dataset ini.

**Tabel 3.** Iter imp variable

iter	itp	variable		
1	1	Mileage	lh	lc
1	2	Mileage	lh	lc
1	3	Mileage	lh	lc
2	1	Mileage	lh	lc
2	2	Mileage	lh	lc
2	3	Mileage	lh	lc
3	1	Mileage	lh	lc
3	2	Mileage	lh	lc
3	3	Mileage	lh	lc
4	1	Mileage	lh	lc
4	2	Mileage	lh	lc
4	3	Mileage	lh	lc
5	1	Mileage	lh	lc
5	2	Mileage	lh	lc
5	3	Mileage	lh	lc

Dataset yang digunakan berisi variabel *Mileage*, *lh*, *lc* yang mengalami *missing value*. Berdasarkan analisis *missing value* menggunakan metode imputasi *multiple imputation by chained equations* (MICE), dilakukan proses imputasi pada variabel-variabel yang mengalami kekosongan data. Proses ini dilakukan dalam beberapa iterasi seperti yang terlihat pada tabel *iter imp variable* di bawah. Tabel ini menunjukkan proses imputasi dalam 5 iterasi untuk setiap variabel yang mengalami *missing value*, yaitu *Mileage*, *lh*, dan

lc. Proses imputasi menggunakan metode MICE menunjukkan bahwa metode ini dapat menangani data yang hilang secara efektif dengan mempertimbangkan hubungan antar variabel. Dari *output* iterasi yang ditampilkan, kita dapat melihat bahwa imputasi pada variabel *Mileage*, *lh*, dan *lc* dilakukan secara konsisten di setiap iterasi. Hal ini mengindikasikan bahwa ada korelasi yang cukup kuat antar variabel yang dimanfaatkan dalam proses imputasi, sehingga data yang hilang dapat diprediksi dengan baik.

```

```{r}
print(impute)
```
Class: mids
Number of multiple imputations: 3
Imputation methods:
   fm Mileage   lh   lc   mc   State
   ""  "pmm"   "pmm" "pmm" ""    ""
    
```

Pada bagian ini diberikan penjelasan tentang proses dan hasil imputasi menggunakan metode *Predictive Mean Matching* (PMM). Pada data yang ditampilkan, ada tiga kali proses imputasi yang dilakukan, dengan beberapa variabel yang diimputasi, yaitu *Mileage*, *lh*, *lc*. Proses imputasi yang digunakan adalah metode PMM, yang bertujuan untuk mengatasi masalah *missing value* dengan mengimputasi nilai yang mirip secara statistik dengan nilai yang hilang. Pada tabel dan plot di atas, terlihat bahwa beberapa variabel seperti *Mileage*, *lh*, dan *lc* diimputasi menggunakan metode ini. Terlihat ada tiga kali iterasi imputasi yang dilakukan, dan hal ini memberikan kita tiga set nilai yang diimputasi. Berikut untuk hasil proses imputasi tersebut.

**Tabel 4.** Dataset sebelum diimputasi

| vehicle | fm    | Mileage | lh    | lc     | mc     | State |
|---------|-------|---------|-------|--------|--------|-------|
| <int>   | <int> | <int>   | <dbl> | <dbl>  | <dbl>  | <chr> |
| 15      | 2     | 38019   | 3.6   | 0.0    | 119.66 | MS    |
| 814     | 19    | 25538   | 2.6   | 74.57  | 119.66 | CA    |
| 1163    | 2     | 20746   | 20.2  | 148.79 | 119.66 | WI    |
| 510     | 11    | 6442    | 7.5   | 30.68  | 322.18 | OR    |
| 676     | 12    | 2361    | 2.2   | 235.38 | 55.8   | AZ    |
| 1474    | 17    | 67      | 2.1   | 245.15 | 119.66 | FL    |
| 1180    | 19    | 6074    | 0.9   | 44.0   | 119.66 | LA    |
| 407     | 14    | 470     | 1.9   | 19.04  | 136.36 | GA    |
| 281     | 16    | 35033   | 2.2   | 118.6  | 119.66 | WA    |
| 1494    | 11    | 67507   | 0.0   | 995.3  | 119.66 | PA    |
| 1442    | 14    | 75587   | 1.2   | 232.06 | 119.66 | TX    |
| 6       | 11    | 35178   | 5.1   | 0.0    | 175.46 | LA    |
| 437     | 9     | 30701   | 0.8   | 57.5   | 119.66 | FL    |
| 542     | 6     | 5       | 0.8   | 382.77 | 206.14 | OH    |
| 1342    | 5     | 23172   | 2.4   | 173.13 | 175.46 | MA    |
| 544     | 9     | 52983   | 2.3   | 48.55  | 175.46 | VA    |
| 108     | 6     | 4644    | 1.9   | 186.3  | 119.66 | OH    |
| 822     | 3     | 2888    | 3.7   | 74.4   | 206.14 | CO    |
| 418     | 11    | NA      | 2.4   | 55.68  | 175.46 | FL    |
| 1024    | 1     | NA      | 1.0   | 218.74 | 119.66 | NH    |
| 1064    | 5     | 33466   | 3.1   | 139.17 | 119.66 | IN    |

Data di atas adalah dataset yang menunjukkan informasi tentang kendaraan sebelum dilakukan imputasi. Terdapat beberapa nilai yang hilang (*missing values*) pada data ke-19 dan ke-20, yang dapat mempengaruhi analisis lebih lanjut. Imputasi diperlukan untuk mengisi nilai-nilai yang hilang ini agar dataset dapat digunakan secara optimal dalam proses analisis data selanjutnya.

**Tabel 5.** Dataset setelah diimputasi

| vehicle | fm    | Mileage | lh    | lc     | mc     | State |
|---------|-------|---------|-------|--------|--------|-------|
| <int>   | <int> | <int>   | <dbl> | <dbl>  | <dbl>  | <chr> |
| 15      | 2     | 38019   | 3.6   | 0.0    | 119.66 | MS    |
| 814     | 19    | 25538   | 2.6   | 74.57  | 119.66 | CA    |
| 1163    | 2     | 20746   | 20.2  | 148.79 | 119.66 | WI    |
| 510     | 11    | 6442    | 7.5   | 30.68  | 322.18 | OR    |
| 676     | 12    | 2361    | 2.2   | 235.38 | 55.8   | AZ    |
| 1474    | 17    | 67      | 2.1   | 245.15 | 119.66 | FL    |
| 1180    | 19    | 6074    | 0.9   | 44.0   | 119.66 | LA    |
| 407     | 14    | 470     | 1.9   | 19.04  | 136.36 | GA    |
| 281     | 16    | 35033   | 2.2   | 118.6  | 119.66 | WA    |
| 1494    | 11    | 67507   | 0.0   | 995.3  | 119.66 | PA    |
| 1442    | 14    | 75587   | 1.2   | 232.06 | 119.66 | TX    |
| 6       | 11    | 35178   | 5.1   | 0.0    | 175.46 | LA    |
| 437     | 9     | 30701   | 0.8   | 57.5   | 119.66 | FL    |
| 542     | 6     | 5       | 0.8   | 382.77 | 206.14 | OH    |
| 1342    | 5     | 23172   | 2.4   | 173.13 | 175.46 | MA    |
| 544     | 9     | 52983   | 2.3   | 48.55  | 175.46 | VA    |
| 108     | 6     | 4644    | 1.9   | 186.3  | 119.66 | OH    |
| 822     | 3     | 2888    | 3.7   | 74.4   | 206.14 | CO    |
| 418     | 11    | 11267   | 2.4   | 55.68  | 175.46 | FL    |
| 1024    | 1     | 11251   | 1.0   | 218.74 | 119.66 | NH    |
| 1064    | 5     | 33466   | 3.1   | 139.17 | 119.66 | IN    |

Dataset yang ditampilkan di atas telah mengalami pembaruan melalui proses imputasi, khususnya pada baris 19 dan 20, yang sebelumnya memiliki nilai yang hilang pada kolom *Mileage*. Setelah dilakukan imputasi, nilai pada baris 19 kini menjadi 11267, dan pada baris 20 menjadi 11251. Proses ini penting untuk memastikan bahwa semua data yang diperlukan tersedia, sehingga analisis yang dilakukan dapat menjadi lebih akurat dan valid. Dengan adanya data yang terisi, dataset ini sekarang siap untuk digunakan dalam berbagai analisis lebih lanjut, termasuk perhitungan rata-rata, analisis regresi, dan pemodelan prediktif, tanpa adanya gangguan dari nilai yang hilang.

#### 4. Simpulan

Penelitian ini menunjukkan bahwa metode *multiple imputation by chained equations* (MICE) merupakan teknik yang efektif dalam menangani masalah *missing value* pada dataset yang kompleks. Dengan menggunakan MICE, variabel yang memiliki nilai hilang, seperti *Mileage*, *lh*, dan *lc*, dapat diimputasi secara akurat melalui beberapa iterasi yang memanfaatkan hubungan antar variabel. Proses imputasi ini memungkinkan dataset menjadi lengkap tanpa mengorbankan kualitas data yang ada. Analisis terhadap hasil visualisasi menunjukkan adanya pola distribusi yang cenderung mengandung *outliers*, khususnya pada variabel *lc*. Meskipun demikian, metode MICE mampu memperhitungkan hal ini dalam proses imputasi, sehingga hasil akhirnya dapat diandalkan untuk keperluan analisis lanjutan.

---

**Daftar Pustaka**

- Alruhaymi, A. Z., & Kim, C. J. (2021). Why Can Multiple Imputations and How (MICE) Algorithm Work? *Open Journal of Statistics*, 11(05), 759–777. <https://doi.org/10.4236/ojs.2021.115045>
- Azur, M. J., Stuart, E. A., Frangakis, C., & Leaf, P. J. (2011). Multiple imputation by chained equations: what is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20(1), 40–49. <https://doi.org/10.1002/mpr.329>
- Groothuis-oudshoorn, K. (2011). *mice : Multivariate Imputation by Chained*. 45(3).
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley. <https://doi.org/10.1002/9781119013563>
- Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley. <https://doi.org/10.1002/9780470316696>