|  |  |  |  |
| --- | --- | --- | --- |
|  | Data1 (MCAR) | Data2 (MAR) | Data3 (MNAR) |
| Listwise deletion | |  |  | | --- | --- | | **Mean(B1)** | **5.694141** | | **Var(B1)** | **0.04886532** | | **Mean(B2)** | **3.768373** | | **Var(B2)** | **0.1863263** | | **Mean(mse)** | **474.8636** | | **Var(mse)** | **851.5889** | | **B1 CI coverage** | **99.4%** | | **B2 CI coverage** | **99.1%** | | |  |  | | --- | --- | | Mean(B1) | 6.062758 | | Var(B1) | 0.06788733 | | Mean(B2) | 4.165709 | | Var(B2) | 0.1855791 | | Mean(mse) | 497.5082 | | Var(mse) | 775.333 | | B1 CI coverage | 83.9% | | B2 CI coverage | 97.6 | | |  |  | | --- | --- | | Mean(B1) | 5.179553 | | Var(B1) | 0.05314329 | | Mean(B2) | 2.992119 | | Var(B2) | 0.1600603 | | Mean(mse) | 433.1715 | | Var(mse) | 745.1286 | | B1 CI coverage | 79.8% | | B2 CI coverage | 76.2 | |
| Pairwise deletion | |  |  | | --- | --- | | Mean(B1) | 5.72203 | | Var(B1) | 0.05890984 | | Mean(B2) | 3.824874 | | Var(B2) | 0.240652 | | Mean(mse) | 477.2147 | | Var(mse) | 762.2969 | | B1 CI coverage | 97.7% | | B2 CI coverage | 97.3% | | |  |  | | --- | --- | | **Mean(B1)** | **5.575796** | | **Var(B1)** | **0.06066919** | | **Mean(B2)** | **3.782697** | | **Var(B2)** | **0.2621779** | | **Mean(mse)** | **503.3784** | | **Var(mse)** | **797.5962** | | **B1 CI coverage** | **99.3%** | | **B2 CI coverage** | **97.7** | | |  |  | | --- | --- | | Mean(B1) | 5.622921 | | Var(B1) | 0.05282692 | | Mean(B2) | 2.528733 | | Var(B2) | 0.2085408 | | Mean(mse) | 436.8206 | | Var(mse) | 701.896 | | B1 CI coverage | 99.1% | | B2 CI coverage | 36.5% | |
| Arithmetic mean imputation | |  |  | | --- | --- | | Mean(B1) | 4.589361 | | Var(B1) | 0.04033997 | | Mean(B2) | 2.852796 | | Var(B2) | 0.1442038 | | Mean(mse) | 500.4164 | | Var(mse) | 554.3332 | | B1 CI coverage | 0.5% | | B2 CI coverage | 56.9% | | |  |  | | --- | --- | | Mean(B1) | 4.479298 | | Var(B1) | 0.04226803 | | Mean(B2) | 2.896221 | | Var(B2) | 0.1466641 | | Mean(mse) | 513.1831 | | Var(mse) | 555.4428 | | B1 CI coverage | 0.2% | | B2 CI coverage | 61.4% | | |  |  | | --- | --- | | Mean(B1) | 4.510003 | | Var(B1) | 0.0344106 | | Mean(B2) | 2.254511 | | Var(B2) | 0.1072829 | | Mean(mse) | 492.4254 | | Var(mse) | 573.7631 | | B1 CI coverage | 0% | | B2 CI coverage | 5.7% | |
| Regression imputation | |  |  | | --- | --- | | Mean(B1) | 5.736118 | | Var(B1) | 0.03827928 | | Mean(B2) | 5.038599 | | Var(B2) | 0.2159695 | | Mean(mse) | 356.0172 | | Var(mse) | 414.7547 | | B1 CI coverage | 91.1% | | B2 CI coverage | 17.7% | | |  |  | | --- | --- | | Mean(B1) | 5.820336 | | Var(B1) | 0.04588616 | | Mean(B2) | 5.304823 | | Var(B2) | 0.2267846 | | Mean(mse) | 356.5974 | | Var(mse) | 460.5254 | | B1 CI coverage | 80.1% | | B2 CI coverage | 8.1% | | |  |  | | --- | --- | | Mean(B1) | 5.487992 | | Var(B1) | 0.03979937 | | Mean(B2) | 4.745784 | | Var(B2) | 0.2415886 | | Mean(mse) | 356.789 | | Var(mse) | 454.0791 | | B1 CI coverage | 92.6% | | B2 CI coverage | 42% | |
| Stochastic regression imputation | |  |  | | --- | --- | | Mean(B1) | 5.774865 | | Var(B1) | 0.04811797 | | Mean(B2) | 3.74872 | | Var(B2) | 0.2647226 | | Mean(mse) | 474.7854 | | Var(mse) | 846.445 | | B1 CI coverage | 91.1% | | B2 CI coverage | 91.8% | | |  |  | | --- | --- | | Mean(B1) | 5.825981 | | Var(B1) | 0.05244835 | | Mean(B2) | 3.960301 | | Var(B2) | 0.2404971 | | Mean(mse) | 477.3178 | | Var(mse) | 865.2544 | | B1 CI coverage | 86.5% | | B2 CI coverage | 91.1% | | |  |  | | --- | --- | | **Mean(B1)** | **5.548324** | | **Var(B1)** | **0.05250273** | | **Mean(B2)** | **3.471093** | | **Var(B2)** | **0.2679515** | | **Mean(mse)** | **475.322** | | **Var(mse)** | **824.6848** | | **B1 CI coverage** | **95.5%** | | **B2 CI coverage** | **84.7%** | |
| Hot-Deck imputation | |  |  | | --- | --- | | Mean(B1) | 5.517064 | | Var(B1) | 0.05670436 | | Mean(B2) | 2.597203 | | Var(B2) | 0.1888859 | | Mean(mse) | 546.4872 | | Var(mse) | 1203.383 | | B1 CI coverage | 95.8% | | B2 CI coverage | 28.6% | | |  |  | | --- | --- | | Mean(B1) | 5.491656 | | Var(B1) | 0.04576221 | | Mean(B2) | 2.628071 | | Var(B2) | 0.1814948 | | Mean(mse) | 530.9809 | | Var(mse) | 996.0477 | | B1 CI coverage | 96.2% | | B2 CI coverage | 29.3% | | |  |  | | --- | --- | | Mean(B1) | 5.373908 | | Var(B1) | 0.04435514 | | Mean(B2) | 2.979624 | | Var(B2) | 0.1919834 | | Mean(mse) | 519.2088 | | Var(mse) | 1085.482 | | B1 CI coverage | 88.6% | | B2 CI coverage | 61.3% | |
| Similar response pattern imputation | |  |  | | --- | --- | | Mean(B1) | 5.585122 | | Var(B1) | 0.03979682 | | Mean(B2) | 4.284299 | | Var(B2) | 0.2110076 | | Mean(mse) | 400.0308 | | Var(mse) | 508.3295 | | B1 CI coverage | 96.8% | | B2 CI coverage | 78.% | | |  |  | | --- | --- | | Mean(B1) | 5.480053 | | Var(B1) | 0.05780924 | | Mean(B2) | 4.711881 | | Var(B2) | 0.3207243 | | Mean(mse) | 400.7218 | | Var(mse) | 641.1758 | | B1 CI coverage | 89.1% | | B2 CI coverage | 42.1% | | |  |  | | --- | --- | | Mean(B1) | 5.198421 | | Var(B1) | 0.05191759 | | Mean(B2) | 3.269251 | | Var(B2) | 0.4416955 | | Mean(mse) | 421.4814 | | Var(mse) | 661.5728 | | B1 CI coverage | 55.5% | | B2 CI coverage | 69.9% | |
| Indicator method imputation | |  |  | | --- | --- | | Mean(B1) | 5.789049 | | Var(B1) | 0.06480644 | | Mean(B2) | 3.401023 | | Var(B2) | 0.4465931 | | Mean(mse) | 528.1518 | | Var(mse) | 6431.818 | | B1 CI coverage | 87.8% | | B2 CI coverage | 78.4% | | |  |  | | --- | --- | | Mean(B1) | 5.305567 | | Var(B1) | 0.5871574 | | Mean(B2) | 3.541938 | | Var(B2) | 0.5297892 | | Mean(mse) | 551.7196 | | Var(mse) | 9214.57 | | B1 CI coverage | 44% | | B2 CI coverage | 81.6% | | |  |  | | --- | --- | | Mean(B1) | 5.341773 | | Var(B1) | 0.2622543 | | Mean(B2) | 2.572897 | | Var(B2) | 1.522363 | | Mean(mse) | 541.0082 | | Var(mse) | 6011.128 | | B1 CI coverage | 60.7 % | | B2 CI coverage | 45.4% | |

B1 is for size, B2 is for rating.

Method that gave the best coverage is highlighted for each dataset.

Pairwise and listwise performed very well for MCAR and MAR, but relatively poor on MNAR.

Arithmetic imputation yielded acceptable mse but CI coverages were poor.

Most of the imputation methods took a long time to run.

Variances of mse for hot deck, indicator method imputation were large. The model may be potential overfitting the data.

Similar response pattern imputation performed fairly with the MNAR data set, comparing its performance with MCAR and MAR data sets.

Overall speaking, arithmetic mean imputation should not be used if inference of coefficients are desired because it gives very poor estimation of the true parameters. Simpler methods like list and pairwise deletion give pretty good parameter estimations. More complex and advanced imputation methods give better average MSEs but the parameter estimates can only be said to be fair, not to mention the highly-possible overfitting of the data which is indicated by high variance of the MSE.