Comparison and Summary

The table below compares the coefficients of the logistic models for each imputation method.

Variables/Methods	Listwise	Mean/Mod	Random	Dummy Var	Hotdeck	Regression	Regression +Noise	Multiple Imputation
Intercept	-1.447	-0.894	-1.095	-2.204	-1.259	-1.197	-1.198	-1.46
Training hours	-3.5 * 10^ -4	-9.4 * 10^-4	-8.5 * 10^-4	-8.7 * 10^-4	-7.7 *10^-4	-11.0 * 10^-4	-11.0 * 10^-4	0.00
Gender: Female	-0.186	0.017	0.071	0.091	0.046	0.075	0.075	0.10
Gender: Other	-0.372	-0.032	-0.090	-0.002	-0.056	-0.009	-0.009	0.01
Relevant experience	0.012	-0.284	-0.291	0.172	-0.286	-0.280	-0.280	-0.28
Last new job	0.019	0.023	0.021	0.060	0.014	0.016	0.016	0.02
University	0.136	0.224	0.229	0.159	0.242	0.226	0.226	0.23
Education level	0.161	0.220	0.194	0.308	0.223	0.202	0.202	0.21
Company size	0.019	-0.075	-0.0001	0.003	0.019	0.023	0.023	0.02
Experience	-0.085	-0.055	-0.006	-0.068	-0.057	-0.057	-0.057	-0.06

The table below compares the standard errors of estimates for each imputation method.

Variables/Methods	Listwise	Mean/Mod	Random	Dummy Var	Hotdeck	Regression	Regression +Noise	Multiple Imputation
Intercept	0.177	0.092	0.092	0.106	0.091	0.092	0.092	0.11
Training hours	4.64* 10^-4	3.22 * 10^-4	2.9 * 10^-4	3.3 *10^-4	2.9 * 10^-4	3.2 * 10^-4	2.9 * 10^-4	0.00
Gender: Female	0.099	0.069	0.060	0.072	0.060	0.059	0.059	0.07
Gender: Other	0.294	0.172	0.152	0.177	0.149	0.145	0.145	0.19
Relevant experience	0.078	0.041	0.041	0.046	0.041	0.041	0.041	0.04
Last new job	0.020	0.125	0.012	0.013	0.012	0.012	0.012	0.01
University	0.040	0.022	0.022	0.023	0.022	0.022	0.022	0.02
Education level	0.046	0.027	0.026	0.027	0.026	0.026	0.026	0.03
Company size	0.013	0.010	0.007	0.011	0.008	0.008	0.008	0.01
Experience	0.005	0.003	0.003	0.004	0.003	0.003	0.003	0.00
AIC	8573.1	20504	20568	19669	20529	20552	20549	N/A

The coefficients of the logistic regression model using the listwise imputation method are used as the reference, and the average of percent change on coefficients are calculated. For mean/mode imputation the average of percent change on coefficients is 348%; for random imputation the average of percent change on coefficients is 309%; for dummy variable imputation the average of percent change on coefficients is 218%; for hotdecking nearest neighborhood the average of percent change on coefficients is 295%; for regression imputation the average of percent change on coefficients is 301%; for regression with noise imputation the average of percent change on coefficients is 301%; for multiple imputation with mi package the average of percent change on coefficients is 294%.

For this dataset, it contains missing values originally, so it is impossible for us to know what the complete dataset looks like. The listwise method deleted more than 9000 observations, which is about 47% of the data. So the coefficients from the listwise imputation is not a good reference for deciding which imputation method is the best. The focus should be on the estimation error or the AIC of the model. The likewise imputation method has the smallest AIC due to much smaller data set size, and the dummy variable method has relatively smaller AIC due to extra dummy predictors. The AICs for other methods are similar. Amongst all imputation methods, the regression with noise and hotdecking nearest neighborhood methods have the smallest standard error in general. I would prefer these two methods in this case. With these two methods, the most important factors in identifying the candidates in this dataset who are truly looking for a job change is gender and relevant experience.