



Chemical Concentrations in Newly Constructed Homes

Will Wu, Benjamin He, Tom Xu, Emily Kim, Fred Peng



Recap From Previous Meeting



- Data imputation was required due to the dataset having small sample size and missing data
- Several imputation methods were examined
- The ability to capture original trend was key indicator of model performance
- Joint Modelling scored the highest, therefore it was used to impute missing data

Method	Coverage #1(%)	Coverage #2(%)
Cross Hot Deck	95.6	80.5
Joint modelling	93.9	93.8
Interpolation	75	38
Copymean	98.9	42.5
FCS_LMM	73	8.9

Linear Mixed Model

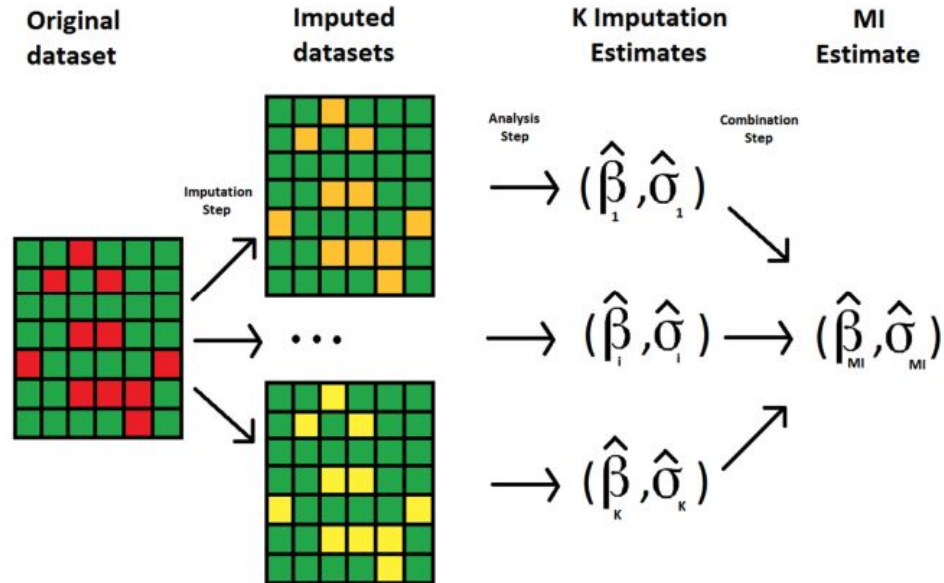
$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_j + \epsilon_{ij}$$

Where:

- Response is chemical concentration
- Fixed effect is Time Period
- With an additional random intercept for different houses
- We will be calling this as Base model

Imputation Methods

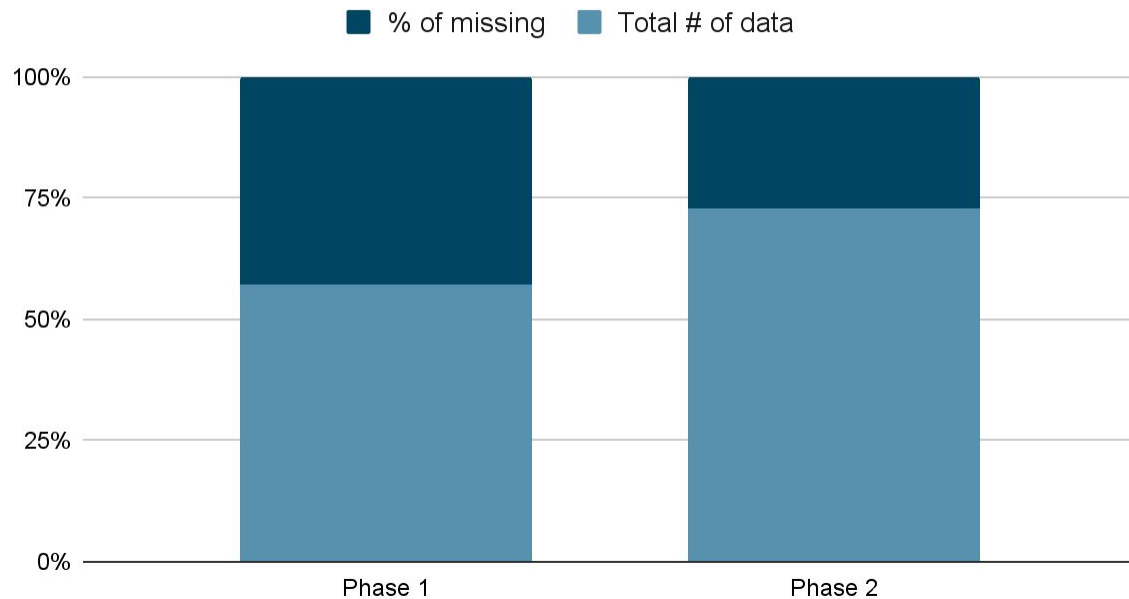
Multiple imputation: Joint Modelling



Imputation



Data Missingness



Imputation model: `lmer(DEP ~ Period + (1|House_ID))`

Model fitting

- After imputation, we obtained three datasets :
 - Phase 1 imputed data (14 houses)
 - Phase 2 imputed data (33 houses)
 - Phase 2 complete case data (15 houses)
- With each dataset, models were fitted and compared with addition of following covariates:
 - low_VOC_home : homes built with low volatile material (don't know : -7 / no : 0 / yes : 1)
 - Paints_yn : paints used inside home during monitoring period (no : 0 / yes : 1)
 - Base_ac_yn : homes with air conditioning (no : 0 / yes : 1)

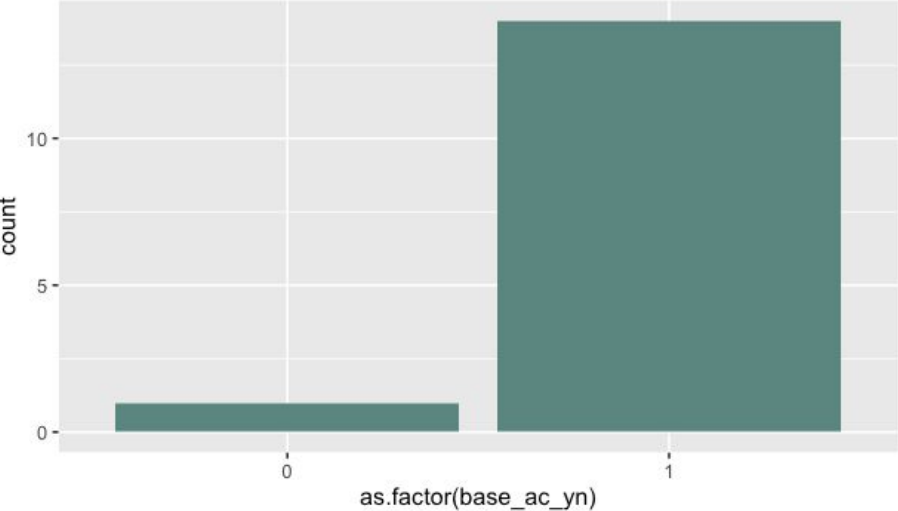


Model comparison in each dataset

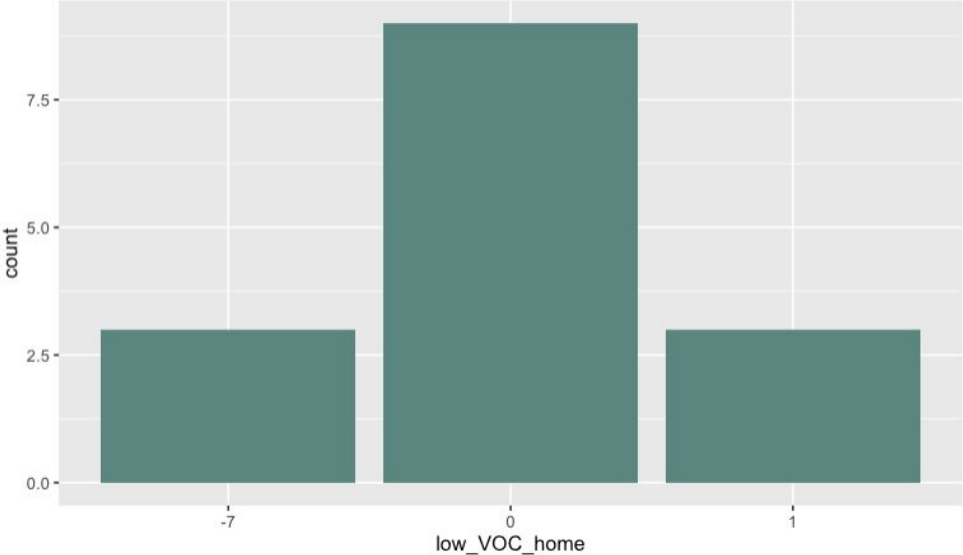
- In Phase 1 and Phase 2 imputed data sets, the model with Period was significantly better than including other covariates.
- In Phase 2 complete data, the model including low_VOC_home and base_ac_yn was significantly better in DiBP concentration
 - In other chemical concentrations, the model with Period was better



Count of base_ac_yn in phase 2 complete data



Count of low_VOC_home in phase 2 complete data





Looking at complete phase 2 data and model

Modelling DiBP			
predictors	Estimate	Std.Error	pvalue
(Intercept)	4.5	1.2	0.00028
Period	0.40	0.071	0.00000075
low_VOC_home0	3.3	0.98	0.0013
low_VOC_home1	-0.69	0.63	0.28
base_ac_yn	-2.0	0.76	0.013

- The low_VOC_home and base_ac_yn had significant effect in the model
- Since we used log-transformation
 - Period : $\exp(0.4) = 1.49$
 - low_VOC_home0: $\exp(3.3) = 27.1$
 - Base_ac_yn: $\exp(2.0) = 7.4$

Fitting model with time effect



Model results of phase1 imputed data

Period estimates			
Chemicals	Estimate	Std.Error	pvalue
DEP	0.039	0.037	0.30
DPP	0.061	0.044	0.18
DiBP	0.049	0.041	0.25
DnBP	0.048	0.061	0.45
BzBP	-0.031	0.039	0.43
DEHP	0.014	0.036	0.71
DnOP	0.16	0.090	0.10
DiNP	0.32	0.078	3.92×10^{-4}

Model results of phase 2 imputed data

Period estimates			
Chemicals	Estimate	Std.Error	pvalue
DEP	0.33	0.050	2.47×10^{-8}
DPP	0.055	0.029	0.063
DiBP	0.28	0.051	1.68×10^{-7}
DnBP	0.31	0.052	4.68×10^{-8}
BzBP	0.056	0.037	0.14
DEHP	0.026	0.033	0.43
DnOP	0.098	0.057	0.092
DiNP	0.10	0.045	0.027

Model results of phase 2 complete data

Period estimates			
Chemicals	Estimate	Std.Error	pvalue
DEP	0.33	0.044	3.56×10^{-9}
DPP	0.060	0.030	0.050
DiBP	0.40	0.072	1.84×10^{-6}
DnBP	0.39	0.069	1.17×10^{-6}
BzBP	0.11	0.047	0.022
DEHP	0.014	0.039	0.73
DnOP	0.14	0.066	0.037
DiNP	0.12	0.048	0.013

Key Discoveries



- For most chemicals, there is evidence supporting that their concentrations increase over time
- After look into the survey data, very few statistical evidence suggests that factors in the survey is influencing the chemical concentrations.

Limitations

- Small sample size and large amount of missing data
- Assumes that houses are independent



Future Studies

- Check if there's statistical evidence of linear trends for other untested chemicals in our dataset
- Consider other factors in technical/survey dataset
- Examine independency between houses



Code

Github link: <https://github.com/DaihaoWu/STA490-Group-5/blob/main/README.md>



Acknowledgement

We wish to acknowledge Dr.Diamond and Sara Vaezafshar for the clarifications and preparation of the dataset, and your help and support throughout the term!

Professor Vianey Leos Barajas for the guidance of the work throughout the year.

Yovna Junglee for the suggestions regarding methodologies and many help during the research process.

We thank you all!



Thank You!

