Demonstration of the final exam for $PM_{2.5}$ exposure-health evaluation and exposure factors

Presenter: Dr. Ming-Chien Mark Tsou

Research Center for Environmental Changes

Academia Sinica

Date: October 13, 2021

Objective of the final exam

- To confirm participants can use the R codes to analysis the data for quantification of $PM_{2.5}$ exposure-health relationships and $PM_{2.5}$ exposure factors
- A total of 4 questions with different setting of scenarios in this exam
  - 2 for $PM_{2.5}$ exposure-health evaluations (Section 1)
  - 2 for $PM_{2.5}$ exposure factors (Section 2)
- The codes which have to modify are indicated in the red box
  - For example:

```
# select no-raining and awake period
PMfinal_A_NR <- PMfinal[which(PMfinal$Precp==0 & PMfinal$Sleep5==4 & PMfinal$Loc_In_All==1),]
```

Modified codes

Raw data sets

- **AS-Lung data (PM data)**
  - The AS-Lung data has already finished the QA/QC procedures by Python and PyCharm.
- **Rooti data (HRV data)**
  - The Rooti data is provided by Rooti company without any process.
- **Questionnaire and time-activity diary data**
- **Meteorological data**

Section 1:
Data analysis on quantification of PM exposure-health evaluation

Section 1

x (the independent variable)	y (the dependent variable)	the added scenario (condition)

Q1. Please evaluate the effects of **personal $PM_{2.5}$ exposure** on **LF/HF ratio** (ratio of low frequency to high frequency) under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments including the indoor transportation modes** (hint: the "Loc_In_All" variable = "1" means subjects were in the indoor environments including the indoor transportation)

Q2. Please evaluate the effects of **personal PM_1 exposure** on **LF** (low frequency power) under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments which do not included the periods in the indoor transportation modes** (hint: the "Loc_In" variable = "1" means subjects were in the indoor environments which is not included the periods in the indoor transportation mode)

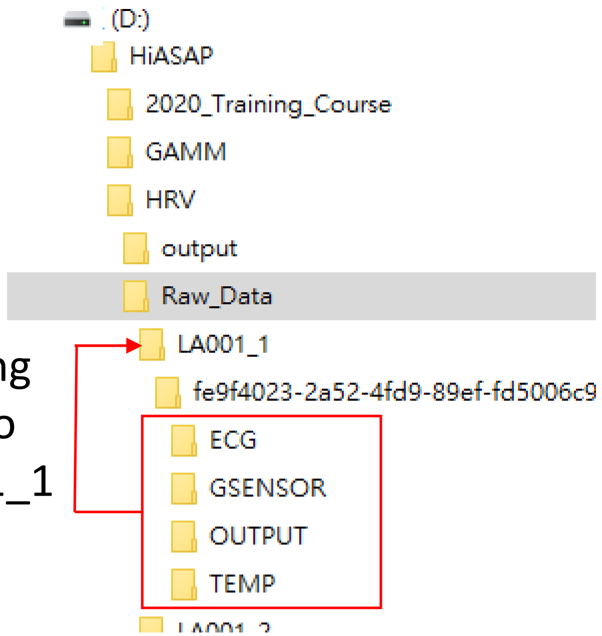
-
- Please provide the output of the models and interpretation the results of models
  - **Quantified effects of $PM_{2.5}$ or PM_1** on HRV indices (expressed as percent changes by interquartile range (IQR) changes) with 95% confidence intervals
  - ***p* values**
  - **Adjusted R^2** of the models.

Q1

Q1. Please evaluate the effects of **personal $PM_{2.5}$ exposure** on **LF/HF ratio** under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments including the indoor transportation modes**

- Step 1: QA/QC of heart rate variability (HRV) data
  - Moving all data files to its own file for every subject
  - Open the R codes for QA/QC of heart rate variability (HRV) (“2020_Trainging_Rooti_HRV.R”) and run
- Step 2: PM data processing
  - Open the R codes for PM data processing for GAMM (“2020_Trainging_PM_for_GAMM.R”) and run
- Step 3: Questionnaire/time-activity diary (TAD) data processing
  - Open the R codes for Questionnaire/time-activity diary (TAD) data processing (“2020_Trainging_Questionnaire_TAD_v2.R”) and run

Moving files to LA001_1



Q1

Q1. Please evaluate the effects of **personal $PM_{2.5}$ exposure** on **LF/HF ratio** under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments including the indoor transportation modes**

- Step 4: Open the R codes for the Generalized Additive Mixed Model (GAMM) ("2020_Trainging_GAMM_v2.R")

- Step 5: Add the selective condition for the new scenario at Line 81 (periods in the indoor environments including the indoor transportation modes)

```
81  PMfinal_A_NR <- PMfinal[which(PMfinal$Precp==0 & PMfinal$Sleep5==4 & PMfinal$Loc_In_All==1),]
```

Q1

Q1. Please evaluate the effects of **personal $PM_{2.5}$ exposure** on **LF/HF ratio** under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments including the indoor transportation modes**

- Step 6: The ““Loc_Out” variable (the variable for the location (outdoor/indoor)) should be removed from the models since we only focus on the periods in the indoor environments including the indoor transportation modes

Before

```
# To run the GAMM for each HRV indices
lg_SDNN<-gamm(lg_SDNN5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(
lg_LFHF<-gamm(lg_LFHF5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(
lg_HRsum<-gamm(lg_HRsum5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=lis
lg_HRmean<-gamm(lg_HRmean5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=l
lg_RMSSD<-gamm(lg_RMSSD5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=lis
lg_LF<-gamm(lg_LF5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_nc
lg_HF<-gamm(lg_HF5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_nc
lg_VLF<-gamm(lg_VLF5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_
lg_TP<-gamm(lg_TP5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_nc
```

After Remove the “Loc_Out” variable from the models

```
# To run the GAMM for each HRV indices
lg_SDNN<-gamm(lg_SDNN5~PM2.5+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_no=~1
lg_LFHF<-gamm(lg_LFHF5~PM2.5+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_no=~1
lg_HRsum<-gamm(lg_HRsum5~PM2.5+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_no=
lg_HRmean<-gamm(lg_HRmean5~PM2.5+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_n
lg_RMSSD<-gamm(lg_RMSSD5~PM2.5+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_no=
lg_LF<-gamm(lg_LF5~PM2.5+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_no=~1),co
lg_HF<-gamm(lg_HF5~PM2.5+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_no=~1),co
lg_VLF<-gamm(lg_VLF5~PM2.5+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_no=~1),
lg_TP<-gamm(lg_TP5~PM2.5+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=list(S_no=~1),co
```

Q1

Q1. Please evaluate the effects of **personal $PM_{2.5}$ exposure** on **LF/HF ratio** under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments including the indoor transportation modes**

- If you do not remove the “Loc_Out” variable, you will see this Error massage

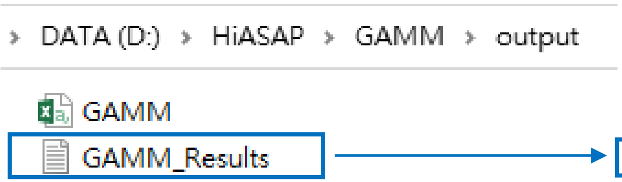
```
> lg_SDNN<-gamm(lg_SDNN5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfi
nal_A_NR,random=list(S_no=~1),correlation=corCAR1(form=~Time|S_no_Day))
Error in MEestimate(lmeSt, grps) :
  Singularity in backsolve at level 0, block 1
>
```

There is only “0” in this variable, which means it only includes data when subjects stayed indoors

Q1

Q1. Please evaluate the effects of **personal $PM_{2.5}$ exposure** on **LF/HF ratio** under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments including the indoor transportation modes**

- Step 7: run the R codes
- Step 8: Open the text file of GAMM results and find the results for LH/HF




```
Family: gaussian
Link function: identity

Formula:
lg_LFHF5 ~ PM2.5 + Season + Age_G + BMI_G + s(Activitymean, bs = c("tp")) +
    Gender + TEM + s(Time, bs = c("cc"))

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.0742264  0.1996909   5.379 8.02e-08 ***
PM2.5        0.0017243  0.0006537   2.638  0.00839 **
Season      -0.1741421  0.0342141  -5.090 3.80e-07 ***
Age_G       -0.4782042  0.1016338  -4.705 2.65e-06 ***
BMI_G        0.1051327  0.0504808   2.083  0.03737 *
Gender       0.1296601  0.0939994   1.379  0.16788
TEM         -0.0215323  0.0067797  -3.176  0.00151 **
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

Approximate significance of smooth terms:
                  edf Ref.df     F  p-value
s(Activitymean) 3.861  3.861 9.608 1.54e-07 ***
s(Time)         6.399  8.000 8.930 2.04e-14 ***
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

R-sq.(adj) =  0.421
  Scale est. = 0.088452   n = 3142
```

Q1

Q1. Please evaluate the effects of **personal $PM_{2.5}$ exposure** on **LF/HF ratio** under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments including the indoor transportation modes**

- Step 9: Calculate the percent changes by interquartile range (IQR) changes ($[10^{(\beta*IQR)}-1]*100\%$) with 95% confidence intervals (CI)

  ($[10^{((\beta\pm1.96*\text{standard error})*IQR)}-1]*100\%$)

β standard error

```
Parametric coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.0742264  0.1996909   5.379 8.02e-08 ***
PM2.5        0.0017243  0.0006537   2.638  0.00839 **
Season      -0.1741421  0.0342141  -5.090 3.80e-07 ***
```

75th percentile – 25th percentile

IQR	B*IQR	10^(B*IQR)-1*100
8.29129	0.014296671	3.346713748

Percent changes

B+1.96*SE	B-1.96*SE	(B+1.96*SE)*IQR	(B-1.96*SE)*IQR	((10^((B+1.96*SE)*IQR)-1))*100	((10^((B-1.96*SE)*IQR)-1))*100
0.003005552	0.000443048	0.024919903	0.003673439	5.905838545	0.849428032

95% CI

	$PM_{2.5}$ (adjusted R^2 = 0.421)		
HRV indices	Percentage change	95% CI	*p*-value
LF/HF	3.35	0.85, 5.91	0.008

an increase in $PM_{2.5}$ concentrations of IQR (8.3 $\mu g/m^3$) was associated with a change of 3.35% in LF/HF

Q2

Q2. Please evaluate the effects of **personal PM_1 exposure** on **LF** under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments which do not included the periods in the indoor transportation modes**

- Step 1: QA/QC of heart rate variability (HRV) data
- Step 2: PM data processing
- Step 3: Questionnaire/time-activity diary (TAD) data processing
- Step 4: Open the R code for the Generalized Additive Mixed Model (GAMM)

Step 1 to Step 4 are the same as those in Q1

- Step 5: Add the selective condition for the new scenario at Line 81 (periods in the indoor environments which is not included the periods in the indoor transportation modes)

```
81  PMfinal_A_NR <- PMfinal[which(PMfinal$Precp==0 & PMfinal$Sleep5==4 & PMfinal$Loc_In==1),]
```

Q2

Q2. Please evaluate the effects of **personal PM_1 exposure** on **LF** under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments which do not included the periods in the indoor transportation modes**

- Step 6: The ""Loc_Out" variable (the variable for the location (outdoor/indoor)) should be removed from the models since we only focus on the periods in the indoor environments
- Step 7: Replace PM2.5 by PM1

Before

Remove the "Loc_Out" variable from the models

```
# To run the GAMM for each HRV indices
lg_SDNN<-gamm(lg_SDNN5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfin
lg_LFHF<-gamm(lg_LFHF5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfin
lg_HRsum<-gamm(lg_HRsum5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMf
lg_HRmean<-gamm(lg_HRmean5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=P
lg_RMSSD<-gamm(lg_RMSSD5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMf
lg_LF<-gamm(lg_LF5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A
lg_HF<-gamm(lg_HF5~PM2.5+Loc_Out+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A
```

After

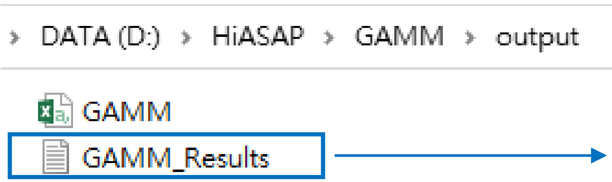
```
# To run the GAMM for each HRV indices
lg_SDNN<-gamm(lg_SDNN5~PM1+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,ranc
lg_LFHF<-gamm(lg_LFHF5~PM1+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,ranc
lg_HRsum<-gamm(lg_HRsum5~PM1+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,ra
lg_HRmean<-gamm(lg_HRmean5~PM1+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,
lg_RMSSD<-gamm(lg_RMSSD5~PM1+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,ra
lg_LF<-gamm(lg_LF5~PM1+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=
lg_HF<-gamm(lg_HF5~PM1+Season+Age_G+BMI_G+s(Activitymean,bs=c("tp"))+Gender+TEM+s(Time,bs=c("cc")),data=PMfinal_A_NR,random=
```

Replace PM2.5 by PM1

Q2

Q2. Please evaluate the effects of **personal PM_1 exposure** on **LF** under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments which do not included the periods in the indoor transportation modes**

- Step 8: run the R codes
- Step 9: Open the text file of GAMM results and find the results for LH




```
Family: gaussian
Link function: identity

Formula:
lg_LF5 ~ PM1 + Season + Age_G + BMI_G + s(Activitymean, bs = c("tp")) +
    Gender + TEM + s(Time, bs = c("cc"))

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.2412303  0.2460926  13.171  < 2e-16 ***
PM1          0.0018386  0.0007835   2.347 0.019008 *
Season      -0.0471197  0.0383016  -1.230 0.218708
Age_G       -0.8108690  0.2278069  -3.559 0.000377 ***
BMI_G        0.0837720  0.0551473   1.519 0.128853
Gender      -0.1345752  0.1878787  -0.716 0.473869
TEM         -0.0132411  0.0076468  -1.732 0.083448 .
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

Approximate significance of smooth terms:
                  edf Ref.df     F  p-value
s(Activitymean) 3.188  3.188 9.589 1.21e-06 ***
s(Time)         3.747  8.000 4.698 5.57e-09 ***
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

R-sq.(adj) =  0.365
  Scale est. = 0.1046    n = 3033
```

Q2

Q2. Please evaluate the effects of **personal PM_1 exposure** on **LF** under the following scenarios: (1) **no-raining periods**, (2) **awake periods**, and (3) **periods in the indoor environments which do not included the periods in the indoor transportation modes**

- Step 10: Calculate the percent changes by interquartile range (IQR) changes ($[10^{(\beta*IQR)}-1]*100\%$) with 95% confidence intervals (CI)

($[10^{((\beta\pm1.96*\text{standard error})*IQR)}-1]*100\%$)

β standard error

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	3.2412303	0.2460926	13.171	< 2e-16	***
PM1	$\boxed{0.0018386}$	$\boxed{0.0007835}$	2.347	0.019008	*
Season	-0.0471197	0.0383016	-1.230	0.218708	

75th percentile – 25th percentile

IQR	B*IQR	10^(B*IQR)-1*100
7.6379	0.014043043	3.286376785

Percent changes

B+1.96*SE	B-1.96*SE	(B+1.96*SE)*IQR	(B-1.96*SE)*IQR	((10^((B+1.96*SE)*IQR)-1))*100	((10^((B-1.96*SE)*IQR)-1))*100
0.00337426	0.00030294	0.02577226	0.002313825	6.113896095	0.534199779

95% CI

	$PM_{2.5}$ (adjusted R^2 = 0.365)		
HRV indices	Percentage change	95% CI	*p*-value
LF	3.29	0.53, 6.11	0.019

an increase in PM_1 concentrations of IQR (7.6 $\mu g/m^3$) was associated with a change of 3.29% in LF

Section 2:
Data analysis on PM exposure factors with environmental and survey data

Section 2

x (the independent variable)	y (the dependent variables)	the scenarios (conditions)

Q3. Please evaluate the **personal $PM_{2.5}$ exposure factors** including **outdoor $PM_{2.5}$ concentration**, **indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

Q4. Please evaluate the **indoor PM_1 exposure factors** including **outdoor PM_1 concentration**, **indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

-
- Please provide the output of the stepwise procedures, models and partial R^2 of each independent variables, and interpretation the results of models
  - **Incremental contribution** to $PM_{2.5}$ or PM_1 of these indoor sources, ventilations and outdoor PM concentration with 95% confidence intervals
  - ***p* values**
  - **Partial R^2 values**
  - **Adjusted R^2** of the models.
- Please also state which variable has **the greatest contribution** to $PM_{2.5}$ or PM_1.

Section 2

- Prior to analyzing the data, please modified the definition of "S_Other2" group in the R codes for TAD data (2020_Trainging_Questionnaire_TAD_v2.R). The "S_Other2" group should include the sources of dust/clean, environmental tobacco smoke (ETS), mosquito coils, aromatic products, open burning, odor of garbage, and other sources. **(from Line 121 to 127)**

Put the sources of ETS into the "S_Other2" group

```
121 S_Other2<-c()
122 for(i in 1:dim(TADr)[1]){
123     if(S_ETS[i]==1|TADr$S_Dust[i]==1|TADr$S_MosquitoCoil[i]==1|TADr$S_Aromatic[i]==1|TADr$S_OpenedBurning[i]==1|TADr$S_Factory[i]==1|TADr$S_Garbage[i]==1|TADr$S_Other[i]==1){
124         S_Other2[i]<-1
125     }else{
126         S_Other2[i]<-0
127     }
128 }
```

Q3

Q3. Please evaluate the **personal $PM_{2.5}$ exposure factors** including **outdoor $PM_{2.5}$ concentration, indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

- Step 1: PM data processing
  - Open the R codes for PM data processing for exposure factor evaluations ("2020_Trainging_PM_for_Exposure_Factor.R") and run
- Step 2: Questionnaire/time-activity diary (TAD) data processing
  - Open the R codes for Questionnaire/time-activity diary (TAD) data processing ("2020_Trainging_Questionnaire_TAD_v2.R") and run
- Step 3: Open the R codes for exposure factor evaluations ("2020_Trainging_Exposure_Factor_v3.R")

Q3

Q3. Please evaluate the **personal $PM_{2.5}$ exposure factors** including **outdoor $PM_{2.5}$ concentration, indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

- Step 4: Change the selective condition for the new scenarios in Line 42
  - **No-raining periods**
  - **At-home periods**
  - **Periods between 06:00 to 24:00**

```
42  PMfinal_A_NR_Home <- PMfinal[which(PMfinal$Precp==0 & PMfinal$Loc_Home==1 & PMfinal$Hour>=6 & PMfinal$Hour<=24),]
```

Q3

Q3. Please evaluate the **personal $PM_{2.5}$ exposure factors** including **outdoor $PM_{2.5}$ concentration, indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

- Step 5: Select the variables which input into the stepwise regression model in Line 46 (do not select the "S_ETS" variable since it has been include in the "S_Other2" group)

Before

```
44 # To determine a final regression model by the stepwise regression method
45 # To remain the variables which you are interested
46 aa <- PMfinal_A_NR_Home[,c(9,13,23,50,51,52,62,63,66,67)]
```

After

```
46 aa <- PMfinal_A_NR_Home[,c(9,13,23,50,51,62,63,66,67)]
```

The column 52 is the "S_ETS" variable, so we do not include this variable in this exam

	Season	ALP_PM2.5	ALO_PM2.5	S_Exhaust	S_Cooking	S_Incense_JPaper	S_Other2	Vent_D3_1	Vent_D3_2
1	0	11.53100	11.4748	0	0	0	0	0	0
2	0	13.27900	12.0708	0	0	0	0	0	0
3	0	14.63600	9.5080	0	0	0	0	0	0
4	0	14.38300	9.9252	0	0	0	0	0	0
5	0	18.15500	10.5212	0	0	0	0	0	0
6	0	14.29100	10.1040	0	0	0	0	0	0
7	0	16.43000	11.3556	0	1	0	0	0	0
8	0	18.82200	9.8060	0	1	0	0	0	0
9	0	15.07300	10.7000	0	1	0	0	0	0

Q3

Q3. Please evaluate the **personal $PM_{2.5}$ exposure factors** including **outdoor $PM_{2.5}$ concentration, indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

- Step 6: Run the R codes for stepwise regression methods (in Line 50 to 51)

```
[1] "Stepwise"
Start:  AIC=8951.41
ALP_PM2.5 ~ Season + ALO_PM2.5 + S_Exhaust + S_Cooking + S_Incense_JPaper +
    S_Other2 + Vent_D3_1 + Vent_D3_2

                   Df Sum of Sq    RSS    AIC
- S_Exhaust         1        11 129982 8949.6
- Vent_D3_2         1        15 129986 8949.7
- Vent_D3_1         1        17 129988 8949.7
<none>                          129971 8951.4
- S_Other2          1       253 130224 8953.7
- Season            1      4599 134570 9025.5
- S_Cooking         1     15433 145404 9194.8
- S_Incense_JPaper  1     33219 163189 9447.2
- ALO_PM2.5         1     57331 187302 9748.6

Step:  AIC=8949.6
ALP_PM2.5 ~ Season + ALO_PM2.5 + S_Cooking + S_Incense_JPaper +
    S_Other2 + Vent_D3_1 + Vent_D3_2

                   Df Sum of Sq    RSS    AIC
- Vent_D3_2         1        16 129998 8947.9
- Vent_D3_1         1        19 130002 8947.9
<none>                          129982 8949.6
+ S_Exhaust         1        11 129971 8951.4
- S_Other2          1       244 130226 8951.7
- Season            1      4588 134570 9023.5
- S_Cooking         1     15685 145667 9196.8
- S_Incense_JPaper  1     33584 163566 9450.2
- ALO_PM2.5         1     58961 188943 9765.6
```

```
Step:  AIC=8947.88
ALP_PM2.5 ~ Season + ALO_PM2.5 + S_Cooking + S_Incense_JPaper +
    S_Other2 + Vent_D3_1

                   Df Sum of Sq    RSS    AIC
- Vent_D3_1         1        18 130016 8946.2
<none>                          129998 8947.9
+ Vent_D3_2         1        16 129982 8949.6
+ S_Exhaust         1        13 129986 8949.7
- S_Other2          1       250 130248 8950.1
- Season            1      4748 134746 9024.3
- S_Cooking         1     15679 145678 9194.9
- S_Incense_JPaper  1     33610 163609 9448.8
- ALO_PM2.5         1     59296 189294 9767.7

Step:  AIC=8946.17
ALP_PM2.5 ~ Season + ALO_PM2.5 + S_Cooking + S_Incense_JPaper +
    S_Other2
```

Final model

```
                   Df Sum of Sq    RSS    AIC
<none>                          130016 8946.2
+ Vent_D3_1         1        18 129998 8947.9
+ S_Exhaust         1        15 130001 8947.9
+ Vent_D3_2         1        14 130002 8947.9
- S_Other2          1       255 130271 8948.5
- Season            1      5074 135090 9027.9
- S_Cooking         1     15662 145678 9192.9
- S_Incense_JPaper  1     33656 163672 9447.6
- ALO_PM2.5         1     60301 190317 9777.5
```

Q3

Q3. Please evaluate the **personal $PM_{2.5}$ exposure factors** including **outdoor $PM_{2.5}$ concentration, indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

- Step 6: Run the R codes for stepwise regression methods (in Line 50 to 51)

Final model

```
Step:  AIC=8946.17
ALP_PM2.5 ~ Season + ALO_PM2.5 + S_Cooking + S_Incense_JPaper +
    S_Other2

                   Df Sum of Sq    RSS    AIC
<none>                          130016 8946.2
+ Vent_D3_1         1        18 129998 8947.9
+ S_Exhaust         1        15 130001 8947.9
+ Vent_D3_2         1        14 130002 8947.9
- S_Other2          1       255 130271 8948.5
- Season            1      5074 135090 9027.9
- S_Cooking         1     15662 145678 9192.9
- S_Incense_JPaper  1     33656 163672 9447.6
- ALO_PM2.5         1     60301 190317 9777.5
```

- In this case, there is no issue about the dummy variables in the final model obtained from the stepwise method
- You can skip Line 53 to 56

Skip Line 53 to 56

```
48 # To use the stepwise regression method to identify and select the most useful explanatory variables from a list of
49 # several plausible independent variables
50 Final_Model <- step(lm(ALP_PM2.5~.,data=aa),direction="both")
51 summary(Final_Model)
52
53 # In order to explain the set of dummy variables of trinary variable (window open, window closed and AC-on),
54 # we add the dummy variable of "Vent_D3_2" into the final model
55 Final_Model_2 <- lm(ALP_PM2.5 ~ ALO_PM2.5+S_Incense_JPaper+S_Other2+Vent_D3_1+Vent_D3_2+Season, data = aa)
56 summary(Final_Model_2)
```

Q3

Q3. Please evaluate the **personal $PM_{2.5}$ exposure factors** including **outdoor $PM_{2.5}$ concentration, indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

- Step 7: Run the remained R codes

```r
# To calculate 95% confidence interval
CI <- confint(Final_Model,level=0.95)

# To determine the partial R2 of each independent variable
# install.packages("rsq") (only for first time to use)
library("rsq")

Partial_R2 <- rsq.partial(Final_Model)

# To print out final model results to txt file
sink("Exposure_Factor_Results.txt") # redirect console output to a
print("Stepwise")
print(step(lm(ALP_PM2.5~.,data=aa),direction="both"))
print("Final model")
print(summary(Final_Model))
print("95% CI")
print(CI)
print("Partial R2")
print(Partial_R2)
sink()  # return output to the terminal
```

```
[1] "95% CI"
                      2.5 %      97.5 %
(Intercept)       1.1304395  2.45835440
Season           -4.3850830 -2.84763238
ALO_PM2.5         0.6119106  0.69232868
S_Cooking         7.5078081  9.57451514
S_Incense_JPaper 20.9072297 24.66875793
S_Other2         -3.6187682 -0.09596588
```

95% confidence intervals

```
[1] "Final model"

Call:
lm(formula = ALP_PM2.5 ~ Season + ALO_PM2.5 + S_Cooking + S_Incense_JPaper +
    S_Other2, data = aa)

Residuals:
    Min      1Q  Median      3Q     Max
-27.539  -3.185  -1.103   2.475  87.957

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.7944     0.3386   5.300 1.28e-07 ***
Season            -3.6164     0.3920  -9.225  < 2e-16 ***
ALO_PM2.5          0.6521     0.0205  31.805  < 2e-16 ***
S_Cooking          8.5412     0.5269  16.209  < 2e-16 ***
S_Incense_JPaper  22.7880     0.9591  23.761  < 2e-16 ***
S_Other2          -1.8574     0.8982  -2.068   0.0388 *
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 7.721 on 2181 degrees of freedom
Multiple R-squared:  0.5327,    Adjusted R-squared:  0.5317
F-statistic: 497.3 on 5 and 2181 DF,  p-value: < 2.2e-16
```

β (coefficients)

p value

Adjusted R^2

```
[1] "Partial R2"
$adjustment
[1] FALSE

$variable
[1] "Season"           "ALO_PM2.5"        "S_Cooking"        "S_Incense_JPaper" "S_Other2"

$partial.rsq
[1] 0.03755760 0.31684614 0.10751292 0.20563096 0.00195682
```

Partial R^2

Q3

Q3. Please evaluate the **personal $PM_{2.5}$ exposure factors** including **outdoor $PM_{2.5}$ concentration, indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

	Personal $PM_{2.5}$ exposure (μg/m^3) (adjusted R^2 = 0.531)			
Variables	Coefficient	95% CI	Partial R^2	*p*-value
Outdoor $PM_{2.5}$ concentration (μg/m^3)	0.652	0.612, 0.692	0.317	<0.001
Burning of incense/joss-paper	22.8	20.9, 24.7	0.206	<0.001
Cooking	8.54	7.51, 9.57	0.108	<0.001
Other sources	-1.86	-3.62, -0.0960	0.002	0.039
Season	-3.62	-4.39, -2.85	0.038	<0.001

- Burning of incense/joss-paper had on average the highest 5-min $PM_{2.5}$ increments (22.8 μg/m^3) to personal $PM_{2.5}$ exposure.

Q4

Q4. Please evaluate the **indoor PM_1 exposure factors** including **outdoor PM_1 concentration**, **indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

- Step 1: PM data processing
- Step 2: Questionnaire/time-activity diary (TAD) data processing
- Step 3: Open the R codes for exposure factor evaluations ("2020_Trainging_Exposure_Factor_v3.R")
- Step 4: Change the selective condition for the new scenarios in Line 42

```
42  PMfinal_A_NR_Home <- PMfinal[which(PMfinal$Precp==0 & PMfinal$Loc_Home==1 & PMfinal$Hour>=6 & PMfinal$Hour<=24),]
```

Step 1 to Step 4 are the same as those in Q3

Q4

Q4. Please evaluate the **indoor PM_1 exposure factors** including **outdoor PM_1 concentration**, **indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

- Step 5: Select the variables which input into the stepwise regression model in Line 46
  - Do not select the “S_ETS” variable
  - Select the “ALI_PM1”(indoor PM1) and “ALO_PM1” (outdoor PM1) variables

```
44 # To determine a final regression model by the stepwise regression method
45 # To remain the variables which you are interested
46 aa <- PMfinal_A_NR_Home[,c(9,17,22,50,51,62,63,66,67)]
```

	Season	ALI_PM1	ALO_PM1	S_Exhaust	S_Cooking	S_Incense_JPaper	S_Other2	Vent_D3_1	Vent_D3_2
1	0	11.27740	10.8126	0	0	0	0	0	0
2	0	11.11025	11.5218	0	0	0	0	0	0
3	0	11.36255	8.7835	0	0	0	0	0	0
4	0	11.45805	9.3942	0	0	0	0	0	0
5	0	11.01290	9.7882	0	0	0	0	0	0
6	0	11.54390	9.7094	0	0	0	0	0	0
7	0	11.99355	10.4186	0	1	0	0	0	0
8	0	12.59695	9.1578	0	1	0	0	0	0
9	0	12.91025	10.2610	0	1	0	0	0	0

	Vent_D1	Vent_D2
Window open	0	0
Window closed	1	0
AC-on	0	1

Q4. Please evaluate the **indoor PM_1 exposure factors** including **outdoor PM_1 concentration**, **indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

- Step 6: Run the R codes for stepwise regression methods (in Line 50 to 51)

```
48  # To use the stepwise regression method to identify and select the most useful explanatory variables from a list of
49  # several plausible independent variables
50  Final_Model <- step(lm(ALI_PM1~.,data=aa),direction="both")
51  summary(Final_Model)
```

Be careful. The dependent variable (y) is the indoor PM_1

Put both 2 dummy variables ("Vent_D3_1" and "Vent_D3_2") for ventilations in the final model

Final model obtained from stepwise method

```
Step:  AIC=9620.05
ALI_PM1 ~ Season + ALO_PM1 + S_Cooking + S_Incense_JPaper + S_Other2 +
    Vent_D3_1

                   Df Sum of Sq    RSS     AIC
<none>                          176774  9620.0
- S_Other2          1       270 177043  9621.4
+ S_Exhaust         1        28 176745  9621.7
+ Vent_D3_2         1        15 176759  9621.9
- Vent_D3_1         1       408 177182  9623.1
- Season            1      6789 183562  9700.5
- S_Cooking         1     14157 190930  9786.5
- S_Incense_JPaper  1     41851 218624 10082.7
- ALO_PM1           1     90156 266930 10519.3
```

Only include 1 dummy variable for ventilations

```
[1] "Final model"

Call:
lm(formula = ALI_PM1 ~ ALO_PM1 + S_Cooking + S_Incense_JPaper +
    S_Other2 + Vent_D3_1 + Vent_D3_2 + Season, data = aa)

Residuals:
    Min      1Q  Median      3Q     Max
-26.823  -2.728  -0.304   2.550  97.133

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -0.23068    0.43900  -0.525   0.5993
ALO_PM1           0.94783    0.02848  33.275   <2e-16 ***
S_Cooking         8.15027    0.61682  13.213   <2e-16 ***
S_Incense_JPaper 25.42828    1.12004  22.703   <2e-16 ***
S_Other2         -1.88394    1.04387  -1.805   0.0712 .
Vent_D3_1         1.13757    0.50267   2.263   0.0237 *
Vent_D3_2         0.48800    1.15145   0.424   0.6717
Season           -4.37339    0.48507  -9.016   <2e-16 ***
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 9.007 on 2179 degrees of freedom
Multiple R-squared:  0.5289,    Adjusted R-squared:  0.5273
F-statistic: 349.4 on 7 and 2179 DF,  p-value: < 2.2e-16
```

Q4

Q4. Please evaluate the **indoor PM_1 exposure factors** including **outdoor PM_1 concentration**, **indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

- Step 7: Run the remained R codes

```r
# To calculate 95% confidence interval
CI <- confint(Final_Model_2,level=0.95)

# To determine the partial R2 of each independent variable
# install.packages("rsq") (only for first time to use)
library("rsq")

Partial_R2 <- rsq.partial(Final_Model_2)

# To print out final model results to txt file
sink("Exposure_Factor_Results.txt") # redirect console output to a file
print("Stepwise")
print(step(lm(ALI_PM1~.,data=aa),direction="both"))
print("Final model")
print(summary(Final_Model_2))
print("95% CI")
print(CI)
print("Partial R2")
print(Partial_R2)
sink()  # return output to the terminal
```

```
[1] "95% CI"
                        2.5 %     97.5 %
(Intercept)        -1.0915893  0.6302252
ALO_PM1             0.8919679  1.0036881
S_Cooking           6.9406595  9.3598843
S_Incense_JPaper   23.2318259 27.6247357
S_Other2           -3.9310158  0.1631449
Vent_D3_1           0.1518117  2.1233357
Vent_D3_2          -1.7700682  2.7460596
Season             -5.3246305 -3.4221466
```

95% confidence intervals

```
[1] "Final model"

Call:
lm(formula = ALI_PM1 ~ ALO_PM1 + S_Cooking + S_Incense_JPaper +
    S_Other2 + Vent_D3_1 + Vent_D3_2 + Season, data = aa)

Residuals:
    Min      1Q  Median      3Q     Max
-26.823  -2.728  -0.304   2.550  97.133

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      -0.23068    0.43900  -0.525   0.5993
ALO_PM1           0.94783    0.02848  33.275   <2e-16 ***
S_Cooking         8.15027    0.61682  13.213   <2e-16 ***
S_Incense_JPaper 25.42828    1.12004  22.703   <2e-16 ***
S_Other2         -1.88394    1.04387  -1.805   0.0712 .
Vent_D3_1         1.13757    0.50267   2.263   0.0237 *
Vent_D3_2         0.48800    1.15145   0.424   0.6717
Season           -4.37339    0.48507  -9.016   <2e-16 ***
---
Signif. codes:  0 ?**?0.001 ?*?0.01 ??0.05 ??0.1 ??1

Residual standard error: 9.007 on 2179 degrees of freedom
Multiple R-squared:  0.5289,    Adjusted R-squared:  0.5273
F-statistic: 349.4 on 7 and 2179 DF,  p-value: < 2.2e-16
```

β (coefficients)

p value

Adjusted R^2

```
[1] "Partial R2"
$adjustment
[1] FALSE

$variable
[1] "ALO_PM1"          "S_Cooking"        "S_Incense_JPaper" "S_Other2"         "Vent_D3_1"        "Vent_D3_2"
[7] "Season"

$partial.rsq
[1] 3.369286e-01 7.418206e-02 1.912941e-01 1.492571e-03 2.344860e-03 8.242241e-05 3.596408e-02
```

Partial R^2

Q4

Q4. Please evaluate the **indoor PM_1 exposure factors** including **outdoor PM_1 concentration**, **indoor PM sources (vehicle exhaust, cooking, burning of incense/joss-paper and other sources)** and **ventilations** under the following scenarios: (1) **no-raining periods**, (2) **at-home periods**, and (3) **periods between 06:00 to 24:00**.

	Personal PM_1 exposure (μg/m³) (adjusted R^2 = 0.527)			
Variables	Coefficient	95% CI	Partial R^2	*p*-value
Outdoor PM_1 concentration (μg/m³)	0.948	0.892, 1.00	0.337	<0.001
Burning of incense/joss-paper	25.4	23.2, 27.6	0.191	<0.001
Cooking	8.15	6.94, 9.36	0.074	<0.001
Other sources	-1.88	-3.93, 0.163	0.002	0.0712
Window-closed	1.14	0.152, 2.12	0.002	0.0237
AC-on	0.488	-1.77, 2.75	0.00008	0.6717
Season	-4.37	-5.32, -3.42	0.036	<0.001

- Burning of incense/joss-paper had on average the highest 5-min PM_1 increments (25.4 μg/m³) to indoor PM_1 exposure.

Thank you for your attention