The University of Hong Kong
School of Public Health

CMED 6040 Advanced Statistical Methods II
TUTORIAL 3

Indoor particulate matter and airway macrophages black carbon content

A longitudinal study recruited 200 adults aged 60 to 69y to investigate the relation between indoor $PM_{2.5}$ exposure and black carbon content in airway macrophages (AMBC). Subjects were visited for 5 times in the study. The dataset was saved in 'ambc.csv' with the following variables:

Variable	Data label	Remark
id	Subject identification number	
age	Age in years	
male	Male sex	0: Female; 1: Male
time	Follow-up occasion	
pm	Mean indoor $PM_{2.5}$ concentration ($\mu g/m^3$)	
ambc	Mean area of black carbon in airway macrophages (μm^2)	

(a) Read the dataset into R. Plot the pattern of pattern of indoor $PM_{2.5}$ and AMBC over time. Do you think AMBC are correlated within subjects?

(b) Explore the relationship between indoor $PM_{2.5}$ and AMBC content for some subjects.

(c) Fit an ordinary regression model to predict AMBC by age, male, and $PM_{2.5}$. Summarize your results in the following table:

Variable	***Estimated coefficient***	***95% CI****
Male		
Age		
$PM_{2.5}$		

*CI: confidence interval

(d) Fit a generalized estimating equation (GEE) model to predict AMBC by age, male, and $PM_{2.5}$, by assuming independence, exchangeable, AR(1) and unstructured correlation structure.

(e) Select the best model from (d).

(f) Summarize your results from the selected model in (e) in the following table:

Variable	***Estimated coefficient***	***95% CI****
Male		
Age		
$PM_{2.5}$		

*CI: confidence interval

(g) Fit a linear mixed model with random intercept to predict AMBC by age, male, and $PM_{2.5}$. Summarize your results in the following table:

Variable	*Estimated coefficient*	*95% CI**
Male		
Age		
$PM_{2.5}$		

*CI: confidence interval

(h) Calculate the intraclass correlation (ICC) for the model in (g).

(i) Fit a linear effects model with random intercept and slope to predict AMBC by age, male, and $PM_{2.5}$. Summarize your results in the table below and compare your results with that in (g).

Variable	*Estimated coefficient*	*95% CI**
Male		
Age		
$PM_{2.5}$		

*CI: confidence interval

(j) Compare the results in (a)-(i) from different methods.

(k) Suppose it is also interesting to investigate how the previous mean indoor $PM_{2.5}$ concentrations will affect the AMBC at the 5^{th} visit. Fit a linear regression for the purpose and summarize your results in the table below:

[hint: convert data to wide format]

Variable	*Estimated coefficient*	*95% CI**
Male		
Age		
$PM_{2.5}$ (visit 1)		
$PM_{2.5}$ (visit 2)		
$PM_{2.5}$ (visit 3)		
$PM_{2.5}$ (visit 4)		
$PM_{2.5}$ (visit 5)		

*CI: confidence interval

(l) Perform a LASSO regression and select the optimal λ by cross-validation method.