Particular Matter Exposure

Thomas Jensen and Mitchell Pudil

March 2019

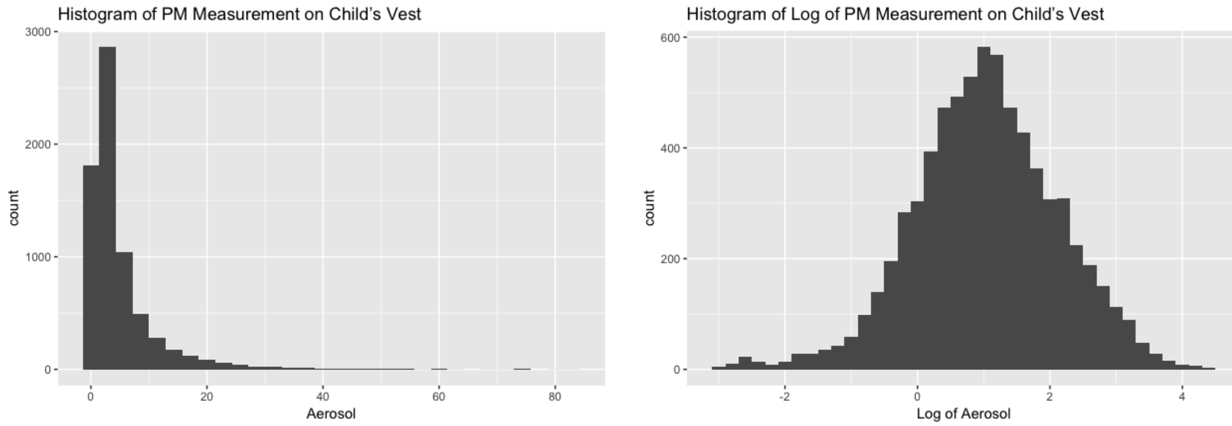
Executive Summary

Can activities and indoor stationary measurements of particulate matter (PM) help explain the actual level of PM inhaled by children? Are these effects child specific? In order to answer these questions, we utilize a generalized least squares (GLS) model with an AR(1) correlation structure. We find that stationary PM readings are a helpful predictor for actual exposure, though ultimately not statistically significant. While our joint hypothesis tests show that activities help predict the level of PM that a child is exposed to, individual hypothesis tests show that the only activity that is statistically different from playing on the computer is playing on the furniture. Additionally, we find that PM exposure indoors does differ from child to child (perhaps better put as house-to-house) and care must be taken to promote lower PM exposure.

1 Introduction

Particulate matter (PM) is defined as a mixture of solid and liquid particles that float around in the air. This complex mixture includes both organic and inorganic particles, such as dust, pollen, soot, smoke, and liquid droplets. Recently, there has been great interest in better understanding the health effects of PM exposure, as PM has been linked to several potentially fatal illnesses. The object of this study was to collect PM data from indoor air samples and determine how the level of PM breathed in is affected by the type of activity performed, if the stationary measurement in the home is a good predictor of the actual amount breathes in, and if any of these effects vary by child. To determine this, researchers recruited 60 children to participate in the study where two PM readings were recorded every minute for nearly two hours. The first PM reading was from a vest that the children wore and the second from a stationary monitor set up in the house.

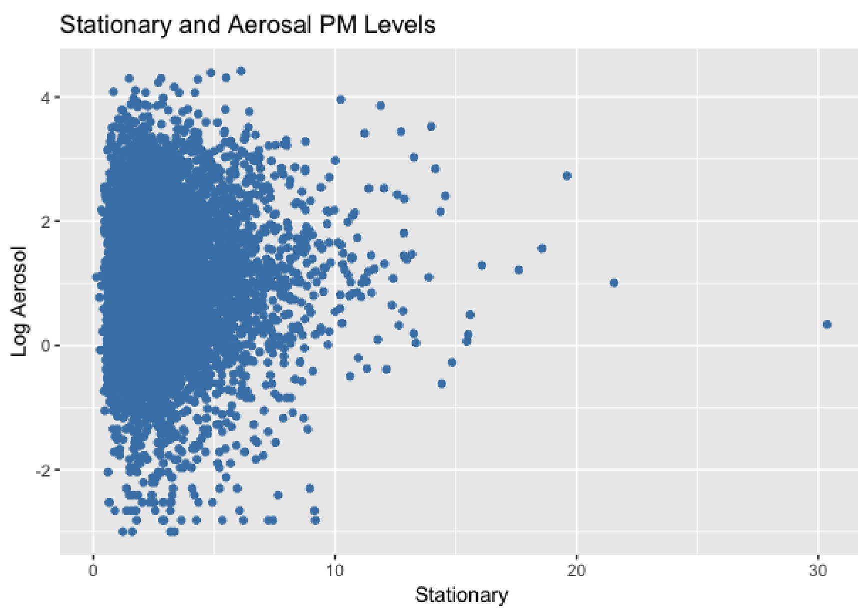
The explanatory variables in the data collected include a unique ID number for each child participant, PM readings from the stationary monitors set up in each of the children's houses, and the activity the child



engaged in for each minute of the nearly two-hour-long study. The variable of interest, the response variable, is the PM reading (Aerosol) near the mouth from the measuring device on the subject's vest.

Upon initial exploration of the data, there did not appear to be a significant relationship between the stationary PM reading and aerosol reading. However, a simple BoxCox statistic on the initial independent model suggested a log transformation of the aerosol reading. Side-by-side histograms of the non-logged aerosol and the logged aerosol are shown above.

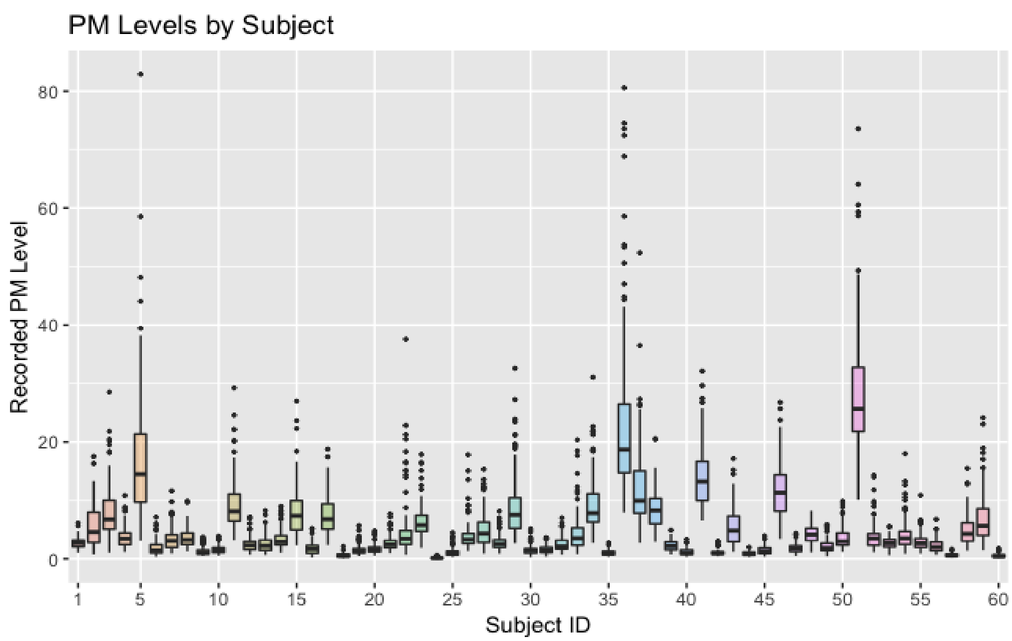
After taking the log transformation of aerosol, we plotted that variable against the stationary measurements. This scatterplot is shown below.



From this graph, there appears to be a weak linear relationship between the stationary and log transformed values of aerosol PM. A possible issue we may have here now presents itself. The spread of stationary readings against the logged values of aerosol readings may not be evenly distributed. There are two potential problems with the data that we will want to consider in building model. The first is determining whether the variance of the data residuals about the fitted (predicted) values is equal. The second is determining whether or not there exists correlation between residual values. If either of these two problems are not accounted for in our model building, the standard errors produces by our model will be inaccurate and predictions will not perform well.

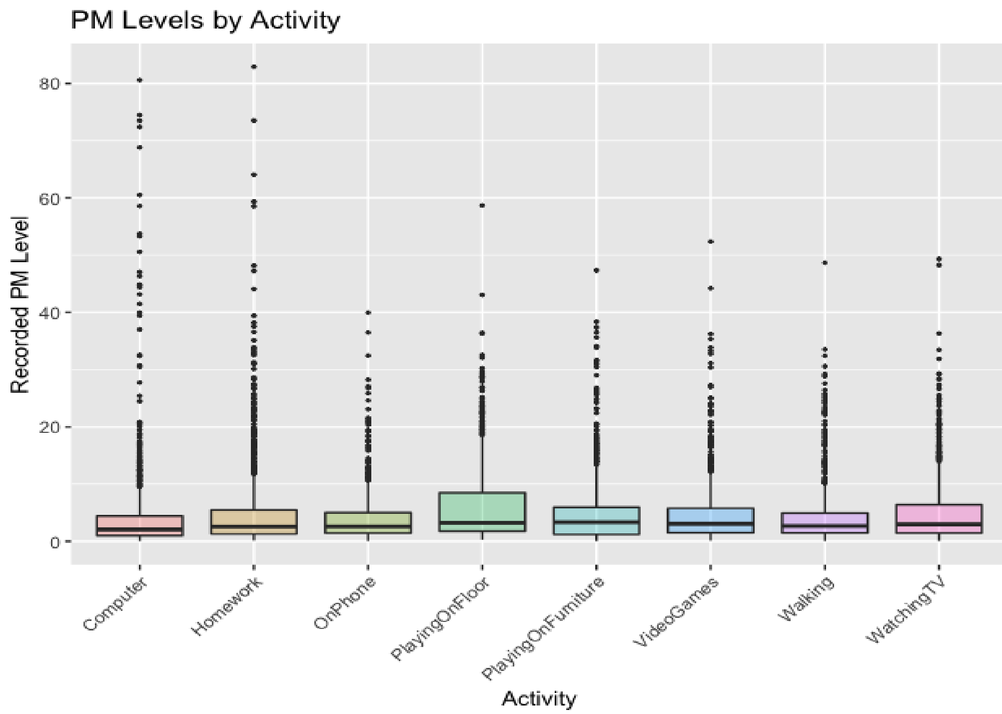
We address the problems of unequal variance and residual correlation in the Model Validation section below.

Additionally, we looked at the level of aerosol each child was exposed to using a side-by-side boxplot with each box representing a different ID. The graph is shown below.



As is evident from the above boxplot, certain subjects including 5, 36, and 51 (to name the most obvious examples) recorded a much larger range of PM readings than the other subjects. A limitation in this study is the fact that researchers were unable to control the initial air purity from house to house of the subjects. Therefore, a noteworthy question we will analyze in this report is whether or not PM exposure levels are subject-specific (as suggested by the graphic above).

We are interested to know if certain activities lead to higher PM exposure for children on average, so we next block by activity. The boxplot below shows the spread of aerosol PM readings across all activities. At first glance there appears to be reasonably equal spread, with the exception of a few outliers.



Having explored the data and obtaining an idea of potential complexities we face in building our model, we proceed with utilizing a generalized least squares model structure for this PM data. We use generalized least squares to allow us to address the possibility of unequal variance and correlation in the residuals.

2 Statistical Model

In order to determine the model to use, we first used a BoxCox statistic on a basic linear model, as explained earlier, leading us to realize that we should take the log of the aerosol measurement. This ultimately allowed for better model fit and a lower AIC. Initially, we fit an independent linear model on the data, but proved our concerns mentioned above when we discovered a considerable amount of residual correlation. Thus, we identified which correlation structure best fit the error terms in the longitudinal data. The structure with the lowest AIC value of the ones we looked at (including AR(1), MA(1), and compound symmetric), was the AR(1) model. Below is the statistical model we will be using in mathematical form.

$$\mathbf{y} \sim MVN(\mathbf{X}\boldsymbol{\beta}, \sigma^2\mathbf{B})$$

where

$$
\mathbf{y} = \begin{pmatrix} \log(\text{Aerosol})_{1,1} \\ \log(\text{Aerosol})_{1,2} \\ \vdots \\ \log(\text{Aerosol})_{60,118} \end{pmatrix}
$$

$$
\mathbf{X} = \begin{pmatrix} 1 & \text{Stationary}_{1,1} & \text{Homework}_{1,1} & \dots & \text{WatchingTV}_{1,1} & \text{ID} = 1_{1,1} & \dots & \text{ID} = 60_{1,1} & \dots \\ 1 & \text{Stationary}_{1,2} & \text{Homework}_{1,2} & \dots & \text{WatchingTV}_{1,2} & \text{ID} = 1_{1,2} & \dots & \text{ID} = 60_{1,2} & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & \text{Stationary}_{60,118} & \text{Homework}_{60,118} & \dots & \text{WatchingTV}_{60,118} & \text{ID} = 1_{60,118} & \dots & \text{ID} = 60_{60,118} & \dots \end{pmatrix}
$$

where the subscript i,j for each term in matrices $\mathbf{y}$ and $\mathbf{X}$ refer to the i^{th} subject and the j^{th} minute. Note the X matrix extends to include ID*Activity fixed effects and ID*Stationary interactions.

$$
\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_{\text{Stationary}} \\ \beta_{\text{Homework}} \\ \vdots \\ \beta_{\text{WatchingTV}} \\ \beta_{\text{ID=1}} \\ \vdots \\ \beta_{\text{ID=60}} \\ \beta_{\text{ID=1}*Homework} \\ \vdots \\ \beta_{\text{ID=60}*WatchingTV} \\ \beta_{\text{ID=1}*Stationary} \\ \vdots \\ \beta_{\text{ID=60}*Stationary} \end{pmatrix}, \mathbf{B} = \begin{pmatrix} \mathbf{R} & 0 & 0 & \dots & 0 \\ 0 & \mathbf{R} & 0 & \dots & 0 \\ 0 & 0 & \mathbf{R} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \mathbf{R} \end{pmatrix} = \text{diag}(\mathbf{R}, \dots, \mathbf{R})
$$

In the $\boldsymbol{\beta}$ matrix, when β is composed of an ID=i * Activity, references the interaction effect of a particular subject (ID) and activity. For the $\mathbf{B}$ matrix, we define $\mathbf{R}$ to be the following:

$$\mathbf{R} = \begin{pmatrix} 1 & \rho(1,2) & \rho(1,3) & \dots & \rho(1,118) \\ \rho(2,1) & 1 & \rho(2,3) & \dots & \rho(2,118) \\ \rho(3,1) & \rho(3,2) & 1 & \dots & \rho(3,118) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho(118,1) & \rho(118,2) & \rho(118,3) & \dots & 1 \end{pmatrix}$$

$$\rho(t_1, t_2) = \text{Correlation between time } t_1 \text{and } t_2$$

Since we are using an AR(1) correlation, we can mathematically represent the correlation function as:

$$\text{AR(1): } \rho(t_1, t_2) = \phi^{|t_1, t_2|}$$

In the model mathematically defined above, $\mathbf{y}$ is a vector of our response variable, the log of the PM measurement on the child's vest, for each subject-minute observation (60 subjects, 118 minutes per subject). The $\mathbf{X}$ matrix includes first a column of 1's which allow for our intercept, β_0. Following the column of 1's, we have a column for each of the explanatory variables we are using, which includes Stationary (the PM measurement of the stationary monitor), as well as dummy variables for the following activities: homework, being on the phone, playing on the floor, playing on furniture, video games, walking, and watching TV. We use being on the computer as a baseline. For each activity, ID, or ID*Activity, observation, we put a 1 if the participant was performing the particular activity, was given that particular ID, or both, respectively, and a 0 otherwise. For ID*Stationary, we put the value of the child's stationary measurement if the child was received the particular ID, and a 0 otherwise. The $\boldsymbol{\beta}$ matrix is a vector of the intercept and slopes (estimated effect sizes) for each of the variables (including any of the particular interactions).

Finally, $\sigma^2\mathbf{B}$ is a matrix which tells us the covariance of $\mathbf{y}$. Now, each element in the diagonal of $\mathbf{B}$, denoted as $\mathbf{R}$, is a 118 by 118 symmetric matrix, where the i^{th} row and the j^{th} column is the correlation of time i with time j.

Now, this model is built under certain assumptions. First, we are assuming that there is a linear relationship between The log of the PM measurement on the child's vest and the explanatory variables, or at least that a linear relationship does no better than any other model at summarizing the relationship. Since activities are categorical variables, we really only need to worry about its relationship with the PM measurement of the stationary monitor. If this relationship is linear, then we will have avoided one potential source of bias in our estimations of $\boldsymbol{\beta}$.

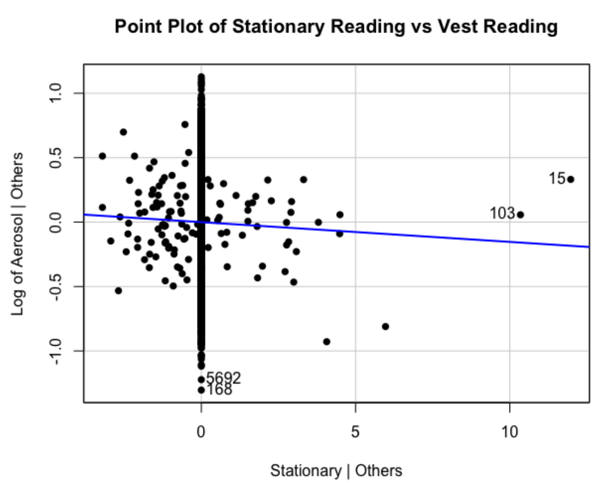
Next, we assume independence between subjects; however, serial correlation across time should be resolved through use of the AR(1) model as explained above. That is, once we account for AR(1) correlated residuals in the linear model, we assume that any dependence between subjects and time has been accounted for. We will show this to be the case in the next section. By decorrelating the residuals, we avoid another source of bias in our β estimates and avoid incorrect standard errors, which may otherwise lead to incorrect inference. We also assume normality of the residuals after we account for the correlation across time to help avoid biased parameters.

Unfortunately, one source of bias that is difficult to fully account for is endogeneity, specifically through omitted variables. It is likely that there are variables hidden within the residuals that are difficult to account for that are related to both our explanatory variables and our dependent variables. Such factors could include cleaning habits of the parents or children or the demographics or other characteristics of the family. Since this data is not available, it is likely that our estimates will be slightly biased.

The final assumption in this model is equal variance of the error term. Here, we assume that the value of the error term is not different depending on the value of our explanatory variable. If instead we have heteroskedasticity, then our standard errors will be wrong and the model will be inefficient. Again, this would lead to incorrect inference. We will now justify these assumptions through the use of graphics, summary statistics, and hypothesis tests.

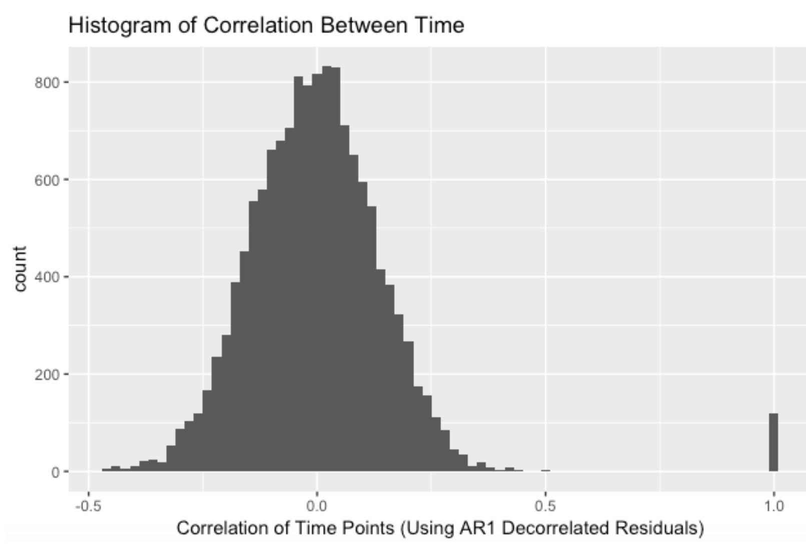
3 Model Validation

As mentioned in the previous section, the validity of our model depends on several assumptions holding true. The first assumption we will look at is linearity of the relationship between independent and dependent variables. To do so, we will look at an added variable plot of stationary against the log of the aerosol variable. This plot is shown below:



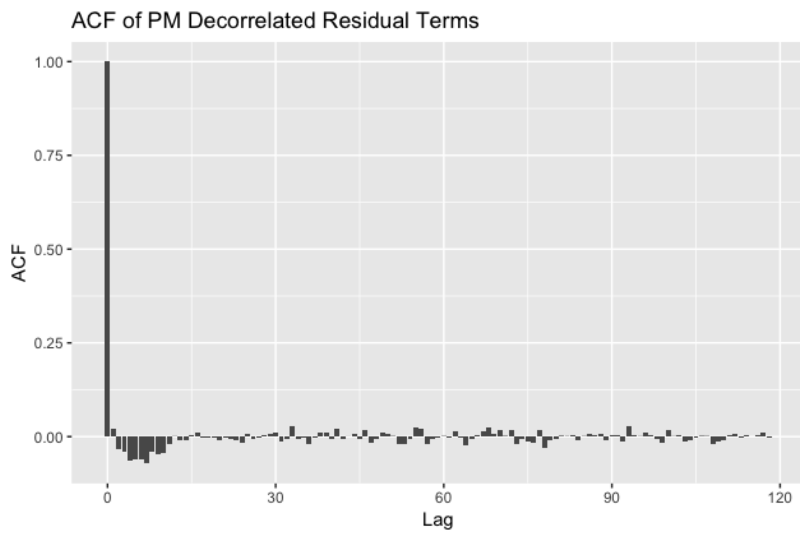
It appears that a linear relationship between the log of aerosol and stationary would perform just as well or better than any other model. We therefore conclude that the linearity assumption is met, and continue to analyze the validity of the assumptions by determining whether or not the AR(1) model is sufficient to account for the correlation of the residuals in the model.

Next, we examine the independence assumption. Understanding intuitively that there would be a correlation across time, but not across subjects, we used an AR(1) model to capture the correlation across time. This is, at the very least, a decent model intuitively to account for such correlation since the relationship between two minutes that are close together (e.g. the first and second minute) should be more correlated than two minutes further apart. That is, as the difference in time increases, the correlation between the minutes decreases. Through AIC measurements and ACF graphs, we determine that the AR(1) model is, in fact an appropriate model to use to account for the minute-to-minute correlation. Specifically, the AIC value (approximately 4,123) using the AR(1) model was much lower than the AIC from other models tested. Note that the histogram double-counts correlations by counting the correlation of minute i and minute j, as well as the correlation of minute j and minute i separately. Not counting the correlation of the same minutes (where correlation would be 1), we find that the mean of the decorrelated residuals is -0.008, with standard deviation 0.132.



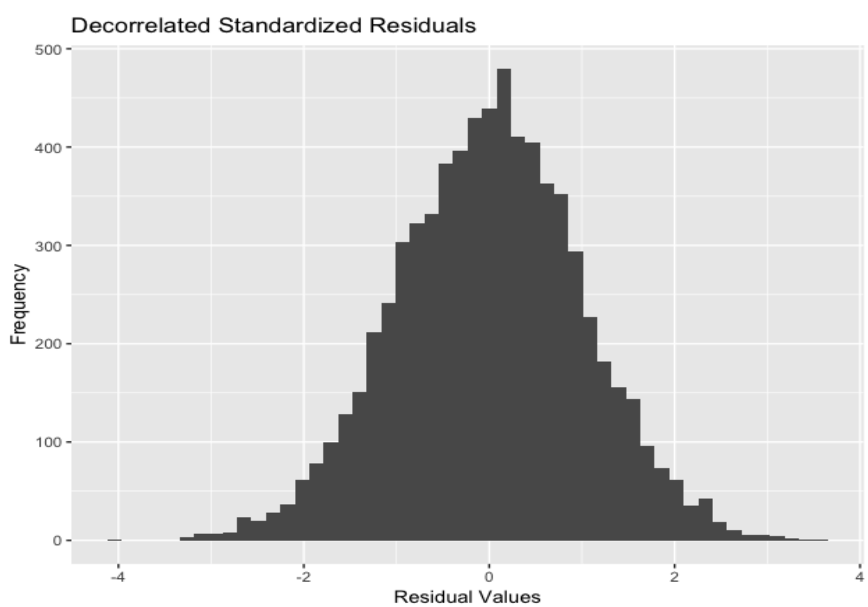
The ideal residual correlation values should have the majority of correlations range somewhere between -0.2 and 0.2, though this threshold is largely left to the discretion of the researcher to determine. As is shown in the histogram below, the correlation values after incorporating the autoregressive correlation structure had about 87% of the correlations (not including correlations of the same minute) range between -0.2 and 0.2.

We also looked at the ACF plot to ensure that there were no obvious correlation trends after accounting for AR(1) correlation. Below is the ACF of decorrelated residuals.



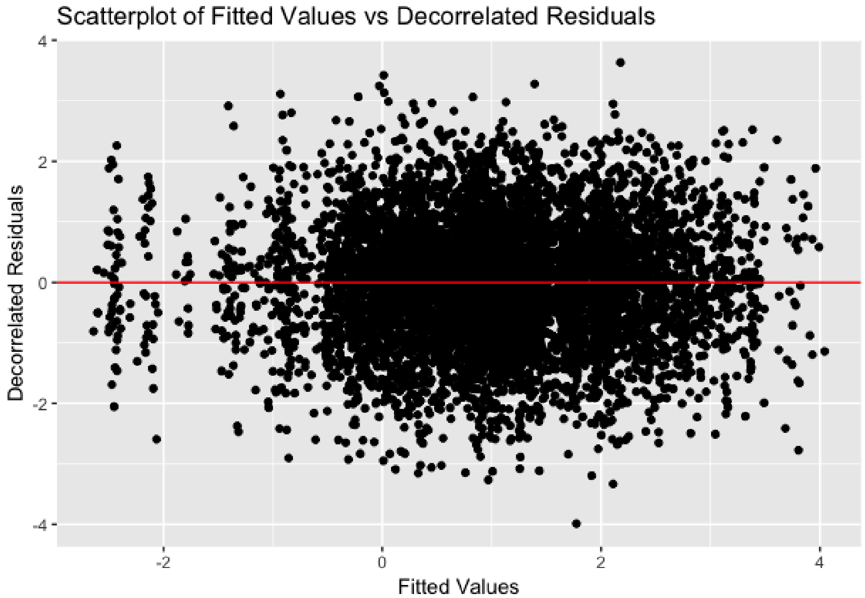
While there appears to still be a small correlation between the first minute or so, the correlation across time is low enough to suggest that the AR(1) model successfully controlled for the majority of the correlation.

Next, we will determine whether or not the error terms are normally distributed.



The histogram above displays the spread of the decorrelated residual values which appears to be effectively normal. A K-S test resulting in a p-value of 0.5827 tells us that the error terms are, in fact, normally distributed.

The final assumption to look at is equal variance. To begin, we look at a graph of the fitted values against the decorrelated residual terms. This scatterplot is shown below.



Visually, it appears that there is no relationship between the fitted values and the observed variance of the residual terms. A (manually performed) Breusch-Pagan test results in a p-value of 0.1863, allowing us to conclude that we have homoskedasticity. Seeing that our model fulfills all four of our LINE assumptions (Linearity, Independence, Normality of residuals, and Equal variance), we can conclude that our model is sufficient for producing estimates that are unbiased and of minimum variance.

4 Analysis Results

Having satisfied all the necessary assumptions of our model, we move into the results of our analysis using our model.

The first question of interest was whether or not stationary readings of PM levels alone performed well at explaining the aerosol readings of PM levels. We compared our model against a model only containing the variable Stationary as predictor using a simple ANOVA test. The p-value returned was < 0.0001, indicating that the two models' explanation of the response is statistically different. The AIC value of our model, again, is approximately 4,123, whereas the AIC value of the model containing only Stationary is approximately 7,025. Therefore, we conclude that Stationary alone is an insufficient predictor of Aerosol readings.

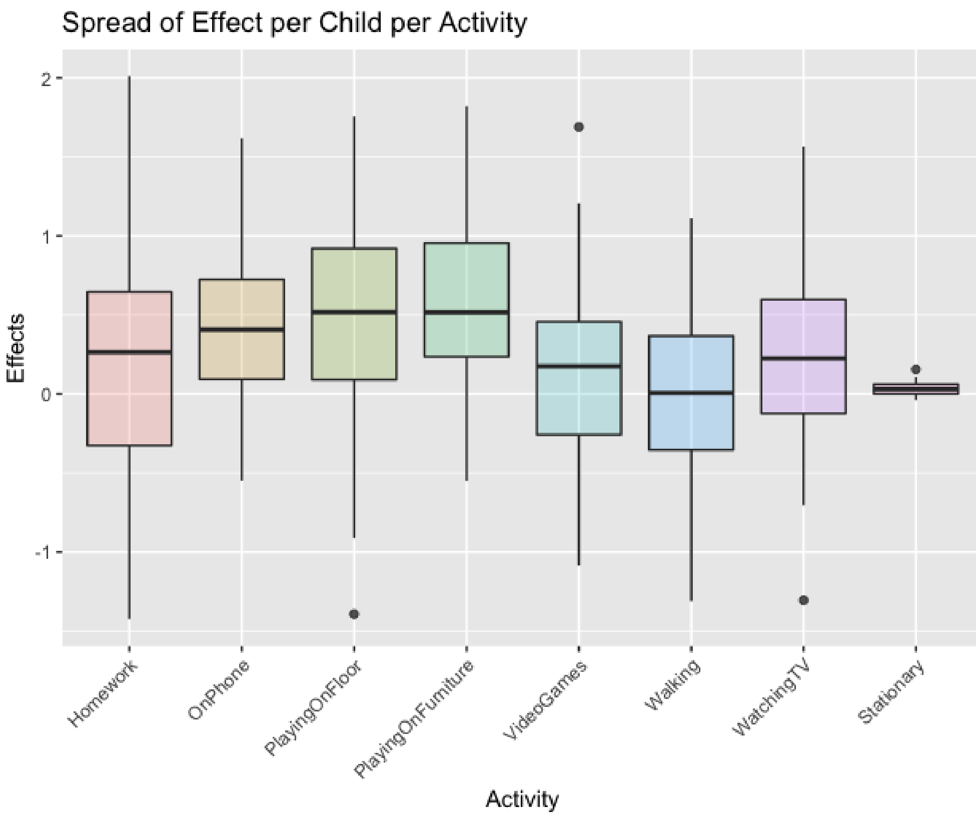
The second question of interest was whether or not the variable Activity, in addition to Stationary, helped explain the Aerosol reading. We executed another ANOVA test comparing our model to a model absent of

Activity as a predictor. The p-value returned was also < 0.0001, indicating a significant difference between the models in explaining the values of the Aerosol reading. The model excluding Activity as a predictor has an AIC value of approximately 5,750, still much higher than our model's AIC value of 4,123. We conclude that the variable Activity, in addition to Stationary, did help explain the values of the Aerosol reading.

In order to determine exactly which activity(ies) is(are) significant, we calculated point estimates and 95% confidence intervals of each effect size of the activities included in our model and only found playing on furniture to be statistically significant in reference to playing on the computer. The calculated effect size for playing on furniture was -0.401 with 95% confidence interval (-0.762, -0.041) and p-value 0.029, suggesting that playing on furniture results in less PM being breathed in than being on the computer. However, there was not enough evidence to suggest that any other activity resulted in a statistically significant difference from being on the computer.

The final question of interest in this report is whether or not the effects of the activities and stationary readings on PM exposure were child-specific. We refer back to when we first introduced this notion with the boxplot showing the unequal spread of aerosol PM readings across the 60 subjects. To best answer this question, we utilized another ANOVA test comparing our model to a model that excludes all interaction effects between ID and Activity and ID and Stationary. The calculated p-value was < 0.0001, confirming a difference between the models in explaining the aerosol PM readings. The model absent of interaction terms returned an AIC value of approximately 5,921, which is still considerably greater than our model's AIC value of 4,123. We conclude once more that our model, which includes the interaction effects between ID and Activity and ID and Stationary, explains the response values better than the model without any interaction effects.

Given our conclusions, we can reasonably assume that the effects of Activity and, to a lesser extent, stationary, on the aerosol PM readings are child-specific. To help visualize the variation in the effect sizes of the activities and stationary readings on each child, we supply the following boxplot.



As clearly shown in the boxplot above, only the effect of stationary appears to be reasonably constant across each child. Since we are only interested to know whether or not the interaction between ID and Stationary help explain the logged values of aerosol for each child, we perform our last ANOVA test to compare our model against a model with all the same predictors except the interaction between ID and Stationary. The ANOVA test returned a p-value of < 0.0001, which confirms a difference in the two model's ability to explain the data. The model without ID and Stationary interaction term produced an AIC value of 4,455, which is larger than our model's AIC value of 4,123. We conclude that the interaction between ID and Stationary is appropriate to leave in our model. We then calculated the actual variance in the effect sizes of each activity. The table below provides estimates of the variance of effects across students (variance of ID*Activity or ID*Stationary), alongside the average difference across the subjects (mean of ID*Activity or ID*Stationary term).

Activity/Stationary	Average Difference in Effect of Activity	Variance of Effect
Homework	0.220	0.552
On Phone	0.379	0.202
Playing on Floor	0.527	0.452
Playing on Furniture	0.563	0.317
Video Games	0.113	0.253
Walking	-0.0147	0.274
Watching TV	0.209	0.272
Stationary	0.033	0.00155

In reference to playing on the computer, doing homework appears to have the greatest variation in effect sizes of aerosol PM readings per child.

5 Conclusions

Using the model we created and validated for our supplied PM Exposure data set, our analysis concludes that stationary PM readings help explain the actual PM exposure of each subject. The validity of this claim, however, is arguably challenged by the results displayed in the last graphic of our report. The effect of stationary readings for each child appears to be essentially zero, which challenges the notion that stationary readings are of any use in explaining the actual PM exposure of each child subject in this study. We suggest that researchers in the future replicate this experiment with both including multiple stationary monitors operating during each trial and retesting each child subject more than once.

While we also find that activities help explain the level of PM that a child is exposed to, individual hypothesis tests reveal that the only activity statistically different from being on the computer is playing on the furniture. One possible confounding factor in this conclusion, however, is the difference in initial air quality in each household of the 60 subjects. We suggest that researchers implement a method for blocking subjects according to initial readings of PM levels inside each house. By including such blocking, researchers will better see the effects of each activity across the different subjects because any extreme PM readings will be contained within their respective blocks and will not influence the results of all other subjects.

Code Appendix

```
# Import Data and Libraries -------------------
library(ggplot2)
library(nlme)
install.packages('mvtnorm')
library(mvtnorm)
library(car)
install.packages('multcomp')
library(multcomp)
library(astsa)
library(tidyverse)
library(magrittr)
source("~/Desktop/Winter_2019/STAT_469/stdres.gls.R")

#Read in data  -------------------
pm <- read.table("https://mheaton.byu.edu/Courses/Stat469/Topics/2%20-%20
                 TemporalCorrelation/3%20-%20Project/Data/BreathingZonePM.txt",header = TRUE)
pm$ID <- as.factor(pm$ID)

#Justifying log transformation of Aerosol  -----------
ggplot(data=pm, mapping=aes(x=Stationary, y=Aerosol)) + geom_point(colour='steel blue') +
  ggtitle('Stationary and Aerosal PM Levels') +
  geom_abline(intercept=0, slope = 0.1, color='red', linetype='dashed')

initmod <- lm(Aerosol ~ ., data = pm)
boxCox(initmod)
ggplot(pm) + geom_histogram(aes(x=Aerosol),bins = 40) +
  ggtitle("Spread of Aerosol") + xlab("Aerosol") + ylab("Frequency")
ggplot(pm) + geom_histogram(aes(x=log(Aerosol)),bins = 40) +
  ggtitle("Spread of Log of Aerosol") + xlab("Log of Aerosol") + ylab("Frequency")
pm$LogAero <- log(pm$Aerosol)

#Exploratory graphics  -------------------
ggplot(pm,aes(x=Activity,y=Aerosol, fill = Activity)) +
  geom_boxplot(alpha=0.3, outlier.alpha = 1, outlier.size = 0.5) +
  ggtitle("PM Levels by Activity") +
  theme(legend.position = "none", axis.text.x = element_text(angle = 45, hjust = 1)) +
  ylab("Recorded PM Level")
ggplot(pm,aes(x=as.factor(ID),y=Aerosol, fill = ID)) +
  geom_boxplot(alpha = 0.3, outlier.alpha = 1, outlier.size = 0.5) +
  ggtitle("Spread of Log of Aerosol Reading for Each Subject") +
  xlab("Subject ID") + theme(legend.position = "none") +
  ylab("Log of Aerosol") + scale_x_discrete(breaks = c(1,seq(0,60,by = 5)))
ggplot(data=pm, mapping=aes(x=Stationary, y=logaerosol)) +
  geom_point(colour='steel blue') +
  labs(title='Stationary and Aerosal PM Levels', y='Log Aerosol')

#Correlated residual model
pmmod1 <- lm(LogAero ~ ID*Stationary + ID*Activity, data = pm)

#Residual correlations
rmat <- matrix(pmmod1$residuals, ncol = length(unique(pm$Minute)), byrow = TRUE)
cors <- cor(rmat)
```

```
corvec <- matrix(cors, ncol = 1, nrow = 118*118, byrow = FALSE)
ggplot() + geom_histogram(aes(x=corvec), bins = 100) +
  ggtitle("Residual Correlation Between Subject's Aerosol Reading Per Minute") +
  xlab("Residual Correlation Values") + ylab("Frequency")
summary(cors)
hist(cors)
dim(re)

#Model selection  -------------------
pm_ar <-   gls(LogAero ~ Stationary*ID + Activity*ID, data = pm,
               correlation = corARMA(form = ~Minute|ID, p = 1, q = 0),method = "ML")
pm_ma <-   gls(LogAero ~ Stationary*ID + Activity*ID, data = pm,
               correlation = corARMA(form = ~Minute|ID, p = 0, q = 1),method = "ML")
pm_arma <- gls(LogAero ~ Stationary*ID + Activity*ID, data = pm,
               correlation = corARMA(form = ~Minute|ID, p = 1, q = 1),method = "ML")
AIC(pm_ar)
AIC(pm_ma)
AIC(pm_arma)

#Determining the percentage of correlation values between -0.2 and 0.2
decor_resids <- stdres.gls(pm_ar)
decor.resids.mat <- matrix(decor_resids, ncol = 118, byrow = TRUE)
dec.cors <- cor(decor.resids.mat)
dec.cors2 <- as.data.frame(matrix(ncol=1, data=dec.cors))

dec.cors.other <- subset(dec.cors2, V1!=1)
mean(dec.cors.other$V1)
sd(dec.cors.other$V1)

dec.res.good <- subset(dec.cors.other, abs(V1) < 0.2)
dim(dec.res.good)[1] / dim(dec.cors.other)[1]

#Creating ACF plot of decorrelated residuals

my.ACF <- acf(stdres.gls(pm_ar), lag.max=118)
ACF.dframe <- data.frame(Lag=my.ACF$lag, ACF=my.ACF$acf)
ggplot(data=ACF.dframe, aes(x=Lag, y=ACF)) + geom_col() +
  ggtitle('ACF of PM Decorrelated Residual Terms')

#Chosen Model
mod <- gls(LogAero ~ Stationary*ID + Activity*ID, data = pm,
           correlation = corAR1(form = ~Minute|ID),method = "ML")
AIC(mod)

#linearity  -------------------
avplotmod <- lm(LogAero ~ Stationary*ID + Activity*ID, data = pm)
avPlot(avplotmod, variable = "Stationary",
       main = "Point Plot of Stationary Reading vs Vest Reading",
       pch = 16, xlab = "Stationary | Others", ylab = "Log of Aerosol | Others")

#Normality  -------------------
dstdr <- stdres.gls(mod)
ks.test(dstdr,"pnorm")
ggplot() + geom_histogram(aes(x=dstdr),bins = 50) +
```

```
  ggtitle("Decorrelated Standardized Residuals") +
  xlab("Residual Values") + ylab("Frequency")

#Independence  -------------------
dcrmat <- matrix(dstdr, ncol = 118, byrow = TRUE)
cors <- cor(dcrmat)
corvec <- matrix(cors, ncol = 1, nrow = 118*118, byrow = FALSE)
ggplot() + geom_histogram(aes(x=corvec),bins = 60)

#Equal Variance  -------------------
ggplot() + geom_point(aes(x=mod$fitted, y=dstdr)) +
  geom_hline(yintercept = 0, colour = "red")

# BP Test

u <- pm_ar$residuals
pm$usq.AR <- u^2
bp.lm <- lm(data=pm, formula=usq.AR ~ Stationary + Activity + ID*Stationary)
summary(bp.lm) # There is NO heteroskedasticity (p-value: 0.1863)


#Sec4.1  -------------------
onlystat <- gls(LogAero ~ Stationary, data = pm,
                correlation = corAR1(form = ~Minute|ID),method = "ML")
anova(mod,onlystat)
#Sec4.2  -------------------
noact <- gls(LogAero ~ Stationary*ID, data=pm,
             correlation = corAR1(form = ~Minute|ID),method = "ML")
anova(noact, mod)
confint(mod,level = 0.95)["ActivityHomework",] #-0.0124873001
confint(mod,level = 0.95)["ActivityOnPhone",] #-0.2265575621
confint(mod,level = 0.95)["ActivityPlayingOnFloor",] #-0.0751911609
confint(mod,level = 0.95)["ActivityPlayingOnFurniture",] #-0.4010977439
confint(mod,level = 0.95)["ActivityVideoGames",] #0.1071637381
confint(mod,level = 0.95)["ActivityWalking",] #0.0971668000
confint(mod,level = 0.95)["ActivityWatchingTV",] #0.0295899248
#Sec4.3  -------------------
nointer <- gls(LogAero ~ ID + Stationary + Activity + Minute, data = pm,
               correlation = corAR1(form = ~Minute|ID),method = "ML")
anova(mod,nointer)

#For boxplots of effects across children
ts <- summary(mod)$tTable
homeworkeffect <- subset(ts, grepl("*:ActivityHomework*",rownames(ts)))[,"Value"]
phoneeffect <- subset(ts, grepl("*:ActivityOnPhone*",rownames(ts)))[,"Value"]
furneffect <- subset(ts, grepl("*:ActivityPlayingOnFurniture*",rownames(ts)))[,"Value"]
flooreffect <- subset(ts, grepl("*:ActivityPlayingOnFloor*",rownames(ts)))[,"Value"]
walkeffect <- subset(ts, grepl("*:ActivityWalking*",rownames(ts)))[,"Value"]
vgeffect <- subset(ts, grepl("*:ActivityVid*",rownames(ts)))[,"Value"]
watchtveffect <- subset(ts, grepl("*:ActivityWatch",rownames(ts)))[,"Value"]
stationeffect <- subset(ts, grepl("*:Stationary",rownames(ts)))[,"Value"]

hwdf <- data.frame("Homework",homeworkeffect)
phdf <- data.frame("OnPhone",phoneeffect)
```

```
fndf <- data.frame("PlayingOnFurniture",furneffect)
fldf <- data.frame("PlayingOnFloor",flooreffect)
wkdf <- data.frame("Walking",walkeffect)
vgdf <- data.frame("VideoGames",vgeffect)
wtdf <- data.frame("WatchingTV",watchtveffect)
stdf <- data.frame("Stationary",stationeffect)

names(hwdf) <- c("Activity","Effects")
names(phdf) <- c("Activity","Effects")
names(fndf) <- c("Activity","Effects")
names(fldf) <- c("Activity","Effects")
names(wkdf) <- c("Activity","Effects")
names(vgdf) <- c("Activity","Effects")
names(wtdf) <- c("Activity","Effects")
names(stdf) <- c("Activity","Effects")

dat <- rbind(hwdf,phdf,fldf,fndf,vgdf,wkdf,wtdf,stdf)

rownames(dat) <- NULL
boxplot(Effects ~ Activity, data = dat)
#Boxplot
ggplot(dat,aes(x=Activity,y=Effects, fill=Activity)) +
  geom_boxplot(alpha = 0.2, outlier.alpha = 1) +
  theme(legend.position = "none", axis.text.x = element_text(angle = 45,hjust = 1)) +
  ggtitle("Spread of Effect per Child per Activity")

onlyactinter <- gls(LogAero ~ Stationary + Activity*ID, data = pm,
                    correlation = corAR1(form = ~Minute|ID),method = "ML")
anova(mod, onlyactinter)

#Calculate Mean and Variance of individual effects of activity/stationary
dat %>% group_by(Activity) %>% summarize("Var" = mean(Effects))
dat %>% group_by(Activity) %>% summarize("Var" = var(Effects))
```