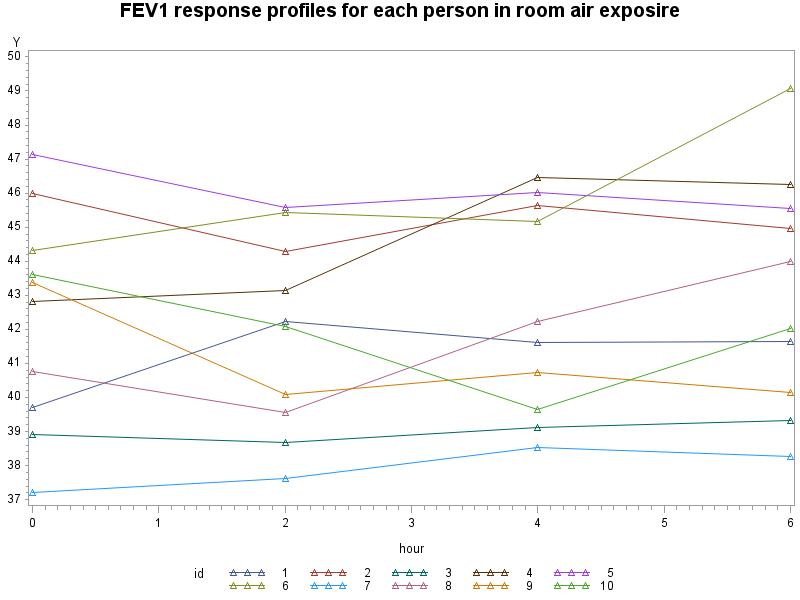
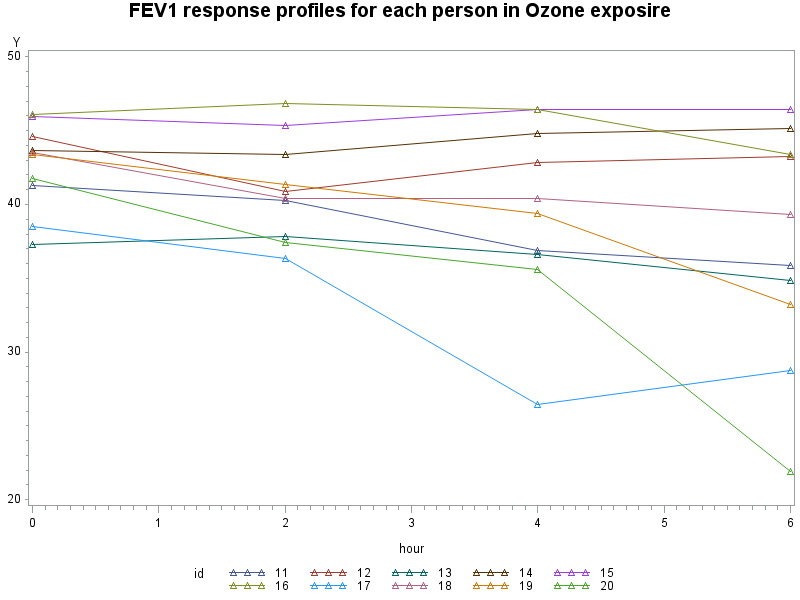
1. **Descriptive Analyses:**

**a.Describe key aspects of the longitudinal design and completeness of data.**

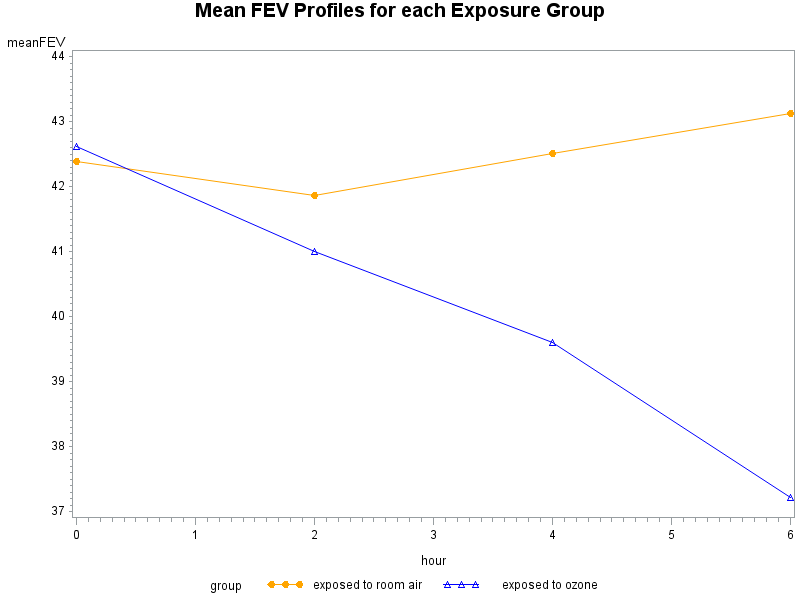
We have a randomized study design with every subjects randomly assigned to either ozone exposure or room air exposure. And we have a balanced design with all subjects measured at the same measurement times (hour 0,1,2,3,4,5, and 6). Also, our data used in this analysis which only contains hour 0, 2, 4, and 6, are complete (no missing data).

**b. Plot the FEV1 response against hour for each subject exposed to room air(overlaid on the same plot) and for each subject exposed to ozone (overlaid on a second plot). Comment on any patterns in the data or other notable aspects**

**of the data.**

In group exposed to room air, the pattern is not monotonically increasing nor decreasing, during hour 0 to 2, many subjects decreased, while after hour 2, most of them increased a little bit. Overall, FEV1 response levels don’t change dramatically.

In group exposed to ozone, some subjects stays at a relative stable level, while some subjects shows a dramatic decrease, and the variation at the end of the study is discernibly larger than that of group exposed to room air

**c. Obtain the mean FEV1 value at each hour of measurement for ozone and room air subjects separately. Plot the means against hour. Comment on the pattern of change in mean FEV1 with hour for ozone and room air subjects.**

At baseline, the mean FEV1 are at the same level for the two exposure groups, this is concord with the randomization at baseline. After baseline (hour 0), for the room air exposure group, the mean FEV1 decreases a little at hour 2, and then increases somewhat in a linear fashion from hour 2 to 6; After baseline (hour 0), for the ozone exposure group, the mean FEV1 decreases dramatically in a linear pattern until the end (hour 6).

1. **Fitting a “Maximal” Model and Evaluating Variance-Covariance Structure:**

**a.Define a reasonable “maximal” mean model for this study. Fit this model using an unstructured variance-covariance matrix. Comment on the variance structure and on the correlation structure. What simplified variance-covariance structure(s) might be reasonable? Justify your answer.**

Since the study design is balanced and no missing data, and also the number of measurement occasions is 4 which is not too large, we can choose a “saturated” model for the mean response, with main effects of time and group, and also the way-way interaction of time and group.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimated R Matrix for id 1 | | | | |
| Row | **Col1** | **Col2** | **Col3** | **Col4** |
| 1 | 9.3296 | 8.0468 | 11.1510 | 11.1966 |
| 2 | 8.0468 | 9.7035 | 12.9840 | 14.9012 |
| 3 | 11.1510 | 12.9840 | 23.4591 | 24.2632 |
| 4 | 11.1966 | 14.9012 | 24.2632 | 36.8041 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimated R Correlation Matrix for id 1 | | | | |
| Row | **Col1** | **Col2** | **Col3** | **Col4** |
| 1 | 1.0000 | 0.8457 | 0.7537 | 0.6042 |
| 2 | 0.8457 | 1.0000 | 0.8606 | 0.7885 |
| 3 | 0.7537 | 0.8606 | 1.0000 | 0.8257 |
| 4 | 0.6042 | 0.7885 | 0.8257 | 1.0000 |

Variance increases over time, and variance appears to be much larger (almost 4 times) by the end of the study when compared to the variance at baseline.

Correlations decrease as the time separation between the repeated measures increases.

Based on the above observations, heterogeneous variance should be assumed. As for correlation, compound symmetry is not suitable since it ignores the apparent decreasing pattern over time; Toeplitz is not the best option since it has 3 correlation parameters + 4 variance parameters, it’s too large relative to our sample N=20. Instead, first-order autoregressive with heterogeneous variance is the most appropriate model for the covariance structure, since we have equal-spaced measurements, and AR(1) incorporates the decreasing pattern of correlation among repeated measures and at the same time only has 1 correlation parameters +4 variance parameters. The heterogeneous autoregressive covariance model is optimal for this case.

**b. Keeping the same maximal mean model, evaluate whether your suggestion(s) for the variance-covariance structure from question 2a as well as the following models for the variance-covariance structure provide an adequate fit to the data compared with an unstructured variance-covariance:**

**i. compound symmetry**

**ii. heterogeneous compound symmetry**

**iii. 1st-order autoregressive**

**iv. heterogeneous 1st-order autoregressive.**

**Using likelihood ratio tests and the AIC criterion as appropriate, identify a model for the variance-covariance structure that provides a good fit to the data. Provide estimates for the parameters used in defining this variance-covariance model. Also provide estimates of the variance-covariance and**

**correlation matrices.**

**i. compound symmetry versus unstructured**

Table: Maximized (REML) log-likelihood

|  |  |
| --- | --- |
| Model | -2(REML)Log-likelihood |
| CS | 394.1 |
| UN | 355.2 |

There is a hierarchy among the models: heterogeneous AR(1) is nested within unstructured. LRT yields with 8 d.f. (p<.0001), so we reject the null hypothesis and conclude that compound symmetry is not a good fit to the data when compared to the unstructured.

This model results in the following estimates of the variance and correlation parameters,

Table: Estimated CS variance-covariance matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimated R Matrix for id 1 | | | | |
| Row | **Col1** | **Col2** | **Col3** | **Col4** |
| 1 | 19.8241 | 13.7571 | 13.7571 | 13.7571 |
| 2 | 13.7571 | 19.8241 | 13.7571 | 13.7571 |
| 3 | 13.7571 | 13.7571 | 19.8241 | 13.7571 |
| 4 | 13.7571 | 13.7571 | 13.7571 | 19.8241 |

Table: Estimated CS correlation matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimated R Correlation Matrix for id 1 | | | | |
| Row | **Col1** | **Col2** | **Col3** | **Col4** |
| 1 | 1.0000 | 0.6940 | 0.6940 | 0.6940 |
| 2 | 0.6940 | 1.0000 | 0.6940 | 0.6940 |
| 3 | 0.6940 | 0.6940 | 1.0000 | 0.6940 |
| 4 | 0.6940 | 0.6940 | 0.6940 | 1.0000 |

**ii. heterogeneous compound symmetry versus unstructured**

Table: Maximized (REML) log-likelihood

|  |  |
| --- | --- |
| Model | -2(REML)Log-likelihood |
| heterogeneous CS | 365.3 |
| UN | 355.2 |

There is a hierarchy among the models: heterogeneous AR(1) is nested within unstructured. LRT yields with 5 d.f. (p= 0.072451), so we can not reject the null hypothesis and conclude that heterogeneous compound symmetry is a good fit to the data when compared to the unstructured.

This model results in the following estimates of the variance and correlation parameters,

Table: Estimated CS variance-covariance matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimated R Matrix for id 1 | | | | |
| Row | **Col1** | **Col2** | **Col3** | **Col4** |
| 1 | 9.8538 | 7.4162 | 11.6505 | 15.2437 |
| 2 | 7.4162 | 9.1445 | 11.2234 | 14.6849 |
| 3 | 11.6505 | 11.2234 | 22.5682 | 23.0695 |
| 4 | 15.2437 | 14.6849 | 23.0695 | 38.6355 |

Table: Estimated CS correlation matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimated R Correlation Matrix for id 1 | | | | |
| Row | **Col1** | **Col2** | **Col3** | **Col4** |
| 1 | 1.0000 | 0.7813 | 0.7813 | 0.7813 |
| 2 | 0.7813 | 1.0000 | 0.7813 | 0.7813 |
| 3 | 0.7813 | 0.7813 | 1.0000 | 0.7813 |
| 4 | 0.7813 | 0.7813 | 0.7813 | 1.0000 |

**iii. 1st-order autoregressive versus unstructured**

Table: Maximized (REML) log-likelihood

|  |  |
| --- | --- |
| Model | -2(REML)Log-likelihood |
| AR(1) | 381.1 |
| UN | 355.2 |

There is a hierarchy among the models: heterogeneous AR(1) is nested within unstructured. LRT yields with 8 d.f. (p= .001092488), so we reject the null hypothesis and conclude that 1st-order autoregressive covariance model is not an adequate fit to the data when compared to the unstructured.

This model results in the following estimates of the variance and correlation parameters,

Table: Estimated CS variance-covariance matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimated R Matrix for id 1 | | | | |
| Row | **Col1** | **Col2** | **Col3** | **Col4** |
| 1 | 21.5176 | 17.8239 | 14.7643 | 12.2299 |
| 2 | 17.8239 | 21.5176 | 17.8239 | 14.7643 |
| 3 | 14.7643 | 17.8239 | 21.5176 | 17.8239 |
| 4 | 12.2299 | 14.7643 | 17.8239 | 21.5176 |

Table: Estimated CS correlation matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimated R Correlation Matrix for id 1 | | | | |
| Row | **Col1** | **Col2** | **Col3** | **Col4** |
| 1 | 1.0000 | 0.8283 | 0.6862 | 0.5684 |
| 2 | 0.8283 | 1.0000 | 0.8283 | 0.6862 |
| 3 | 0.6862 | 0.8283 | 1.0000 | 0.8283 |
| 4 | 0.5684 | 0.6862 | 0.8283 | 1.0000 |

**iv. heterogeneous 1st-order autoregressive versus unstructured (also My suggestion)**

Table: Maximized (REML) log-likelihood

|  |  |
| --- | --- |
| Model | -2(REML)Log-likelihood |
| Heterogeneous AR(1) | 358.2 |
| UN | 355.2 |

There is a hierarchy among the models: heterogeneous AR(1) is nested within unstructured. LRT yields with 5 d.f. (p= 0.69999), so we can not reject the null hypothesis and conclude that heterogeneous 1st-order autoregressive covariance model is an adequate fit to the data when compared to the unstructured.

This model results in the following estimates of the variance and correlation parameters,

Table: Estimated CS variance-covariance matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimated R Matrix for id 1 | | | | |
| Row | **Col1** | **Col2** | **Col3** | **Col4** |
| 1 | 9.1838 | 7.8710 | 10.4343 | 11.1475 |
| 2 | 7.8710 | 9.4883 | 12.5783 | 13.4381 |
| 3 | 10.4343 | 12.5783 | 23.4532 | 25.0563 |
| 4 | 11.1475 | 13.4381 | 25.0563 | 37.6514 |

Table: Estimated CS correlation matrix

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Estimated R Correlation Matrix for id 1 | | | | |
| Row | **Col1** | **Col2** | **Col3** | **Col4** |
| 1 | 1.0000 | 0.8432 | 0.7110 | 0.5995 |
| 2 | 0.8432 | 1.0000 | 0.8432 | 0.7110 |
| 3 | 0.7110 | 0.8432 | 1.0000 | 0.8432 |
| 4 | 0.5995 | 0.7110 | 0.8432 | 1.0000 |

**Compare heterogeneous compound symmetry versus heterogeneous 1st-order autoregressive**

Heterogeneous compound symmetry and heterogeneous 1st-oder autoregressive models are not nested, thus we can not use LRT to compare them. Since they have the same number of parameters (4 variance paramters+1 correlation parameters), we can compare these two models by looking their maximized REML log-likehood:

Table: Maximized (REML) log-likelihood

|  |  |
| --- | --- |
| Model | -2(REML)Log-likelihood |
| heterogeneous CS | 365.3 |
| Heterogeneous AR(1) | 358.2 |

AIC= -2 maximized REML log-likehood +2 number of parameters

According to AIC criteria, we should select the one that minimizes -2 REML log-likelihood, thus heterogeneous 1st-order autoregressive model is a better fit to the data when compared to heterogeneous compound symmetry.

In summary, heterogeneous 1st-order autoregressive model is the best fit to the data when compared to all other models listed above.

1. **Analysis of Response Profiles:**

**Fit the usual model for the analysis of mean profiles using room air exposure as the reference level for group and baseline as the reference group for time. Use the variance-covariance structure identified in your answer to question 2b. Based on this model:**

Saturated Model for the mean:

Model for the covariance: heterogeneous 1st-order autoregressive

**a. Test the null hypothesis that the pattern of means over hours is identical (coincides) for the two exposure groups. What do you conclude?**

We can use a LRT to test that the mean response profiles coincide.

Table: LRT for coincide profiles

|  |  |  |
| --- | --- | --- |
| Model | -2 Maximized ML log-likelihood | Number of parameters |
| Hour only model | 377.2 | 4 |
| Saturated Model | 369.1 | 8 |
| Difference | 8.1 | 4 |

LRT yields with 4 d.f. (p=0.087983), so we can not reject the null hypothesis and conclude that the model with only time effect is an adequate fit to the data compared to the saturated model with additional group and group\*time effects, i.e., the pattern of means over hours is identical (coincides) for the two exposure groups.

**b.Test the null hypothesis that the mean response profiles of the two groups are parallel. What do you conclude?**

1. We can use a LRT to test that the mean response profiles parallel.

Table: LRT for parallel profiles

|  |  |  |
| --- | --- | --- |
| Model | -2 Maximized ML log-likelihood | Number of parameters |
| Hour only model | 376.8 | 5 |
| Saturated Model | 369.1 | 8 |
| Difference | 7.7 | 3 |

LRT yields with 3 d.f. (p=0.052636), no strong evidence to reject the null.

1. We can also use the multivariate Wald test (Method=REML) to test the hypothesis

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Type 3 Tests of Fixed Effects | | | | | | |
| Effect | **Num DF** | **Den DF** | **Chi-Square** | **F Value** | **Pr > ChiSq** | **Pr > F** |
| hour | 3 | 54 | 9.01 | 3.00 | 0.0291 | 0.0382 |
| group | 1 | 18 | 1.86 | 1.86 | 0.1732 | 0.1900 |
| hour\*group | 3 | 54 | 7.80 | 2.60 | 0.0503 | 0.0615 |

The Wald test for hour\*group interaction term yields a chi-squared statistic 7.8 with 3 d.f. (p=0.0503), no strong evidence to reject the null.

Based on both test results, we can not reject the null hypothesis and conclude that the model with only main effects of time and group is an adequate fit to the data compared to the saturated model with additional group\*time interaction effects, that is, the pattern of means over hours is parallel for the two exposure groups.

1. **Fitting a Linear Model in Time:**

**Fit a model that includes hour as a continuous variable, group and their interaction. Use the model for the variance-covariance structure that you identified in question 2b.**

Linear Trend Model for the mean:

Where

Model for the covariance: heterogeneous 1st-order autoregressive

**a. What is the estimated rate of change in mean response for the room air group?**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Solution for Fixed Effects | | | | | | |
| Effect | **group** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** |
| Intercept |  | 41.8571 | 0.8729 | 18 | 47.95 | <.0001 |
| hour |  | -0.05061 | 0.2217 | 58 | -0.23 | 0.8203 |
| group | **exposed to ozone** | 0.7402 | 1.2345 | 18 | 0.60 | 0.5562 |
| hour\*group | **exposed to ozone** | -0.7947 | 0.3136 | 58 | -2.53 | 0.0140 |

The estimated rate of change in mean response for the room air group is -0.05061 (p=0.8203). Thus it would appear that room air group doesn’t have a significant decline in mean response over hours.

**b. What is the estimated rate of change in mean response for the ozone group?**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Estimates | | | | | |
| Label | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** |
| rate of change for ozone group | -0.8453 | 0.2217 | 58 | -3.81 | 0.0003 |

The estimated rate of change in mean response for the ozone group is -0.05061-0.7947= -0.8453 (p=0.0003). Thus the ozone exposure group has a significant decline in mean response over hours.

**c. Test the hypothesis that the rates of change in mean response are identical in the two groups. What do you conclude? Give a possible reason for any difference in conclusion you make from this test and the analogous test based on the mean profiles analysis from 3b.**

We can use a LRT to test that the mean response profiles coincide.

Table: LRT for coincide profiles

|  |  |  |
| --- | --- | --- |
| Model | -2 Maximized ML log-likelihood | Number of parameters |
| Hour only model | 378.1 | 2 |
| Saturated Model | 371.8 | 4 |
| Difference | 6.3 | 2 |

LRT yields with 2 d.f. (p=0.042852), so we reject the null hypothesis and conclude that the model with only time effect is not an adequate fit to the data compared to the saturated model with additional group and group\*time effects, that is, the pattern of means over hours is not identical (coincides) for the two exposure groups.

This is different from mean profile analysis from 3b. This is the problem of multiple testing. In the response profile analysis, the test for group\*hour interaction is quite general. It posits no specific pattern for the difference in the response profiles between groups. The general test for group\*hour interaction has (2-1)\*(4-1)=3 degrees of freedom, the test simultaneously makes 3 comparisons, it becomes less sensitive to detect a significant effect than a single degree of freedom test. This lack of specificity results in the difference of mean response profile analysis and linear trend model.

**d. What is the estimated difference in rate of mean change between the two groups? By calculating a 95% confidence interval for this difference, identify what are plausible values for the underlying true difference.**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Solution for Fixed Effects | | | | | | | | | |
| Effect | **group** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** | **Alpha** | **Lower** | **Upper** |
| Intercept |  | 41.8571 | 0.8729 | 18 | 47.95 | <.0001 | 0.05 | 40.0232 | 43.6910 |
| hour |  | -0.05061 | 0.2217 | 58 | -0.23 | 0.8203 | 0.05 | -0.4945 | 0.3933 |
| group | **exposed to ozone** | 0.7402 | 1.2345 | 18 | 0.60 | 0.5562 | 0.05 | -1.8533 | 3.3337 |
| hour\*group | **exposed to ozone** | -0.7947 | 0.3136 | 58 | -2.53 | 0.0140 | 0.05 | -1.4224 | -0.1669 |

The mean response for room air group is:

The mean response for ozone group is:

The estimated difference in rate of mean change between the two groups is -0.7947 and there is a 95% chance that the interval ( -1.4224, -0.1669) contains the true underlying difference.

1. **Evaluating the Fit of a Linear Model in Time:**

**Does a model with a linear trend in hour for each exposure group adequately describe the pattern of change in the two groups? Justify your answer with appropriate statistical analysis.**

The adequacy of the linear trend model can be assessed by including higher-order polynomial trends.

First, we compare linear trend model and quadratic trend model. The linear trend model is nested within the quadratic model, we can use LRT.

Table: LRT for comparing Linear Trend vs Quadratic Trend

|  |  |  |
| --- | --- | --- |
| Model | -2 Maximized ML log-likelihood | Number of parameters |
| Linear Trend model | 371.8 | 4 |
| Quadratic Trend Model | 369.9 | 6 |
| Difference | 1.9 | 2 |

LRT yields with 2 d.f. (p= 0.38674), so we can not reject the null hypothesis and conclude that the linear trend model is an adequate fit to the data compared to the quadratic trend model.

Next, we compare linear trend model and cubic trend model. The linear trend model is nested within the cubic model, we can use LRT.

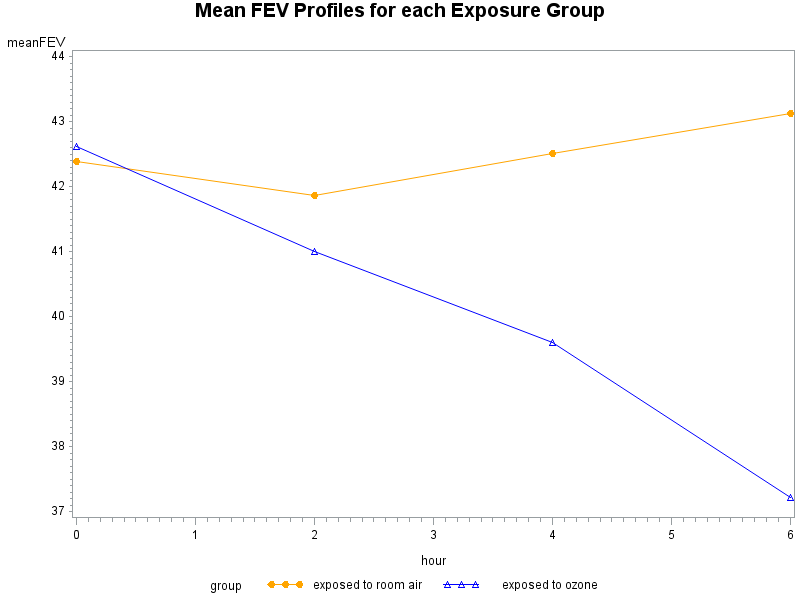
Table: LRT for comparing Linear Trend vs Quadratic Trend

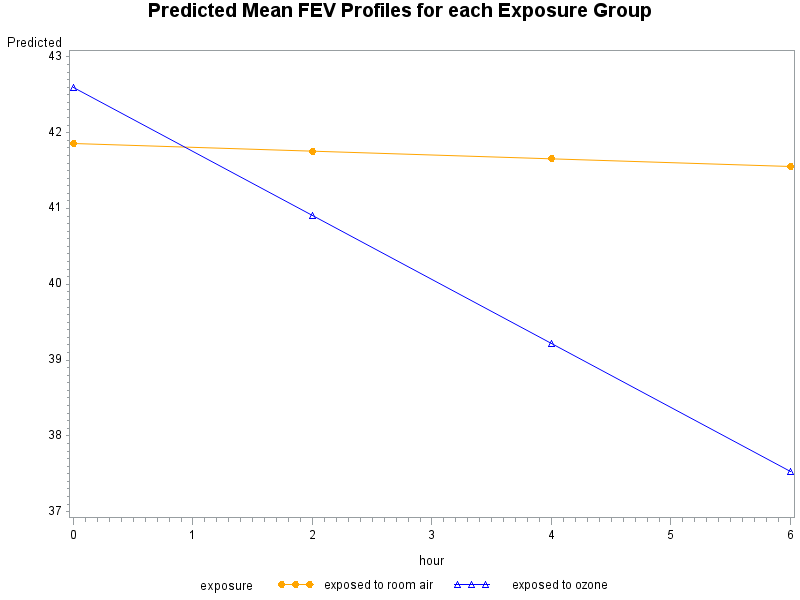
|  |  |  |
| --- | --- | --- |
| Model | -2 Maximized ML log-likelihood | Number of parameters |
| Linear Trend model | 371.8 | 4 |
| Cubic Trend Model | 369.1 | 8 |
| Difference | 2.7 | 4 |

LRT yields with 4 d.f. (p= 0.60921), so we can not reject the null hypothesis and conclude that the linear trend model is an adequate fit to the data compared to cubic trend model.

In summary, the linear trend model appears to be an adequate fit to the data when compared to higher-order polynomial trend models (e.g., quadratic and cubic). We can also assess the linear model by looking at the estimated means and sample means.

|  |  |  |  |
| --- | --- | --- | --- |
| exposure | hour | pred\_mean | sample\_mean |
| exposed to room air | 0 | 41.8571 | 42.383 |
| exposed to room air | 2 | 41.7559 | 41.868 |
| exposed to room air | 4 | 41.6547 | 42.515 |
| exposed to room air | 6 | 41.5535 | 43.124 |
| exposed to ozone | 0 | 42.5973 | 42.617 |
| exposed to ozone | 2 | 40.9068 | 41.009 |
| exposed to ozone | 4 | 39.2163 | 39.600 |
| exposed to ozone | 6 | 37.5258 | 37.211 |

****



1. **Summarizing the Key Results and Conclusions:**

**Write a brief structured abstract (maximum 200 words) summarizing the objective, methods, results and conclusions that might be drawn concerning exposure differences in patterns of pulmonary function over time.**

The objective of the analysis is to determine whether changes in pulmonary function during the 6 hours of exposure were different in the ozone and room air exposed groups. We looked at the mean responses over time for each group, the ozone group appears to have a linear decline trend over hours, while the room air group stays at a stable level. We compared covariance models for repeated measures and concluded that the heterogeneous 1st-order autoregressive is the best fit to the data. To model the mean response, the response profile analysis is adequate since the number of measurement is large relative to the sample size N=20 and it fails to detect discernible differences in systematic patterns of change over time between groups. Finally, we used linear trend model and assessed the model adequacy by comparing it with higher-order polynomial models. We concluded that the estimated rate of change in mean response for the room air group is -0.05061 (p=0.8203) and for the ozone group is -0.8453 (p=0.0003), and estimated difference in rate of mean change between the two groups is -0.7947 and there is a 95% chance that the interval ( -1.4224, -0.1669) contains the true underlying difference.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

SAS Codes Appendix

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*;

\*import dataset;

**proc** **format**;

value exposure

**1**='exposed to ozone'

**0**='exposed to room air';

**run**;

**data** ozone;

infile 'C:\data\Projects\APCD High Cost\Longitudinal\ozone0246.dat';

input id hour group Y;

format group exposure.;

**run**;

**proc** **print** data=ozone;**run**;

\*Q1b:Plot all response profiles on 2 plots(one for each group);

**proc** **gplot** data=ozone;

symbol1 interpol=join value=triangle;

symbol2 interpol=join value=triangle;

symbol3 interpol=join value=triangle;

symbol4 interpol=join value=triangle;

symbol5 interpol=join value=triangle;

plot y\*hour=id;

where group=**0**;

title 'FEV1 response profiles for each person in room air exposire';

**run**;

**proc** **gplot** data=ozone;

symbol1 interpol=join value=triangle;

symbol2 interpol=join value=triangle;

symbol3 interpol=join value=triangle;

symbol4 interpol=join value=triangle;

symbol5 interpol=join value=triangle;

plot y\*hour=id;

where group=**1**;

title 'FEV1 response profiles for each person in Ozone exposire';

**run**;

\*Q1c: Plot mean response profile;

**proc** **means** data=ozone n mean std nway;

var y;

class group hour;

output out=meandata mean=meanFEV;

**run**;

**proc** **gplot** data=meandata;

symbol1 color=orange interpol=join value=dot;

symbol2 color=blue interpol=join value=triangle;

plot meanFEV\*hour=group;

format group exposure.;

title 'Mean FEV Profiles for each Exposure Group';

**run**;

\*Q2a: Fit Saturated model;

**proc** **mixed** data=ozone;

class id hour group;

model y=hour group hour\*group/s chisq;

repeated hour/type=un subject=id r rcorr;

**run**;

\*Q2b:

\*LRT CS vs UN;

**proc** **mixed** data=ozone;

class id hour group;

model y=hour group hour\*group/s chisq;

repeated hour/type=CS subject=id r rcorr;

**run**;

\* chi-squared p-value ;

**data** pvalue;p=sdf('chisquared',**38.9**,**8**);

**proc** **print**;title 'LRT p-value';**run**;

\*LRT CSH vs UN;

**proc** **mixed** data=ozone;

class id hour group;

model y=hour group hour\*group/s chisq;

repeated hour/type=CSH subject=id r rcorr;

**run**;

\* chi-squared p-value ;

**data** pvalue;p=sdf('chisquared',**10.1**,**5**);

**proc** **print**;title 'LRT p-value';**run**;

\*LRT AR(1) vs UN;

**proc** **mixed** data=ozone;

class id hour group;

model y=hour group hour\*group/s chisq;

repeated hour/type=AR(**1**) subject=id r rcorr;

**run**;

\* chi-squared p-value ;

**data** pvalue;p=sdf('chisquared',**25.9**,**8**);

**proc** **print**;title 'LRT p-value';**run**;

\* LRT ARH(1) vs UN;

**proc** **mixed** data=ozone;

class id hour group;

model y=hour group hour\*group/s chisq;

repeated hour/type=ARH(**1**) subject=id r rcorr;

**run**;

\* chi-squared p-value ;

**data** pvalue;p=sdf('chisquared',**3**,**5**);

**proc** **print**;title 'LRT p-value';**run**;

\*Q3a Coincide profiles;

**proc** **sort** data=ozone;by descending group descending hour;**run**;

\* First: fit ML, Full model;

**proc** **mixed** data=ozone order=data method=ml ;

class id hour group;

model y=hour group hour\*group/s chisq;

repeated hour/type=ARH(**1**) subject=id r rcorr;

title'Full model, ML';

**run**;

\* Second: fit ML, Reduced model;

**proc** **mixed** data=ozone order=data method=ml;

class id hour group;

model y=hour /s chisq;

repeated hour/type=ARH(**1**) subject=id r rcorr;

title 'Reduced Model, ML';

**run**;

\* chi-squared p-value ;

**data** pvalue;p=sdf('chisquared',**149.7**,**4**);

**proc** **print**;title 'LRT p-value';**run**;

\*Q3b parallel profiles;

**proc** **sort** data=ozone;by descending group descending hour;**run**;

\* First: fit ML, Full model;

**proc** **mixed** data=ozone order=data method=ml ;

class id hour group;

model y=hour group hour\*group/s chisq;

repeated hour/type=ARH(**1**) subject=id r rcorr;

title'Full model, ML';

**run**;

\* Second: fit ML, Reduced model;

**proc** **mixed** data=ozone order=data method=ml;

class id hour group;

model y=hour group/s chisq;

repeated hour/type=ARH(**1**) subject=id r rcorr;

title 'Reduced Model, ML';

**run**;

\* chi-squared p-value ;

**data** pvalue;p=sdf('chisquared',**7.7**,**3**);

**proc** **print**;title 'LRT p-value';**run**;

\*Wald test using REML;

**proc** **mixed** data=ozone order=data ;

class id hour group;

model y=hour group hour\*group/s chisq;

repeated hour/type=ARH(**1**) subject=id r rcorr;

title'Full model, REML Wald test';

**run**;

\*Q4abd Linear trend model;

**data** ozone;set ozone;hourc=hour;**run**;

**proc** **sort** data=ozone;by descending group ;**run**;

\* First: fit ML, Full model;

**proc** **mixed** data=ozone order=data ;

class id hourc group;

model y=hour group hour\*group/s chisq cl;

repeated hourc/type=ARH(**1**) subject=id r rcorr;

estimate "rate of change for ozone group" hour **1** hour\*group **1** /e;

title'Linear trend model, REML';

**run**;

\*Q4c:coincide profiles;

**proc** **sort** data=ozone;by descending group ;**run**;

\* First: fit ML, Full model;

**proc** **mixed** data=ozone order=data method=ml ;

class id hourc group;

model y=hour group hour\*group/s chisq;

repeated hourc/type=ARH(**1**) subject=id r rcorr;

title'Linear trend model, ML';

**run**;

\* Second: fit ML, Reduced model;

**proc** **mixed** data=ozone order=data method=ml;

class id hourc group;

model y=hour /s chisq ;

repeated hourc/type=ARH(**1**) subject=id r rcorr;

title 'Reduced Model, ML';

**run**;

\* chi-squared p-value ;

**data** pvalue;p=sdf('chisquared',**6.3**,**2**);

**proc** **print**;title 'LRT p-value';**run**;

\*Q5: linear vs quadratic vs Cubic ;

**data** ozone;set ozone;hour2=hour\*hour;hour3=hour2\*hour;**run**;

**proc** **sort** data=ozone;by descending group ;**run**;

\* First: fit ML, Full model;

**proc** **mixed** data=ozone order=data method=ml ;

class id hourc group;

model y=hour group hour\*group/s chisq;

repeated hourc/type=ARH(**1**) subject=id r rcorr;

title'Linear trend model, ML';

**run**;

\* Second: fit ML, Quadratic model;

**proc** **mixed** data=ozone order=data method=ml;

class id hourc group;

model y=hour hour2 group hour\*group hour2\*group /s chisq ;

repeated hourc/type=ARH(**1**) subject=id r rcorr;

title 'Quadratic Model, ML';

**run**;

\* chi-squared p-value ;

**data** pvalue;p=sdf('chisquared',**1.9**,**2**);

**proc** **print**;title 'LRT p-value';**run**;

\* Third: fit ML, Cubic model;

**proc** **mixed** data=ozone order=data method=ml;

class id hourc group;

model y=hour hour2 hour3 group hour\*group hour2\*group hour3\*group/s chisq ;

repeated hourc/type=ARH(**1**) subject=id r rcorr;

title 'Cubic Model, ML';

**run**;

\* chi-squared p-value ;

**data** pvalue;p=sdf('chisquared',**2.7**,**4**);

**proc** **print**;title 'LRT p-value';**run**;

\*Predicted Mean;

**proc** **mixed** data=ozone order=data method=REML ;

class id hourc group;

model y=hour group hour\*group/s chisq outp=predicted;

repeated hourc/type=ARH(**1**) subject=id r rcorr;

title'Linear trend model, REML';

**run**;

**proc** **sort** data=predicted(keep=group hour pred) nodupkey;by group hour;**run**;

\* sample means;

**proc** **means** data=ozone n mean nway;

var y;

class group hour;

output out=meandata mean=mean;

**run**;

**proc** **sql**;

create table temp as

select a.group as exposure, a.hour as hour, a.pred as pred\_mean, b.mean as sample\_mean

from predicted a join meandata b

on a.hour=b.hour and a.group=b.group;

**quit**;

**proc** **print**;**run**;

**proc** **gplot** data=temp;

symbol1 color=orange interpol=join value=dot;

symbol2 color=blue interpol=join value=triangle;

plot pred\_mean\*hour=exposure;

format group exposure.;

title 'Predicted Mean FEV Profiles for each Exposure Group';

**run**;