Julia Wrobel, Victoria Cooley, and Jiayi Ji

Final Project: P8158

Introduction

Uncontrolled hypertension can lead to serious health problems, including heart attack and stroke. White-coat hypertension, in which patients have artificially elevated blood pressure only in a clinical setting, is a common phenomenon.[1] Ambulatory blood pressure (ABP) monitoring, which measures blood pressure (BP) at regular intervals over 24 hours in a non-clinical setting, is the gold standard for hypertension diagnosis. However, BP is still highly variable within subjects throughout the day and across monitoring periods[2]. Therefore a better understanding of BP variability could lead to more accurate hypertension diagnoses.

Our data comes from the masked hypertension study, which collected ambulatory blood pressure data and measured physical activity using accelerometers in order to better understand what factors contribute to high blood pressure. We have 663 subjects, with 20 systolic blood pressure measurements per subject, which were taken at regular intervals over 24 hours. Other covariates include mean heart rate, smoking status, and coffeeDrinker, which indicates whether the subject drinks caffeine. The data also includes a time-varying covariate, activity level, which looks at the intensity of activity of each subject right before each measurement was taken. Descriptive statistics for these fixed-time covariates are given in Table 1. The dataset has a high proportion of caffeine consumers and a low proportion of smokers. The clinical criteria for hypertension diagnosis is mean systolic BP > 135 mm. Based on this criteria, 87 (13.1%) of subjects in the dataset are considered hypertensive.

In order to analyze the data we used latent growth curve modeling. This analysis approach is appropriate because it considers the longitudinal nature of the data, taking into account extra correlation between repeated measurements from the same subject, and allows us to identify potential clusters in the blood pressure profiles. We considered one, two, three, and four class latent growth curve models with fixed time covariates and chose the best model based on the Bayesian Information Criteria (BIC). We then used the best class structure to fit another latent growth curve model with activity level as a time-varying covariate, allowing us to consider the time-dependent effects of activity on blood pressure. Our modeling approach and results are described below.

Methods

Growth mixture models (GMM) were used to see if latent classes existed in our data, describing longitudinal change within each unobserved sub-population, and allowing identification of differential responders in the studies.[3] GMM was used to assign subjects to their most likely class and to obtain estimates of the model parameters for each class. We first considered one, two, three, and four class latent growth curve models with fixed time covariates, including heart rate, smoker and coffeeDrinker. A second run was performed with different starting values to reduce the chance of local maxima, which would give a false sense of the number of latent classes in the data. The BIC was the primary fit statistics used to determine the number of classes that best fit the data.[4] For a practical reference, smaller BIC values are preferred when choosing the number of latent classes. Very small classes may represent chance findings and thus give a false indication of the number of latent classes within the heterogeneous data. In addition, visual inspection of the latent class trajectories provided insight into the reasonableness of the numbers of latent classes to be considered. Once classes were extracted, class ordering and separation were explored as a final check on the appropriateness of the number of latent classes identified. The number of latent classes extracted were based on a review of the information from fit statistics, the smallest class size, and a visual inspection of the slope trajectories by latent class. We also added a time-varying covariate to the best class model chosen, which allowed us to consider the time-dependent effects of activity on blood pressure and determine whether the same number of classes would be extracted. Descriptive statistics were conducted in R version 3.2.3. Growth mixture models were conducted using Mplus version 7.0.

Results of Primary Analysis

As mentioned in the methods section of this paper, the final latent growth curve model with fixed time covariates for the masked hypertension study was selected by assessing the BIC value and the proportion of subjects assigned to a given class. The final class counts and proportions (for the latent classes based on their most likely latent class membership) and the BIC values for each model were located on the Mplus output and placed into tables (see the appendix section) for the ease of visualization. Based on the information provided in tables 2, 3, and 4, the BIC values for the one-class, two-class, three-class, and four-class models are 96966.218, 96958.627, 96965.652, and 96997.590 respectively.

Due to the fact that the BIC for the two-class model (96958.627) is the lowest out of the four models, it is likely that this model provides a better empirical fit to the data. In addition, the four-class model produced an undesirable class size for the first and third latent classes (0% of subjects were assigned to these classes). Given that the major goal of this analysis was to identify the distinct number of classes that could be developed from this data, the one-class model was also deemed undesirable as individuals were all placed in the same, unified class. While the two-class and three-class models displayed similar proportions of subjects in each class, the two-class model was ultimately selected as its BIC value (96958.627) was slightly lower than that for the three-class model (96965.652).

Based on the above discussion, the subjects from the masked hypertension study can be best categorized into two latent classes (low and high blood pressure groups). The majority of subjects are organized into the first latent class (low blood pressure group) of this two-class model (91.53%) and the remaining 8.47% are organized in the second latent class (high blood pressure group). The corresponding slopes and intercepts along with their p-values for the first and second latent classes are -0.003 ($p = 0.001$), 122.977 ($p = 0.000$) and 0.004 ($p = 0.185$), 136.883 ($p = 0.000$) respectively. Based on the different mean intercepts mentioned in the previous sentence and their corresponding p-values, (the slopes, however, are nearly identical), there is evidence that subjects in the first latent class can potentially represent the low blood pressure group while subjects in the second latent class may be more representative of the high blood pressure group.

The results from the 2-class model are visualized in Figures 1 and 2. These figures highlight differences in blood pressure patterns for subjects from each of the classes. Figure 1 shows the distribution of the subject-specific intercepts from the two-class analysis. On average, subjects in class 1 have a smaller intercept than subjects in class 2, indicating that our analysis may be successful in separating out subjects who are more prone to hypertension, though this separation is not perfect. Adding in the subject-specific slopes on the y-axis (Figure 2), we now get a good separation between classes. Subjects in class 2 tend to have a higher intercept or a higher slope.

Results of Secondary Analysis

After adding the activity time-varying covariate to the two-class model, we notice that 90% of the subjects are organized into the first group and the remaining 10% are organized into the second group (Table 5). The division of subjects is almost the same as in the two-class model that was developed for the fixed time covariates.

Conclusion

In our primary analysis, which excluded consideration of activity as a time-varying covariate, we chose amongst one, two, three, and four - class latent growth curve models. Based on BIC, we chose the two-class model. Analysis of the different classes indicate that subjects are being divided into classes based on the magnitude of their random slopes and intercepts.

Appendix: Tables and Figures.

Covariate	**Summary**
Heart Rate	77 (47, 107)
Smoker	59 (8.9)
Coffee Drinker	544 (82.1)
Activity Level	9 (0, 2972)
Systolic Blood Pressure	123 (50, 203)

Table 1. Baseline characteristics of subjects in the masked hypertension study. For continuous covariates we report median (min, max) and for categorical covariates we report frequency (%).

	One Class Model (K=1)	Two Class Model (K=2)	
	Latent Class 1	*Latent Class 1*	*Latent Class 2*
Mean Slope (S), p-value	-0.003, 0.000	-0.003, 0.001	0.004, 0.185
Mean Intercept (I), p-value	124.476, 0.000	122.977, 0.000	136.883, 0.000
Proportion in Each Class	100%	91.53%	8.47%
Bayesian (BIC)	96966.218	96958.627	

Table 2. Representation of the intercepts and slopes in the different classes (with corresponding significance values), the proportion of individuals in each class, and the BIC value, for a latent growth curve model for the masked hypertension study with K=1 and K=2 classes

	Three Class Model (K=3)		
	Latent Class 1	*Latent Class 2*	*Latent Class 3*
Mean Slope (S), p-value	-0.002, 0.011	-0.021, 0.000	0.005, 0.182
Mean Intercept (I), p-value	121.920, 0.000	141.964, 0.000	137.051, 0.000
Proportion in Each Class	88.81%	4.09%	7.11%
Bayesian (BIC)	96965.652		

Table 3. Representation of the intercepts and slopes in the different classes (with corresponding significance values), the proportion of individuals in each class, and the BIC value, for a latent growth curve model for the masked hypertension study with K=3 classes

	Four Class Model (K=4)			
	Latent Class 1	*Latent Class 2*	*Latent Class 3*	*Latent Class 4*
Mean Slope (S), p-value	-0.003, 0.001	-0.003, 0.001	-0.003, 0.001	0.004, 0.185
Mean Intercept (I), p-value	122.977, 0.000	122.977, 0.000	122.977, 0.000	136.883, 0.000
Proportion in Each Class	0%	90.32%	0%	9.68%
Bayesian (BIC)	96997.590			

Table 4. Representation of the intercepts and slopes in the different classes (with corresponding significance values), the proportion of individuals in each class, and the BIC value, for a latent growth curve model for the masked hypertension study with K=4 classes

	Two Class Model (K=2)	
	Latent Class 1	*Latent Class 2*
Mean Slope (S), p-value	0.000, -------	0.010, -------
Mean Intercept (I), p-value	100.846, -------	125.793, -------
Proportion in Each Class	90.00%	10.00%
Bayesian (BIC)	-------	

Table 5. Representation of the intercepts and slopes in the 2 classes, the proportion of individuals in each class, for a latent growth curve model with time-varying covariate for the masked hypertension study with K=2 classes (The M-convergence criterion of the EM algorithm is not fulfilled.)

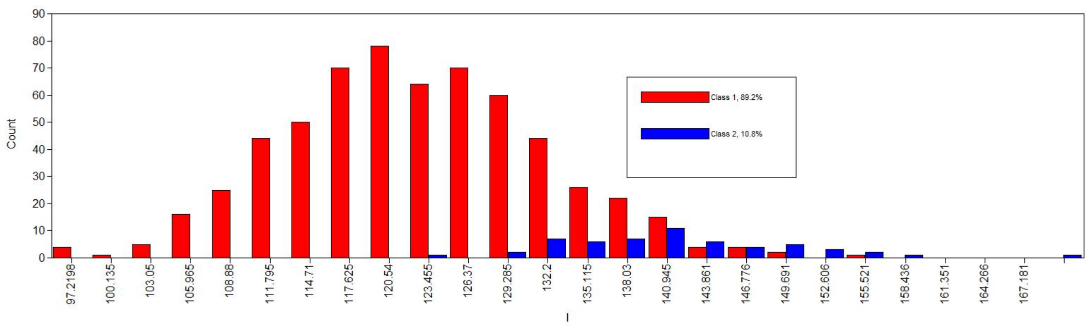

Figure 1. Histogram of subject-specific random intercepts from 2-class model. Intercepts for subjects from class 1 are in red and class 2 are in blue.

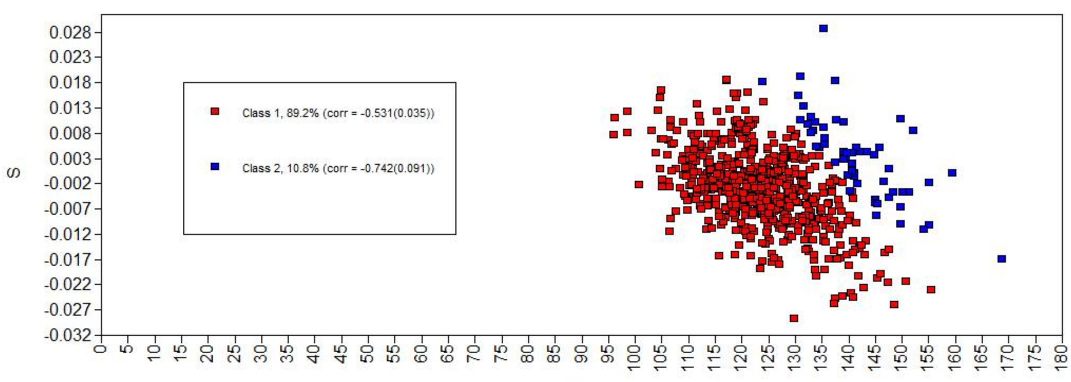

Figure 2. Scatterplot of subject-specific slopes vs. intercepts

References

1. Pickering TG. (1998), White Coat Hypertension: Time for Action. *Circulation*, 98: 1834-1836.

2. Verdecchia P, Angeli F, Gattobigio R (2004), Clinical Usefulness of Ambulatory Blood Pressure Monitoring. *JASN*, 15(1), S30-S33.

3. Muthen, B., C.H. Brown, K. Masyn, B. Jo, S.T. Khoo, C.C. Yang, *et al.* General growth mixture modeling for randomized preventive interventions Biostatistics, 3 (4) (Dec 2002), pp. 459–475

4. Tofighi, D, C.K. Enders. Identifying the correct number of classes in growth mixture models, G.R. Hancock, K.M. Samuelsen (Eds.) Advances in latent variable mixture models (2008), pp. 317–341