

# LA Nurse BP

Case Study 1\*

Kunwu Lyu

Katie St. Clair, STAT 330  
February 13, 2025

## Executive Summary

Here is what I did.

## 1 Introduction

Certain traits such as family history and mood are expected to increase one's ambulatory blood pressure (BP). Goldstein and Shapiro (2000) studied potential factors that contribute to hypertension. They collected information about the participants' BP, activity levels, work status, and mood ratings throughout the day, as well as relevant family history and information about their menstrual phases, to establish links that lead to elevated BP. Towards the end, they sought to uncover preventative measures for individuals who may be at a higher risks of developing hypertension. The objectives of this project is much simpler. I am interested in exactly what traits are *associated* with elevating one's BP, given the longitudinal structure and various time-dependent metrics in the dataset (given by Goldstein & Shapiro, 2000; cited by Roback & Legler, 2021 ).

### 1.1 Methods

The dataset includes repeated measures over the course of two work and off-work days on 203 registered nurses between the ages of 24 and 50 years working in Los Angeles, in the year 2000. Of those 203 nurses, 172 has complete data on all of the variables recorded.<sup>1</sup> BP of the participants were measured 30 minutes before their normal start of work, and measured repeatedly every 20 minutes for the rest of the day. This led to around 40-60 observations per nurse (9573 total observations). Each time the BP was taken, participants were asked to give several mood ratings including happiness, stress, and tiredness. In addition, participants wore an actigraph on their waist to record frequency of movements in one-minute intervals; the researchers obtained an activity measure for the ten-minute periods before each BP reading.

---

\*Code used for this project is available at this [GitHub Repository](#).

<sup>1</sup>In the original paper by Goldstein and Shapiro (pp. 228–29), they reported 171 nurses who completed all sessions. It is likely that this is due to Roback and Legler (2021) or Weiss (2005) excluding other personality variables used in the original study. Because I am not investigating personality traits in this project, 172 suffices being my total subject counts. Additionally, Goldstein and Shapiro mention that “[s]imilar patterns of findings were obtained in the sample of 171 as in the total sample” (p. 229).

## 2 Exploratory Data Analysis

The variables (original and re-parameterized) used in this project are given in [Table 1](#). Variable missingness was explored in [Figure 1](#). We see that the data were missing primarily mood ratings and some activity levels from several participants. Because I am interested in how both of these variables relate to BP, I will proceed the analysis with the missing rows removed, leaving us with a total of 172 participants and 7877 observations. The visuals and statistics below will be given in terms of that subset of the data, unless otherwise specified.

Variable Name	Variable Description
ID (Cluster)	Unique identification number for each participant
BP (Response)	Systolic blood pressure (in mmHg)
<i>Act</i>	Activity level (frequency of movements in 1-minute intervals, over a 10-minute period)
<i>Phase</i>	Menstrual phase (follicular or luteal)
<i>Day</i>	Workday or non-workday
<i>Posture</i>	Position during BP measurement (sitting, standing, or reclining)
<i>HAP</i>	Self-ratings of happiness by each nurse at the time of each BP measurement on a 5-point scale (5 strongest and 1 weakest)
<i>STR</i>	Self-ratings of stress by each nurse at the time of each BP measurement on a 5-point scale (5 strongest and 1 weakest)
<i>TIR</i>	Self-ratings of tiredness by each nurse at the time of each BP measurement on a 5-point scale (5 strongest and 1 weakest)
<i>Age</i>	Age (in years)
<i>Full FH</i>	Family history, coded as either NO (no family history of hypertension), YES (1 hypertensive parent), or YESYES (both parents hypertensive)
<i>Stand</i>	Indicator variable for standing, where it equals 0 if Posture is either sitting or reclining and 1 when Posture is standing
<i>Mood</i>	Combined mood ratings: $HAP - (STR + TIR)/2$
<i>FH</i>	Indicator variable for having family history of hypertension (1 for YES and YESYES, 0 for NO)
<i>Age24</i>	Recentered Age: $Age - 24$

Note: *emphasis* added to re-parameterized variables; colors represent levels.

Table 1: Variable Descriptions

Because of the longitudinal nature of our data, it is worth noting our levels of analysis and to which our variables belong. I will call time-dependent measurements *Level 1* data (variables) and the rest of the subject-level measurements *Level 2* data (variables). Each participant has a unique identification number; this will distinguish between different clusters. Our primary response is systolic blood pressure (BP), and I will use our exploratory data analysis below to guide our model selection process from the bottom-up.

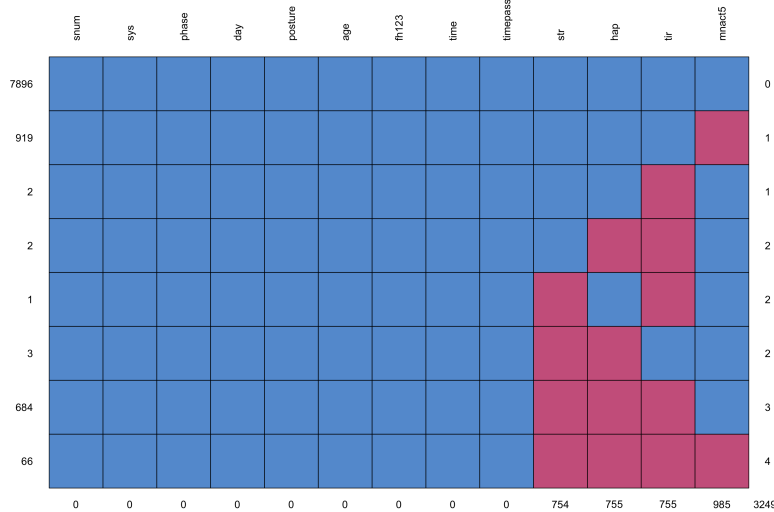


Figure 1: Data Missingness

Note that I created three additional variables from the original dataset to help simplify the modeling process. In the original dataset, **Posture** is a factor variable with three levels: Recline ( $n = 530$ ), Sit ( $n = 3644$ ), and Stand ( $n = 3703$ ). Because the relatively small number of observations where participants were reclining (presumably sleeping at night or resting), I collapsed the three levels into either Standing or non-Standing. Similarly, because of the relatively small number of participants have both parents with hypertension ( $n = 13$ ), I collapsed it together with having one parent with hypertension ( $n = 66$ ) to compare against those without any family history ( $n = 103$ ). Lastly, a general **Mood** measurement was created by subtracting the average of tiredness and stress from happiness.

## 2.1 Level 1 by Clusters

I will first explore how BP vary among participants and how it relates to the level 1 predictors to examine if there is any sign of clustering between different participants (which is the premise of a longitudinal study). From [Figure 2](#), we see that BP readings do tend to vary among different participants, though variations within each participant do not seem to be pronounced. This suggests that a Linear Mixed Model (LMM) might be more appropriate than a regular Multiple Linear Regression (MLR).

Since our data is ordered in time, one might be interested in answering how the participants' BP evolve over time. For sake of brevity, I show a random subsets of subjects and their BP readings over time in [Figure 3a](#). From this plot alone, not much could be deciphered; some trends (e.g., participant 1116) seem to decrease over time while others either slightly increase (e.g., participant 1267) or do not see much fluctuation overall. Indeed, when I fit separate

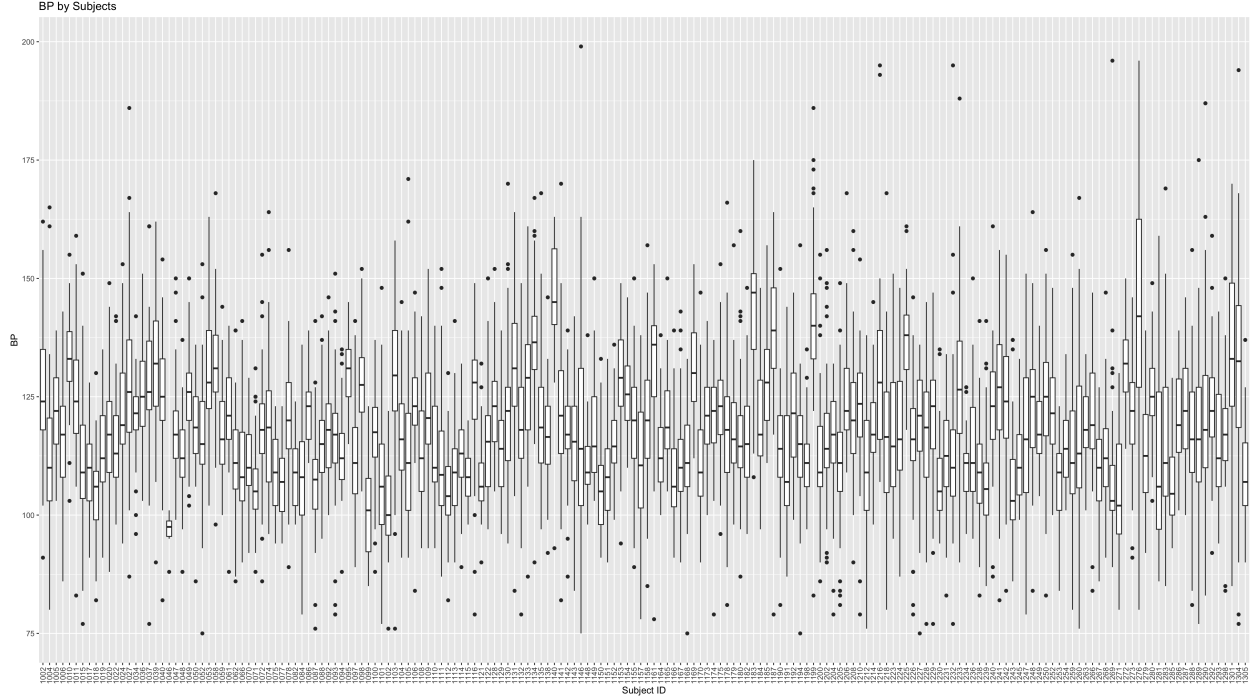


Figure 2: BP by Subjects

linear smoothers for each participant, I see various different intercepts and slopes for each participant (see [Figure 3b](#)). This suggests that a random effect for both the intercept and slope of time.

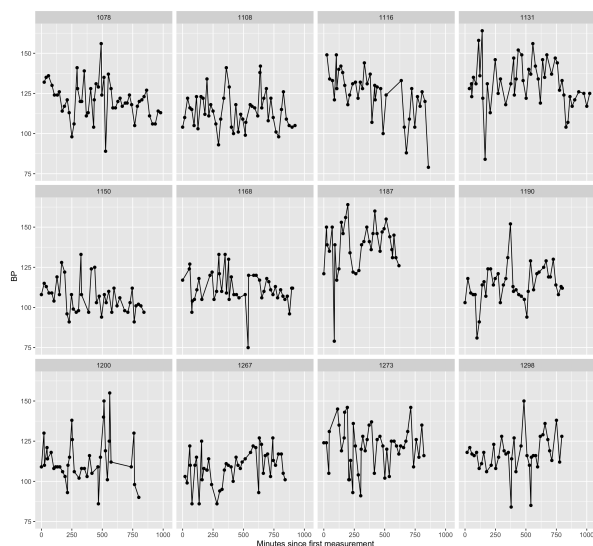
We perform similar analysis for other level 1 covariates (see [Figure 4](#)). We see the most prominent effect in **Mood**, where the slopes and intercepts for each participant differs the most. They all show, to a certain extend, signs of varying slopes and intercepts, but I cannot definitely conclude anything at this point.

## 2.2 Level 2 Covariates

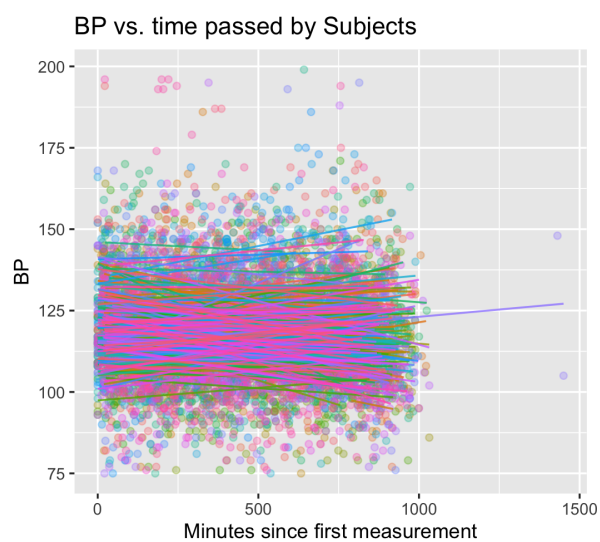
We then want to understand the distributions of our level two covariates before we look at any interactions between the two levels. Because our level 2 covariates contains the same measurement over time (whereas our primary response **BP** is level 1 and hence time-dependent), we will compare the level 2 covariates with the *average* BP of a given participant.

## 2.3 Level 1 by Level 2 Covariates

Next we investigate whether our level 1 covariates depend, in any way, on our level 2 covariates.

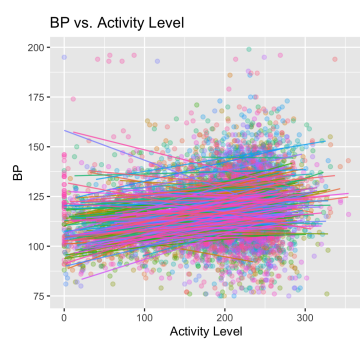


(a) BP vs. Time Passed by Subjects

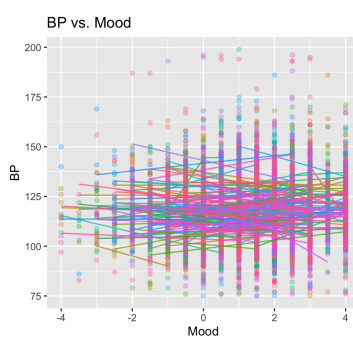


(b) BP vs. Time Passed Separate Lines

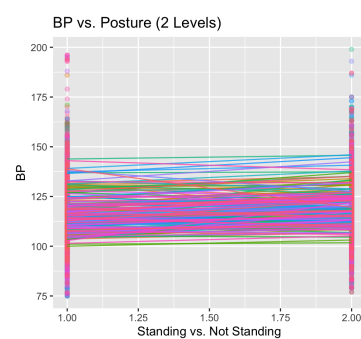
Figure 3: BP vs. Minutes since first measurement



(a) BP vs. Activity Level by Subjects

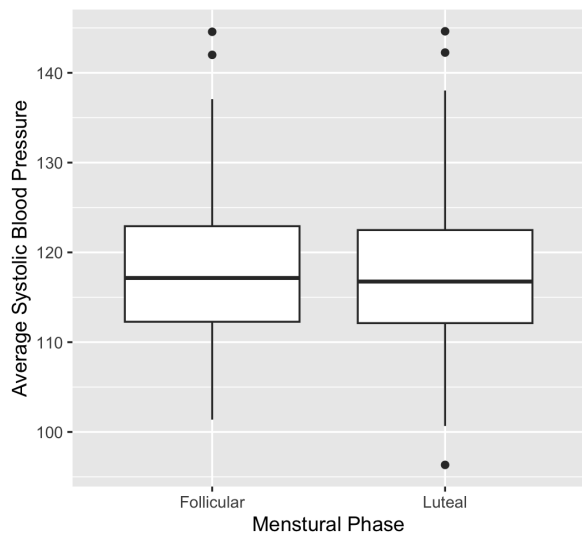


(b) BP vs. Mood by Subjects

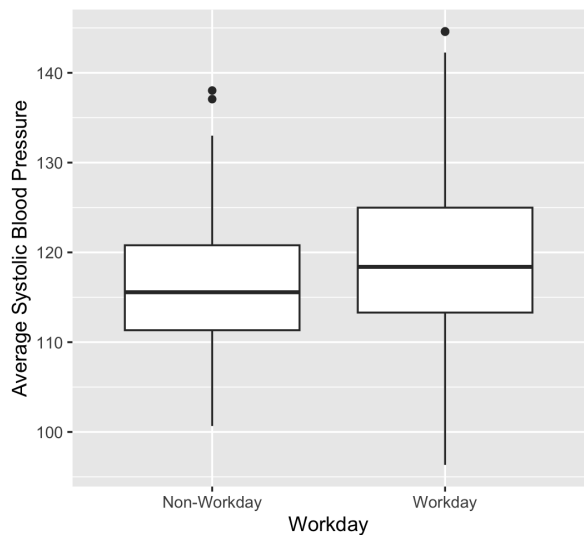


(c) BP vs. Posture by Subjects

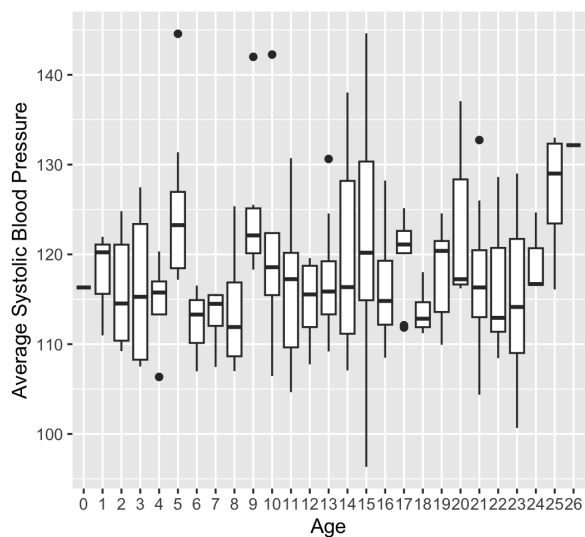
Figure 4: BP vs. Level 1



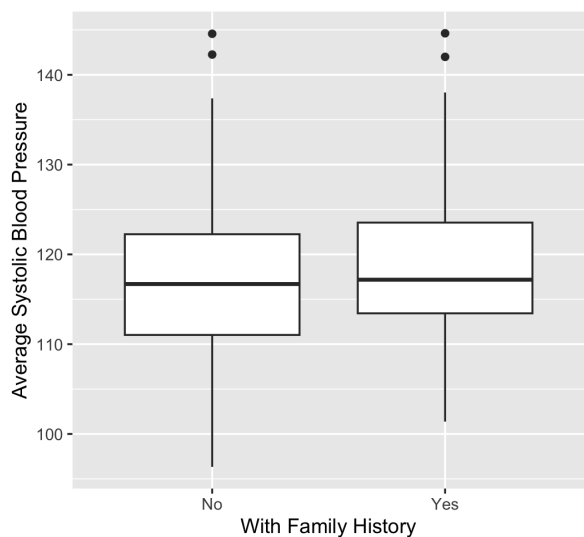
(a) Avg BP vs. Menstrual Phase



(b) Avg BP vs. Workday



(c) Avg BP vs. Age



(d)

Figure 5: Average BP vs. Level 2 Covariates

### 3 Model Selection

### 4 Results

### References

- Goldstein, I. B., & Shapiro, D. (2000). Ambulatory blood pressure in women: Family history of hypertension and personality [Publisher: Taylor & Francis \_eprint: <https://doi.org/10.1080/713690197>]. *Psychology, Health & Medicine*, 5(3), 227–240. <https://doi.org/10.1080/713690197>
- Roback, P., & Legler, J. (2021). *Beyond multiple linear regression* (1st ed.). Chapman; Hall/CRC. Retrieved February 10, 2025, from <https://bookdown.org/roback/bookdown-BeyondMLR/>
- Weiss, R. E. (2005). *Modeling longitudinal data*. Springer. <https://doi.org/10.1007/0-387-28314-5>