

# SURVMETH 687/SURV 617

Homework 1 - Due September 23, 2025

John Kubale and Brady West

2025-08-31

## Overview of data

In a subset of data from the Health and Retirement Study, 311 participants from the “War Babies” cohort (born 1942-1947) had their systolic blood pressure checked every 4 years. This subset was taken from the RAND HRS Longitudinal File 2022. The dataset contains the following variables:

**HHIDPN:** Six character person ID, combining HRS Household ID and HRS person number

**bp\_sys:** Systolic blood pressure (mmHg)

**wave:** HRS wave

**age\_y:** Age in years at a given wave

**year:** Numeric measure of years since first blood pressure measurement

**RAGENDER:** 1 = male, 2 = female

## Study objectives

1. On average, how does systolic blood pressure change over time as people get older?
2. How much variance among participants is there in the trajectories of systolic blood pressure, and can this variance be explained by sex and age?

## Questions

Complete an analysis as described below to address the study objectives using your judgement in order to keep things as simple as possible.

1. Describe the sample in a reasonable way, i.e., construct a table or a graph of descriptive statistics as you may do for a publication. Be sure to remember that there is more than one observation per subject. Comment briefly on the general characteristics of the observations. Keep the study objectives in mind with your descriptive presentation.
2. Fit a multilevel model with systolic blood pressure as the dependent variable and time (year) as the independent variable to explore how systolic blood pressure may have changed over time. Make sure to account for the within-subject correlation in observations on the dependent variable in the multilevel model that you are fitting. Be sure to clearly describe your reasoning.
3. Write out the model that you fit in question 2 using mathematical notation.
4. Interpret the model parameters estimated in part 2, including any estimated variance components. Frame your interpretations with respect to addressing the two research objectives.

5. Briefly describe the steps that one would use to test the null hypothesis that the variance of random year effects is equal to zero, including the computation of the p-value for this test. What is the conclusion of this test as it applies to the model fit in question 2 with respect to the research objectives?
6. Modify the model to account for the sex and age effect on systolic blood pressure.
7. Interpret the fitted multilevel model from question 6 in practical terms.
8. How might one recode the age variable to make the parameters more interpretable? How would this recode change the interpretation of the parameters from 7? Should age and year both be included in the model (be sure to describe your reasoning)?
9. Which model (question 2 or 6) appears to have a better fit? Make use of some simple model diagnostics to help justify your choice.
10. Write a  $\frac{1}{2}$  page summary of the methods and results of the above analyses regarding the study objectives.