Stat 301 Project

Shan Sharif, and Leonard Lange

**Introduction**

This paper will analyze a subset of data from the Framingham Heart Study, a long-term prospective study on the etiology of cardiovascular disease among individuals in Framingham, Massachusetts. The dataset includes laboratory, clinic, questionnaire, and adjudicated event data for 4,434 participants over three examination periods from approximately 1956 to 1968. There were 3 examinations periods, approximately 6 years apart. Each participant was followed for 24 years.

The primary research question is to understand the relationship between various risk factors and the occurrence of cardiovascular events (Angina Pectoris, Myocardial Infarction, Atherothrombotic Infarction, Cerebral Hemorrhage, or death) over a 24-year follow-up period.

Variables include risk factors like blood pressure, blood chemistry, lung function, smoking history, health behaviors, and medication use. Event data indicates whether specific cardiovascular events occurred (0 = did not occur, 1 = did occur).

We created new columns based on the following characteristics as well, agegroup and pre\_cardio\_event (history of cardiovascular events). We will focus on the two created columns as well as educ and sex.

The test that we have chosen is a linear mixed-effects model which is appropriate for analyzing continuous dependent variables with fixed and random effects. The choice of lmer() function is suitable for handling our design with fixed and random effects, ideal for your multifactorial setup and addressing individual variation through randid.

* Null Hypothesis (H0): There is no effect of education level, age group, sex, pre cardio event and their interaction on BMI, after controlling for individual random effects.
* Alternative Hypothesis (H1): There is an effect of education level, age group, sex, pre cardio event and their interaction on BMI, after controlling for individual random effects.

We are looking into the data to determine if education level, what level of education they have achieved, sex, if they are female or male, age group, which age group they are in, or pre cardio event, have they had a previous at-risk event, has a significant effect on BMI.

Summary Statistics Table 1

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Stats** | **sex** | **age** | **bmi** | **heartrte** | **prechd** | **prevmi** | **prevstrk** |
| N | 11627 | 11627 | 11575 | 11621 | 11627 | 11627 | 11627 |
| Mean | 1.57 | 54.79 | 25.88 | 76.78 | 0.07 | 0.03 | 0.01 |
| Std Dev | 0.50 | 9.56 | 4.10 | 12.46 | 0.26 | 0.18 | 0.11 |
| Min | 1 | 32 | 14.43 | 37 | 0 | 0 | 0 |
| Max | 2 | 81 | 56.8 | 220 | 1 | 1 | 1 |
| Range | 1 | 49 | 42.37 | 183 | 1 | 1 | 1 |
| N Missing | 0 | 0 | 52 | 6 | 0 | 0 | 0 |
| Variance | 0.25 | 91.48 | 16.83 | 155.34 | 0.07 | 0.03 | 0.01 |
| Median | 2 | 54 | 25.48 | 75 | 0 | 0 | 0 |

Table 2 Categorical variables



**Methods**

In our statistical analysis, we used Analysis of Variance (ANOVA), nested F-tests, Tukey's Honestly Significant Difference (HSD) test, and General Linear Hypotheses Testing (GLHT) to evaluate the significance of the effects of independent variables on the dependent variables across the different hypothesized models. These methods were selected for assessing the impact and interaction between factors within the dataset. We chose the \*\*(4-way mixed ANOVA) since we have a numerical dependent variable and 4 categorical independent variables. We are checking for significant interactions between the variables. \*\*(don’t want too significant of an interaction.) A limitation of the 4-way ANOVA is that it can create too many interactions between the variables and lead to false positives. Another potential issue with the study is that it is voluntary response and could lead to some significant errors. Also, it is a generational issue, where the participants could have opted out, moved, passed away among other scenarios.

**Conclusion**

When reviewing the analysis of the F values for the 4-way mixed ANOVA there were significant interactions between education and age group as well as sex and age group. (see table 3) To see where the interaction may be developing from we performed 2 more ANOVAs with where we isolated men and women within the test (3-way ANOVA). When isolating for sex there was then no significant interaction between education and age group. This means that sex is the interacting piece between the different independent variables. (see tables 4 and 5)

When reviewing the 4-way results and looking at each independent variable in isolation it is noted that each variable has a significant affect on BMI. Based on the p values for each independent variable we would reject the null hypothesis and can conclude that education, sex, age group, and pre-existing cardio conditions each play a significant affect. Sex also has an interacting affect on each of these. When sex is removed, and we look at only males then age group and pre-existing cardio conditions have a significant effect on BMI for both men and women. In contrast education level does not have a significant effect on BMI for men but it does for women.

The piece that was possibly most interesting was that each of these independent variables had different outcomes between men and women’s BMI. It was believed at the start that it would have a main affect on both men and women, but it does not.

There were a variety of limitations with this study. The study was a voluntary study over generations; thus, participants could opt out at any time. For each individual there were also missing data points which resulted in a smaller sample size, but the sample size was still large enough to be valid.

Table 3: 4 way ANOVA results

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Independent Variables | npar | Sum Sq | Mean sq | F-value |
| educ | 3 | 234.31 | 78.103 | 39.8374 |
| pre\_cardio\_event | 1 | 128.86 | 128.862 | 65.7282 |
| sex | 1 | 28.67 | 28.669 | 14.6231 |
| agegroup | 5 | 556.72 | 111.343 | 56.7922 |
| educ:pre\_cardio\_event | 3 | 11.19 | 3.728 | 1.9017 |
| educ:sex | 3 | 103.36 | 34.454 | 17.5736 |
| educ:agegroup | 14 | 31.21 | 2.229 | 1.137 |
| pre\_cardio\_event:sex | 1 | 0.08 | 0.078 | 0.0396 |
| pre\_cardio\_event:agegroup | 4 | 11 | 2.751 | 1.403 |
| sex:agegroup | 5 | 87.91 | 17.582 | 8.968 |
| educ:pre\_cardio\_event:sex | 3 | 4.52 | 1.508 | 0.7692 |
| educ:pre\_cardio\_event:agegroup | 10 | 12.87 | 1.287 | 0.6566 |
| pre\_cardio\_event:sex:agegroup | 3 | 6.08 | 2.028 | 1.0343 |
| educ:sex:agegroup | 12 | 33.98 | 2.832 | 1.4445 |
| educ:pre\_cardio\_event:sex:agegroup | 6 | 31.31 | 5.219 | 2.6618 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Independent Variables for Men** | **npar** | **Sum Sq** | **Mean sq** | **F-value** |
| educ | 3 | 8.207 | 2.736 | 1.7301 |
| pre\_cardio\_event | 1 | 85.253 | 85.253 | 53.9118 |
| agegroup | 5 | 141.493 | 28.299 | 17.8954 |
| educ:pre\_cardio\_event | 3 | 11.609 | 3.87 | 2.4472 |
| educ:agegroup | 12 | 34.745 | 2.895 | 1.831 |
| pre\_cardio\_event:agegroup | 4 | 4.585 | 1.146 | 0.7248 |
| educ:pre\_cardio\_event:agegroup | 10 | 29.905 | 2.99 | 1.8911 |

Table 3 – Isolated ANOVA for Women

Table 4 – Isolated ANOVA for Men

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Independent Variables for women** | **npar** | **Sum Sq** | **Mean sq** | **F-value** |
| educ | 3 | 300.32 | 100.108 | 44.6483 |
| pre\_cardio\_event | 1 | 50.27 | 50.268 | 22.4195 |
| agegroup | 5 | 506.08 | 101.217 | 45.1429 |
| educ:pre\_cardio\_event | 3 | 5.16 | 1.72 | 0.7671 |
| educ:agegroup | 14 | 29.1 | 2.079 | 0.9271 |
| pre\_cardio\_event:agegroup | 3 | 8.77 | 2.925 | 1.3044 |
| educ:pre\_cardio\_event:agegroup | 6 | 13.75 | 2.291 | 1.0219 |