BIOS 6643 Final Project

Samantha Bothwell | November 17th, 2020

1 Introduction

Individuals who classified as overweight or obese were enrolled into the study to understand factors that contribute to weight loss. Participants were asked to step on a bluetooth scale once a day over the course of the study. Within the study, there are 3 cohorts. These cohorts indicate participants who started the study around the same time.

The research questions of interest are :

- What is the trajectory of weight over the duration of time in the study?
- Is there a relationship between month of study and weight loss, when accounting for sex and race?

2 Methods

Data Cleaning

The data received was already cleaned for prior analysis.

Subjects were asked to step on the Bluetooth scale daily. For subjects who missed a day, we added that missing day to the dataset. This allows us to use an AR(1) covariance structure in our mixed model and allows us to fit functional principle components analysis (FPCA) to the data. To analyze trajectories over a year, we truncated the data to 365 days. To account for the effect of month on weight loss, we used the date variable to determine which month the measurement fell into.

Data Analysis

Data cleaning and visualization will be performed in R, version 4.0.2 (The R Foundation, Vienna University). Longitudinal modeling will be performed in SAS 9.4 (SAS Institute Inc., Cary, NC, USA).

To understand the trajectory of weight over the duration of time in the study we will use the fpca.sc method for regular data in the R refund package, version 0.1-21. We will summarize the results of the mean trajectory for each cohort.

To assess if a relationship exists between month of study and weight loss, when accounting for sex and race, we will fit a linear mixed model in SAS with PROC MIXED. Since time points are equally spaced, we will use an AR(1) covariance structure.

3 Results

Table 1, below, displays the cohort characteristics at baseline stratified by Sex. Categorical variables are represented as N(%) and continuous variables are represented as Median [IQR] for consistency and because the data is not normally distributed. Of the 4,434 participants in the study, 2,490 are female and 1,944 are male. Overall, 447 (10.1%) subjects had their last follow-up at period 1, 724 (16.3%) subjects had their last follow-up at period 2 and the

remaining 3,263 (73.6%) had 3 periods of follow-up. A higher proportion of men were current smokers at their first visit compared to women (60.4% compared to 40.4%). Men and women had the same proportion of subjects who did not experience a stroke (96.8%).

Table 1. Cohort Characteristics

	Cohort 1 (N=29)	Cohort 2 (N=26)	Cohort 3 (N=36)	Overall (N=91)
Age				
Mean (95% CI)	42 (27.13, 55.1)	42 (22.21, 54.7)	44 (22.96, 54.47)	42 (22.55, 54.89)
Missing	0 (0%)	1 (3.8%)	0 (0%)	1 (1.1%)
Sex				
Female	21 (72.4%)	19 (73.1%)	29 (80.6%)	69 (75.8%)
Male	8 (27.6%)	6 (23.1%)	7 (19.4%)	21 (23.1%)
Missing	0 (0%)	1 (3.8%)	0 (0%)	1 (1.1%)
Race				
Asian	0 (0%)	1 (3.8%)	4 (11.1%)	5 (5.5%)
Black	0 (0%)	3 (11.5%)	4 (11.1%)	7 (7.7%)
White	26 (89.7%)	21 (80.8%)	28 (77.8%)	75 (82.4%)
Other	3 (10.3%)	0 (0%)	0 (0%)	3 (3.3%)
Missing	0 (0%)	1 (3.8%)	0 (0%)	1 (1.1%)
Baseline Weight				
Mean (95% CI)	210 (167.07, 277.78)	210 (156.47, 293.77)	220 (163.28, 292.97)	210 (158.57, 291.95)
Total Measures				
Mean (95% CI)	310 (127.6, 525.4)	310 (97, 459.25)	130 (35.88, 212.5)	240 (42.25, 487.25)
Time Span				
Mean (95% CI)	550 (324.8, 600)	410 (233.38, 433)	180 (144.25, 202)	360 (159.25, 600)

Figure 1. Weight Trajectories over time

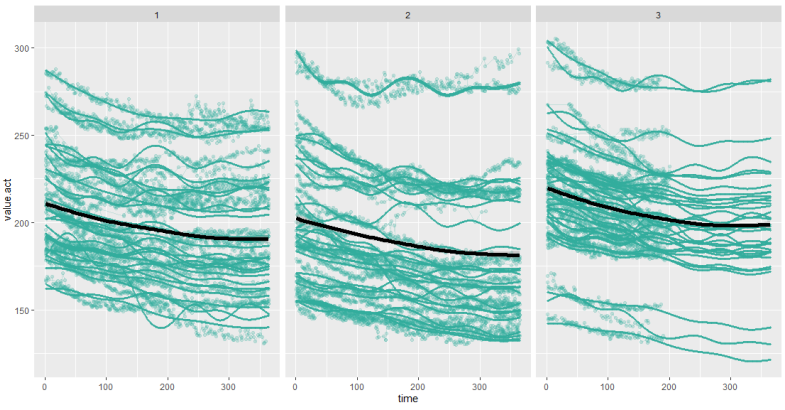

Figure 1 shows the actual weights vs the smoothed FPCA line for all individuals in their respective cohort. From the curves, there is a general decline in weights over time. The mean trajectories for cohorts 1 and 2 are fairly comparable. The mean trajectory is higher for cohort 3 than the other two cohorts, indicating that this group had higher weights on average. A reason for this could be that cohort 3 was sampled during the Covid-19 pandemic, while the cohorts 1 and 2 were sampled prior to the pandemic. All three mean trajectory lines follow a similar pattern. As

time increases, the slope of the line tends to flatten. The FPCA method fills in missing data from full data trends. For cohort 3 the subjects do not have data after day 200, but FPCA fills in data for the rest of the year to show what they expect the trajectory to be for these subjects.

Table 2 represents the summary output from the linear mixed model. Based on the output, the variables that have a significant relationship with weight are Sex, Race, Days in Study and Month. Females weigh, on average, 29.7 lbs less than males (p=0.0001). An Asian participant weighs, on average, 35.1 lbs less than a White participant (p=0.0147). For every additional day in the study, a participant is expected to weight 0.06 lbs less, on average (p<0.0001). Compared to the month of December, participants weigh significantly less in August, October and November (maybe December is feeling the effects of Turkey season). In August, participants weigh, on average, 1.1 lb less than in December (p=0.0051). In October, participants weigh, on average, 0.7 lbs less than in December (p=0.0224). In November, participants weigh, on average, 0.6 lbs less than in December (p=0.0032). There is a significant overall effect of month in the model (Num DF = 11, Den DF = 752, F = 2.8, p = 0.0014).

Table 2. Linear Mixed Model Effect Estimates (AIC = 60749.4)

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	243.73	8.0060	83	30.44	<.0001
Cohort					
Cohort 1	-6.6402	7.8774	83	-0.84	0.4017
Cohort 2	-13.5468	7.6993	83	-1.76	0.0822
Sex = Female	-29.7363	7.3467	83	-4.05	0.0001
Race					
Asian	-35.0862	14.0876	83	-2.49	0.0147
Black/African American	22.1371	11.8563	83	1.87	0.0654
Other	-6.2564	17.7543	83	-0.35	0.7254
Days in Study	-0.06345	0.0087	15E3	-7.28	<.0001
Month					
Jan	0.3610	0.2236	752	1.61	0.1068
Feb	-0.1148	0.2905	752	-0.40	0.6927
Mar	-0.0625	0.3355	752	-0.19	0.8524
Apr	-0.2761	0.3692	752	-0.75	0.4548
May	-0.6381	0.4047	752	-1.58	0.1153
Jun	-0.8124	0.4179	752	-1.94	0.0523
Jul	-0.5340	0.4182	752	-1.28	0.2020
Aug	-1.1282	0.4016	752	-2.81	0.0051
Sep	-0.5131	0.3629	752	-1.41	0.1579
Oct	-0.6813	0.2977	752	-2.29	0.0224
Nov	-0.6385	0.2159	752	-2.96	0.0032

Note: Reference levels - Cohort = 3, Sex = Male, Race = White, Month = December

Table 3. Type 3 Tests for Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Cohort	2	83	1.55	0.2183
Sex	1	83	16.38	0.0001
Race	3	83	3.64	0.0160
Days in Study	1	15E3	52.96	<.0001
Month	11	752	2.80	0.0014

Code Appendix

```
** Models **;
/* AIC = 97794.7 */
PROC MIXED DATA = wt ;
      CLASS participant_id cohort sex race(ref = "5") month;
      MODEL wt_lb = cohort sex age race study_days month / solution;
      RANDOM intercept / SUBJECT = participant_id(cohort) type=ar(1);
RUN;

/* AIC = 61608.7 */
PROC MIXED DATA = wt ;
      CLASS participant_id cohort sex race(ref = "5") month;
      MODEL wt_lb = sex age race study_days month / solution ;
      REPEATED / SUBJECT = participant_id(cohort) type=ar(1);
RUN;

/* AIC = 61590.6 */
PROC MIXED DATA = wt ;
      CLASS participant_id cohort sex race(ref = "5") month;
      MODEL wt_lb = cohort sex age race study_days month / solution ;
      REPEATED / SUBJECT = participant_id type=ar(1);
RUN;

/* AIC = 60749.4 */
PROC MIXED DATA = wt ;
      CLASS participant_id cohort sex race(ref = "5") month;
      MODEL wt_lb = cohort sex race study_days month / solution ;
      REPEATED / SUBJECT = participant_id type=ar(1);
RUN;
```