Homework 3 Solutions

1. The dataset MITgrowth.csv on the course website data are from a prospective study on body fat accretion in a cohort of 162 girls from the MIT Growth and Development Study. The study was designed to look at changes in percent body fat in girls before and after menarche. All subjects had to be pre-menarche and non-obese to enter the study. Observations were taken annually until 4 years after menarche. At each observation percent body fat was measured.

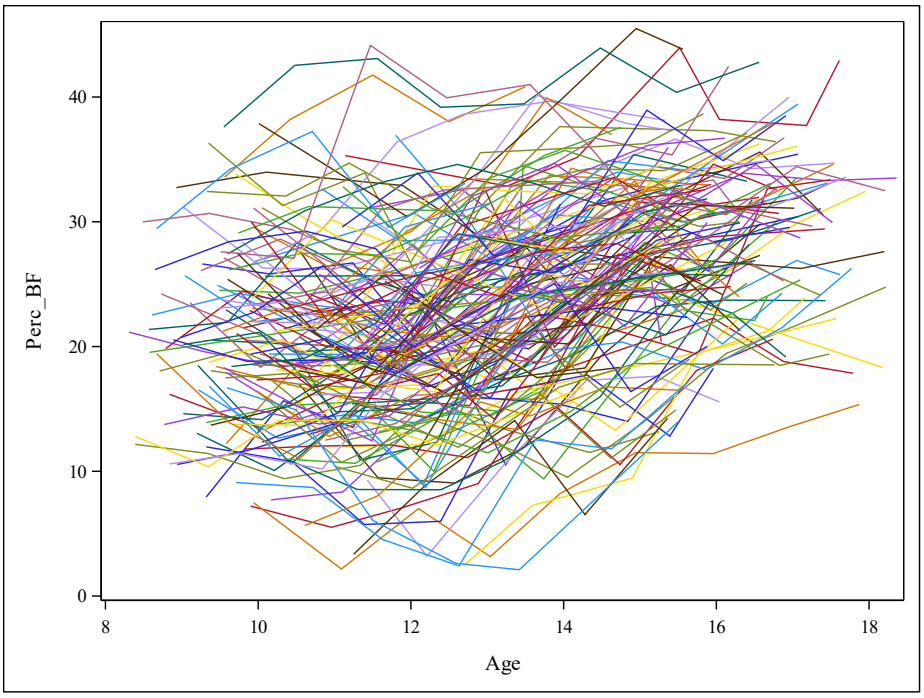
Two time-scales are included: age, and time since menarche (which can be negative). Time since menarche is the more biologically relevant time scale to use. The variables (in order) are: *Subject ID, Current Age (years), Age at Menarche (years), time relative to Menarche (years), Percent Body Fat*.

a. **(15 points) Produce a spaghetti plot of the data using Age and then time relative to menarche. Which time scale appears to have a stronger relationship with percent body fat? Which time scale is best to answer the study question? Comment on the possible parametric methods that could be used to include time in the model.**

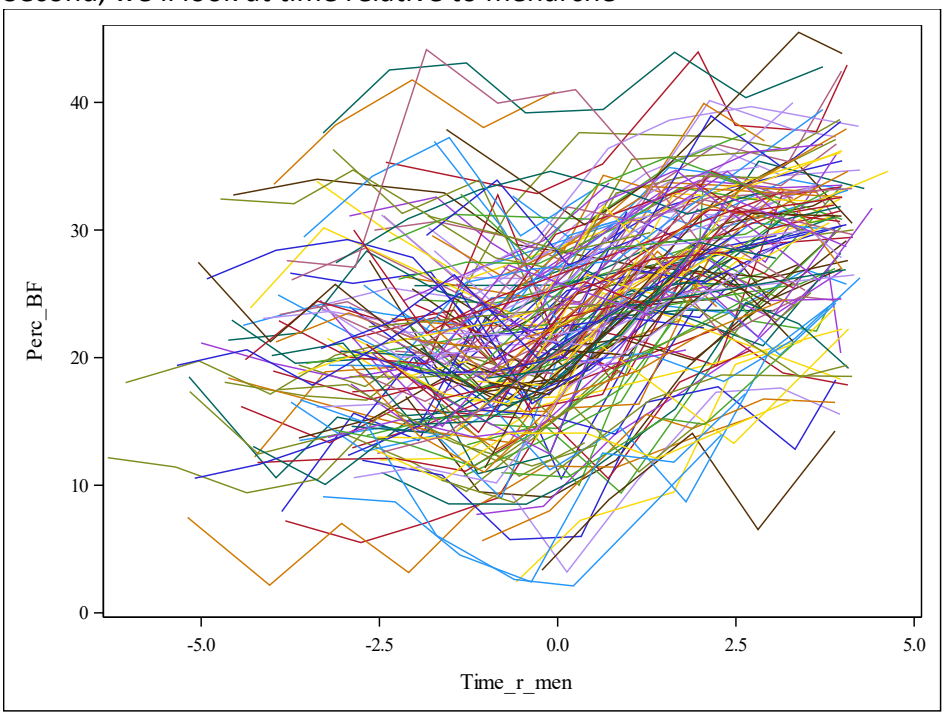
Grading:

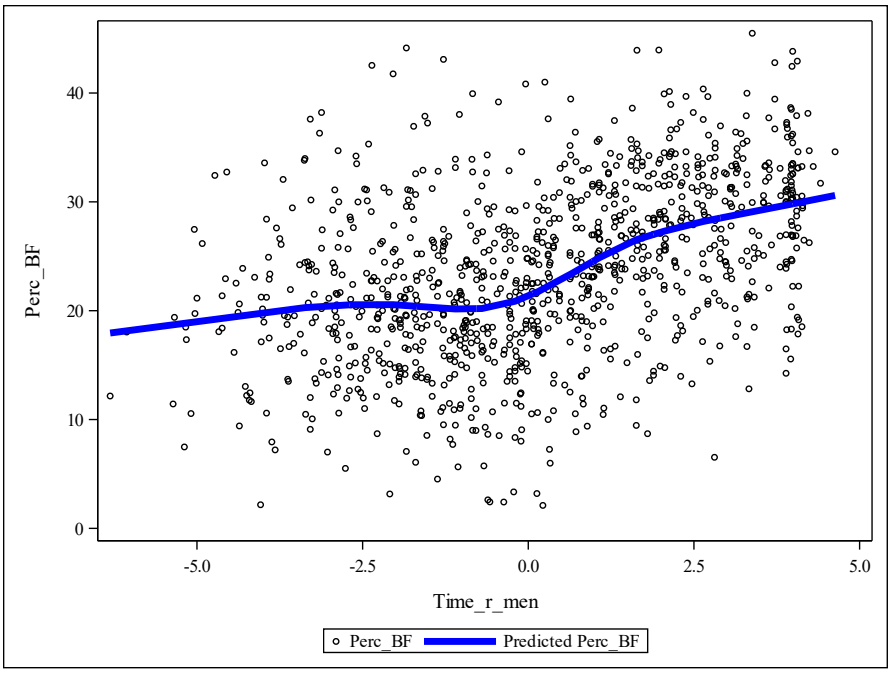
- **4 points for each figure (total of 8 points)**
- **7 points for discussion of how to include time. They don't have to get the linear spline model, but they have to give some justification for what they choose.**

Which time scale is best to answer the study question? Well, there are two "time"" scales in this data: Age and time relative to menarche. Let's look to see which is more closely related to the outcome. First, we'll look at age.



Second, we'll look at time relative to menarche





There's not a big difference in which time variable appears to be more closely related with the outcome, but I would probably say time relative to menarche looks more closely related. Also, since this study is interested in looking at changes in percent body fat in girls before and after menarche, it aligns with the study objectives and is what I'll use going forward.

The possible parametric methods are a linear spline (at time.r.men = 0). This makes a lot of sense since it aligns with the objectives of the study. We could possibly add a log-linear, quadratic, or square-root transformations of time before and after menarche.

b. (15 points) Fit a model with the time effect you found to be most appropriate in (a), with randomly varying intercepts and slopes.

Grading:

- **9 points for correcting fitting the random intercept and slope model. They don't have to fit the model I fit, just the one that corresponds to what they found in a.**
- **2 points each for a, b, and c. Just right or wrong.**

Here I'm going to use the linear spline model. First, I need to add a variable that will allow for a different linear trend before and after menarche.

```
data menarche2;
set menarche;
time2 = max(Time_r_men,0);
```

```
run;

proc mixed data = menarche2;
class ID;
model Perc_BF =  Time_r_men time2/ solution vciry residual outp=BF_pred outpm=BF_pred2
influence(effect=ID est) alpha=0.05;
random intercept Time_r_men/type=UN subject=ID g gcorr v vcorr;
run;
```

i. What is the estimated variance of the random intercepts?

   The variance of the random intercepts is 37.9165

ii. What is the estimated variance of the random slope(s)?

   The variance of the random slope is 0.5330

iii. What is the estimated correlation between the random intercepts and slopes?

   The estimated correlation is -0.2448. The estimated covariance is -1.1006

(c) (10 points) Fit a model with only randomly varying intercepts. What do you think about this model versus the previous model? Should this be used? Why?

Grading:

- **5 points for correcting fitting the random intercept model.**
- **5 points for correcting deciding which model is best.**

Here are the fit statistics from the random intercept and slope model:

Fit Statistics	
-2 Res Log Likelihood	6085.2
AIC (Smaller is Better)	6093.2
AICC (Smaller is Better)	6093.3
BIC (Smaller is Better)	6105.6

Here are the fit statistics from the random intercept model:

```
proc mixed data = menarche2;
class ID;
```

```
model Perc_BF =  Time_r_men time2/ solution vciry residual outp=BF_pred outpm=BF_pred2
influence(effect=ID est) alpha=0.05;
random intercept/type=UN subject=ID g gcorr v vcorr;
run;
```

Fit Statistics

-2 Res Log Likelihood	6155.4
AIC (Smaller is Better)	6159.4
AICC (Smaller is Better)	6159.4
BIC (Smaller is Better)	6165.6

By AIC, BIC or a likelihood ratio test (which has p<0.001), the random intercept and slope model fits better.

(d) (20 points) Give a full description of your findings. Include interpretations of at least two regression coefficients in your description. All descriptions should be in context of the study and the goals of the study.

Grading:

- **Give 5 points for identifying each point given below. There are 5 of them, they only have to say 4 to get full credit and no extra credit.**

Here are the estimates and confidence intervals:

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept	21.3453	0.5206	161	41.01	<.0001	0.05	20.3173	22.3733
Time_r_men	0.4094	0.1291	161	3.17	0.0018	0.05	0.1544	0.6644
time2	2.0545	0.1867	724	11.00	<.0001	0.05	1.6880	2.4211

(1) The study finds that there is a significant increase in body fat percentage before menarche. (2) Specifically, before menarche we expect an increase of 0.409 percentage points for each year they get older. (3) After menarche, there is a significantly higher increase in body fat percentage than before menarche. (4) Specifically, after menarche girls body fact increases 2.055 percentage points more per year than before Menarche. (5) After Menarche girls body fat increases 2.46 (0.409 + 2.055) percentage points per year.

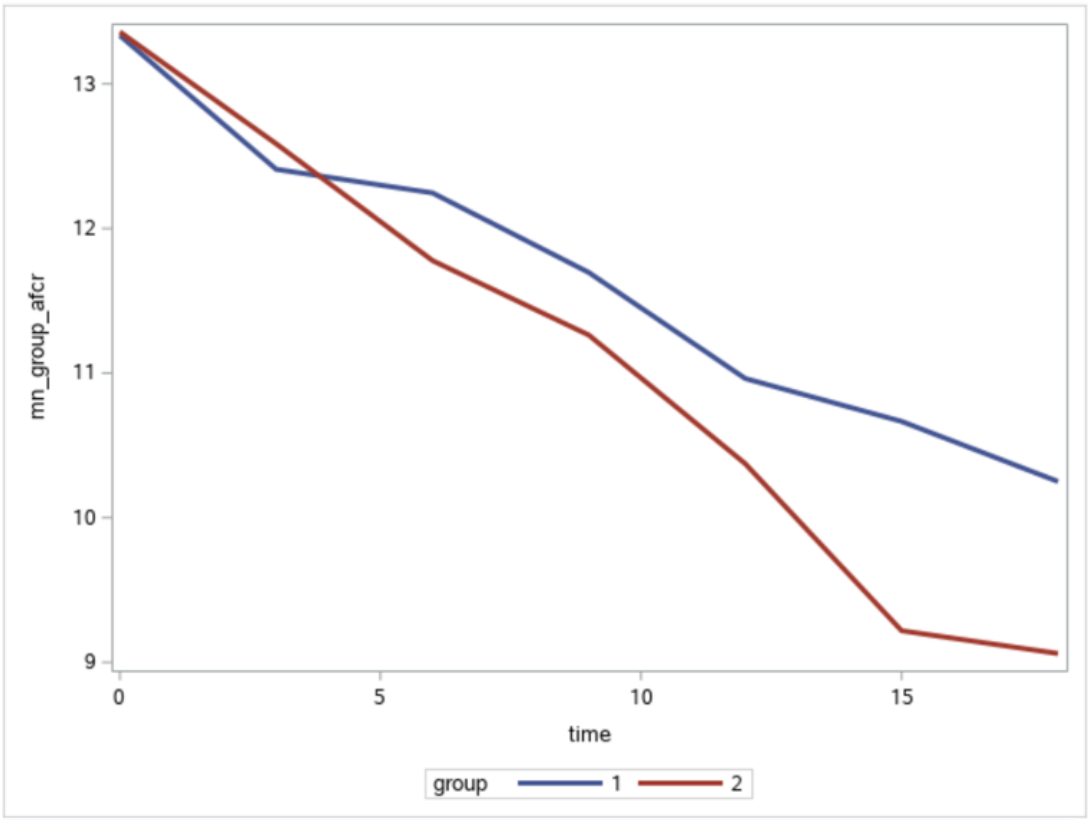
2. **(40 points)** A study was conducted to investigate two treatments for patients suffering from multiple sclerosis. 150 suffers of the disease were recruited into the study, and 75 were randomized to receive azathioprine (AZ) alone (group 1), and 75 were randomized to receive azathioprine plus methylprednisommne (AZ+MP, group 2). For each participant, a measure of auto-immunity, azathioprine AFCR, was planned at clinic visits at baseline (time 0, at initiation of treatment) and at 3, 6, 9, 12, 15, and 18 months thereafter. Multiple sclerosis (MS) affects the immune system. Low values of AFCR (approaching 0) are evidence that immunity is improving, which is hopefully associated with a better prognosis for suffers of MS. Also recorded for each subject was age at entry into the study and an indicator of whether or not the subject had had previous treatment with either of the study agents (0=no, 1=yes). The average age of the men across both treatment groups was 50.45, with SD 6.69. The primary scientific aims of the study are to investigate whether (i) both treatments (AZ or AZ+MP) lower AFCR over the 18 month period and (ii) whether treatment with AZ+MP results in different immune system response than does AZ alone, and, if so, how it is different in terms of response over time. It was also suspected that a subject's age and prior history might be related to their AFCR level at baseline and to the rate at which AFCR changes during the 18 month period. The square root of AFCR is the response variable of interest (square roots were taken so that the AFCR observations better satisfy the assumption of normality).

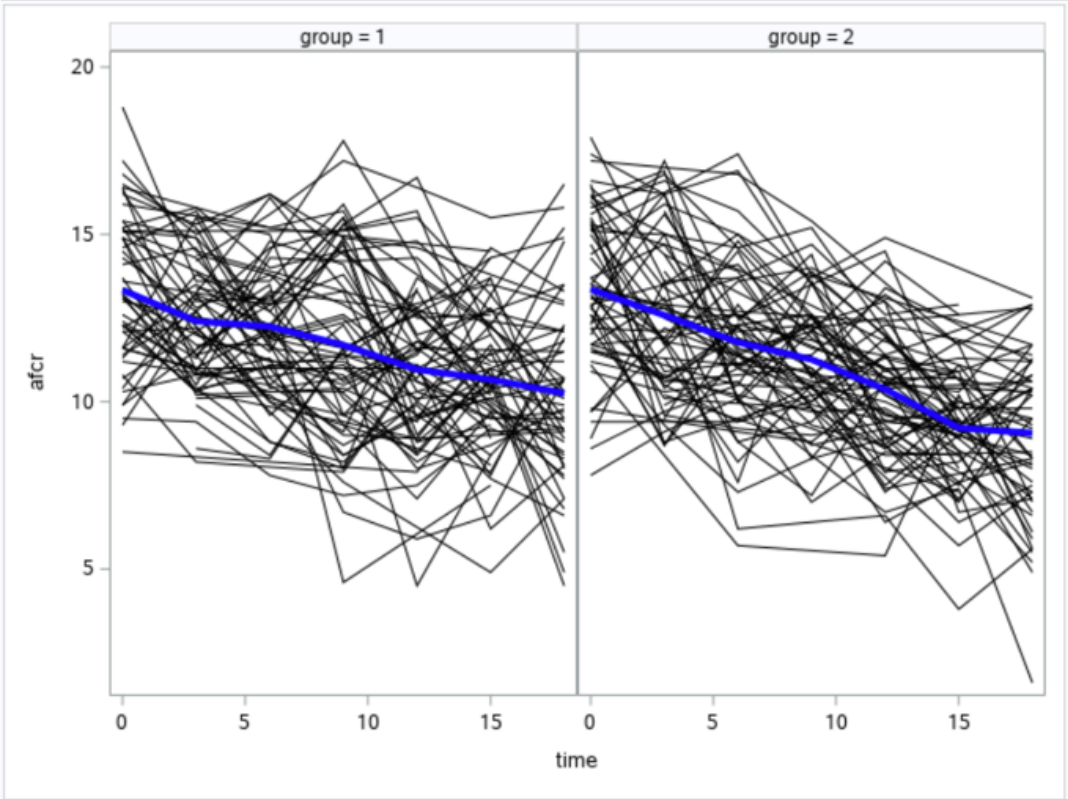
a. Complete an analysis of this data using a linear mixed model that results in the best possible answer to the scientific aims of the study. This should include exploratory data analysis (e.g., figures), model fit, discussion of the random effect(s) variance (or covariance), and a paragraph that accurately summarized your conclusions. It is encouraged to give the code to replicate your findings.

Grading:

- **10 points for exploratory data analysis (figures, etc.)**
- **10 points for model fit. They only need to try different random effect strategies. If they only try different mean models that's OK, but give max 7 points.**
- **5 points for some discussion of the random effect variance.**
- **15 points for the paragraph that accurately summarized the conclusions.**
    - **8 points for finding that there is an interaction,**
    - **2 points for noting the interaction is negative, and**
    - **5 points for interpreting that coefficient correctly.**

The first thing I'd like to do is to look at a plot of the data by group.


mn_group_afcr
13
12
11
10
9
0
5
10
15
time
group
1
2


group = 1
group = 2
afcr
20
15
10
5
0
5
10
15
0
5
10
15
time

There really doesn't appear to be much of an effect of the group variable. It appears that a linear time trend may be sufficient for this data. Also, there may be more heterogeneity in the later months of the study so a random slope may be required.

Random intercept only:

Fit Statistics

-2 Res Log Likelihood	3375.1
AIC (Smaller is Better)	3379.1
AICC (Smaller is Better)	3379.1
BIC (Smaller is Better)	3385.0

Random intercept and slope:

Fit Statistics

-2 Res Log Likelihood	3374.4
AIC (Smaller is Better)	3382.4
AICC (Smaller is Better)	3382.5
BIC (Smaller is Better)	3394.2

Based on this analysis, it appears that the random slope is not required. So, we'll use the model with just the random intercept. Here's a summary of the model.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
UN(1,1)	**id**	2.4026
Residual		3.0725

Fit Statistics

-2 Res Log Likelihood	3375.1
AIC (Smaller is Better)	3379.1
AICC (Smaller is Better)	3379.1
BIC (Smaller is Better)	3385.0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	219.60	<.0001

Solution for Fixed Effects

Effect	group	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper
Intercept		13.0730	0.2464	138	53.05	<.0001	0.05	12.5857	13.5603
time		-0.1578	0.01498	648	-10.53	<.0001	0.05	-0.1872	-0.1283
group	**2**	0.2351	0.3474	648	0.68	0.4989	0.05	-0.4471	0.9173
group	**1**	0	.	.	.	.	.	.	.
time*group	**2**	-0.08893	0.02114	648	-4.21	<.0001	0.05	-0.1304	-0.04743
time*group	**1**	0	.	.	.	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
time	1	648	366.15	<.0001
group	1	648	0.46	0.4989
time*group	1	648	17.70	<.0001

Despite the first figure appearing that there was not an impact of the drug, the final model does show that the treatment was effective. Specifically, the *AZ + MP* group had their azathioprine decrease more rapidly than the group that had AZ alone. In the AZ group (which is group 1, the referent group) we expect that the outcome (sqrt of afcr) will decrease 0.158 units (95% CI -0.187, -0.128) for every month increase. In the *AZ+MP* group, the outcome will decrease 0.089 units (95% CI -0.130, -0.047) more per month than the change in the *AZ* group.

Note that I did not include age or prior into the model. You can adjust for these variables, but it is not necessary because the treatment was randomized. As a result, neither of these variables are related to the treatment group and are thus not confounders.