**BIOST 2050: Longitudinal and Clustered Data Analysis**

**Homework Assignment #3**

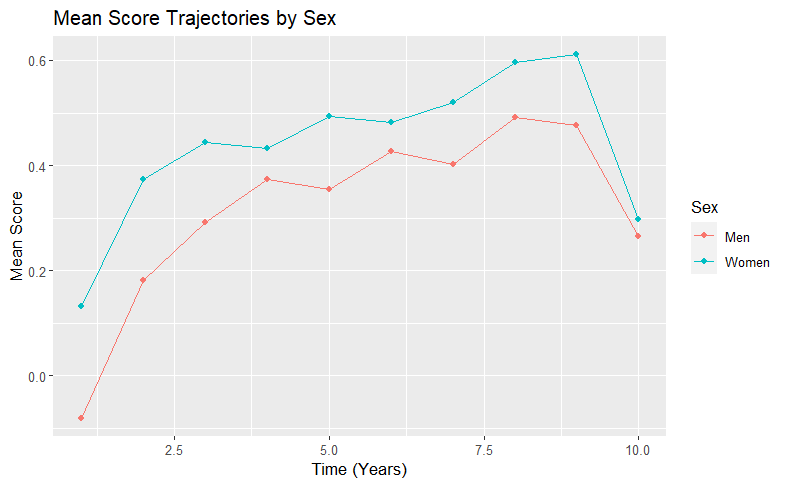
**Due: Wednesday, October 11, 2023**

Answer the questions and justify your answers. Note that you will receive a major deduction if you answer a question by giving only the software output without justifying your answer.

In a study investigating cognitive decline among a cohort of healthy adults approximately aged 60, a total of 1,615 participants were followed for up to 9 years, with a medium follow-up of 4 years. These individuals underwent annual neuropsychological testing. Dataset cognition contains following four variables:

|  |  |
| --- | --- |
| **Variable name** | **Description** |
| studyid | Participant’s unique identification number |
| sexf | Participant’s sex: 0=male, 1=female |
| t | Follow-up year |
| score | Annual neuropsychological test results, where a higher score is associated with better cognition |

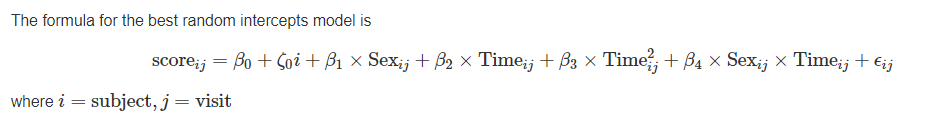
1. Plot the observed mean score trajectories by sex and comment on the shape of the trajectories, highlighting any differences between men and women. You can use the following Stata syntax to label the levels of variable sexf and to generate a variable of quadratic time t2. (6 pts)

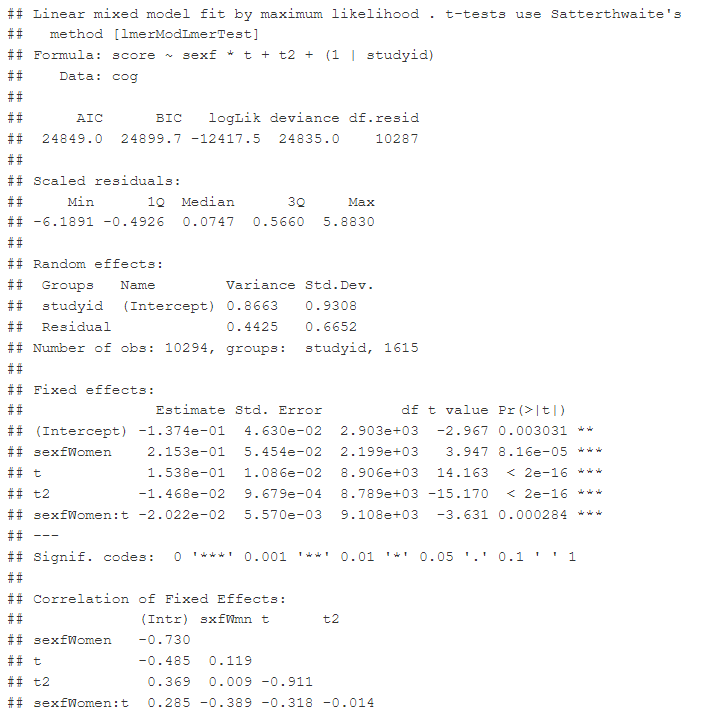


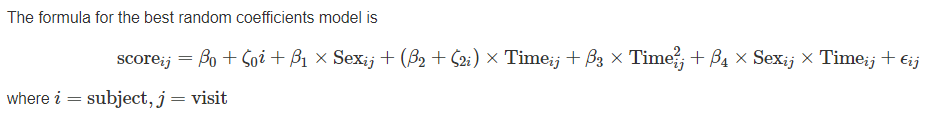
On average, men have an apparently lower score than women at each point in time. Meanwhile, the mean score for each sex follows an inverted-U shape over time. Trajectories for men and women seem mostly aligned.

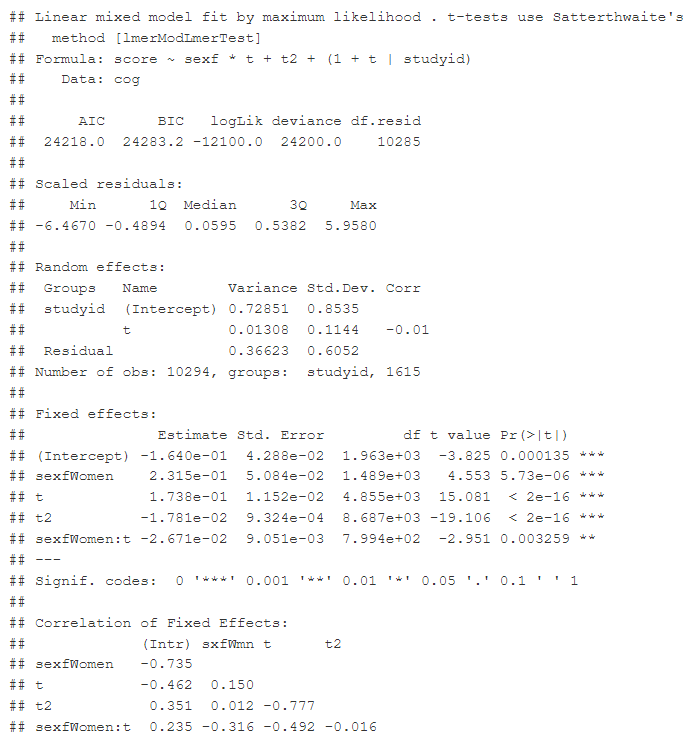
1. Fit both the best random intercepts model and the best random coefficients model, incorporating fixed effects for sex and time, where time is modeled as a quadratic function. In the random coefficients model, random slopes are included solely for the linear time component. Additionally, we account for different rate of change in test scores for men and women. (16 pts)

The following models incorporate these specifications.





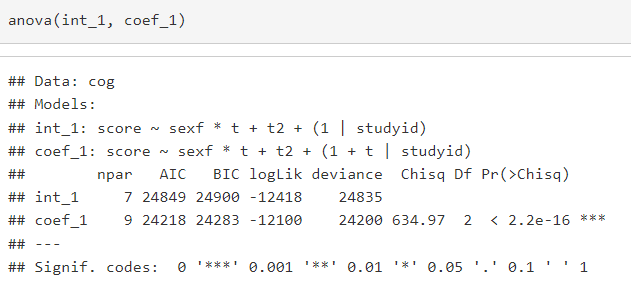




1. From the results of the model fitting in question 2, justify your choice between the best random intercepts and the best random coefficients model for this dataset. (6 pts)

I fit the random intercepts model I fit because this meets the requirements of including fixed sex and time effects (up to quadratic term) and their interaction. Study participants were random effects because their outcomes are correlated across time. Likewise, the random coefficients model I fit met the specification of the fixed effects with random intercept effects and random slopes solely for linear time. For both models, including an interaction between quadratic time and sex was non-significant (did not significantly improve the model).

The sign and magnitude of the fixed effects, whether accounting for time random slopes or not, is largely the same in the two models. However, model comparison reveals that the model performance is better when including random slopes.



1. Based on the model selected in question 2, test whether there are sex differences in score changes over the follow-up period. (8 pts)

According to the linear time main and interaction effects, all individuals have an increased mean score for each unit increase in time. For women, this effect of linear time is slightly less positive in magnitude than it is for men. For each unit increase in quadratic time, the mean score for all individuals decreases slightly.

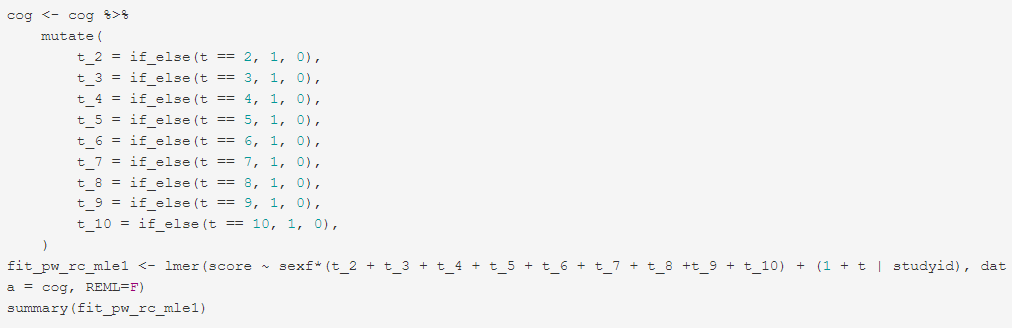
The overall effect of time, at year 5:

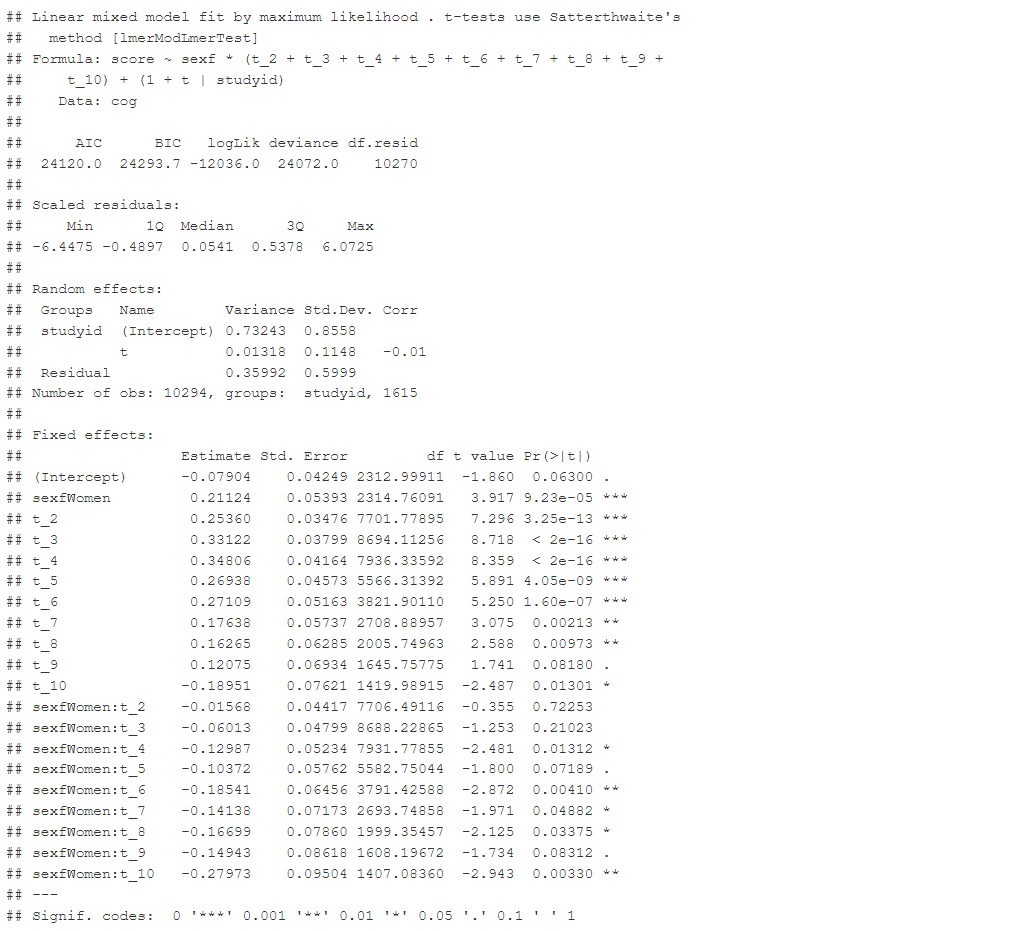
* For women: time contributes to a .1738\*5 - .01781\*25 - .02671\*5 = 0.2902 higher average score
* For men: time contributes to a .1738\*5 - .01781\*25 = 0.42375 higher average score

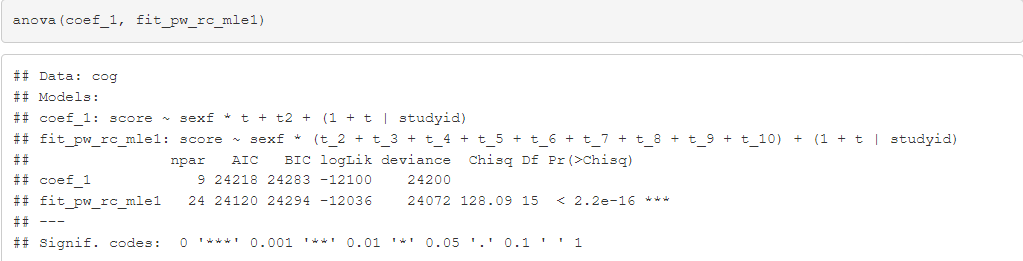
1. Building upon the insights from question 3, identify a more suitable model (called this model hom). Subsequently, with this improved model, fit a piecewise linear time trend with knots at the following eight follow-up times points: 2, 3, 4, 5, 6, 7, 8, and 9 years. Make a choice between the quadratic time trend model (model hom) and the piecewise linear time trend model, providing a justification for your decision. (20 pts)

Set hom = the random coefficients model.

For the piecewise linear model:







The piecewise model summary indicates that the residual variance is about the same as the “hom” model. Similarly, the likelihood ratio test reveals the piecewise model performs incrementally better in AIC, but worse in BIC due to the number of parameters. At the same time, the piecewise model is perhaps a bit more interpretable than the model with quadratic time. That is, each year can be interpreted in its own right without invoking the notion of “quadratic” time. Therefore, I select the piecewise model as my preferred model, but admit that the `hom` model is simpler in terms of the number of parameters and for use in testing level-1 and level-2 heteroskedasticity

1. Make a plot that includes the fitted mean test scores by sex from both the quadratic time trend model and the piecewise linear time trend model (from question 4) on the same figure. (8 pts)

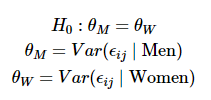


The hom model is smoother than the piecewise model. The linear piecewise model appears more responsive to the less smooth curvature in the average scores observed in the data as it fluctuates around the predicted mean of the hom model. While the predictions mostly overlap between models, it is clear that the average increases more quickly in the piecewise model at the start, then levels out while the hom model continues to rise, then the piecewise model quickly increases again and then quickly decreases as the hom model begins its descent.

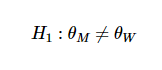
1. Check where there is significant level-1 heteroskedasticity. Explicitly write down the null and alternative hypotheses and perform the test. (15 pts)

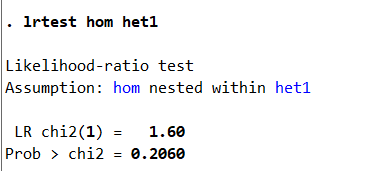
For this question, I will allow level-1 residual variance to vary by sex

H0: The residual variance for men is equal to the residual variance for women.



H1: The residual variance for men is not equal to the residual variance for women





There is no statistical evidence to reject the null hypothesis that the residual variances are the same for men and women.

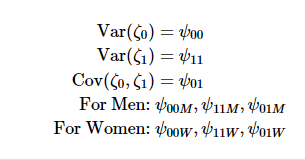
Where het1 = 

And hom = 

1. Check where there is significant level-2 heteroskedasticity. Explicitly write down the null and alternative hypotheses and perform the test. (15 pts)

For this question, I will allow level-2 random effects to vary by sex

H0: The variance of the random intercept for men is equal to the variance of the random intercept for women AND the variance of the random slope for time for men is equal to the variance of the random slope for time for women AND the covariance of the random intercept and random slope for time for men is equal to the covariance of the random intercept and random slope for time for women

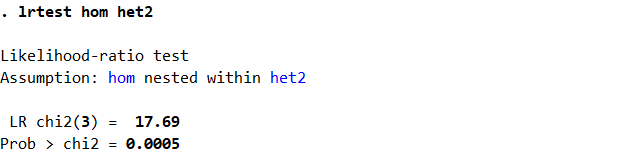




H1: Not (H0) = at least one of the equalities is an inequality.



Let het2 = mixed score sexf t t2 t\_sexf || studyid: t\_sexm sexm, noconstant cov(uns) || studyid: t\_sexf sexf, noconstant cov(uns) reml



There is statistically significant reason to reject the null hypothesis that all the random effects are equal between men and women.

1. Based on the results of questions 1 to 8, summarize your findings. (6 pts)

My exploratory and statistical analysis makes it clear there is a negative, approximately quadratic curvature in mean scores over time. At first they rise, and then towards the end of the study the mean scores start to fall, while never quite returning to the initial averages. This pattern is apparent in both men and women. I also conclude that while women tend to have higher average scores than men, their improvements due to time are slightly less than that of men, a pattern that continues even as average scores start decreasing year over year (i.e., in the last year women have an even greater decrease in average scores than the men do). I also have statistical reason to conclude that residual variances do not differ between men and women in my models, but that random effects of time vary between men and women.