Final Exam

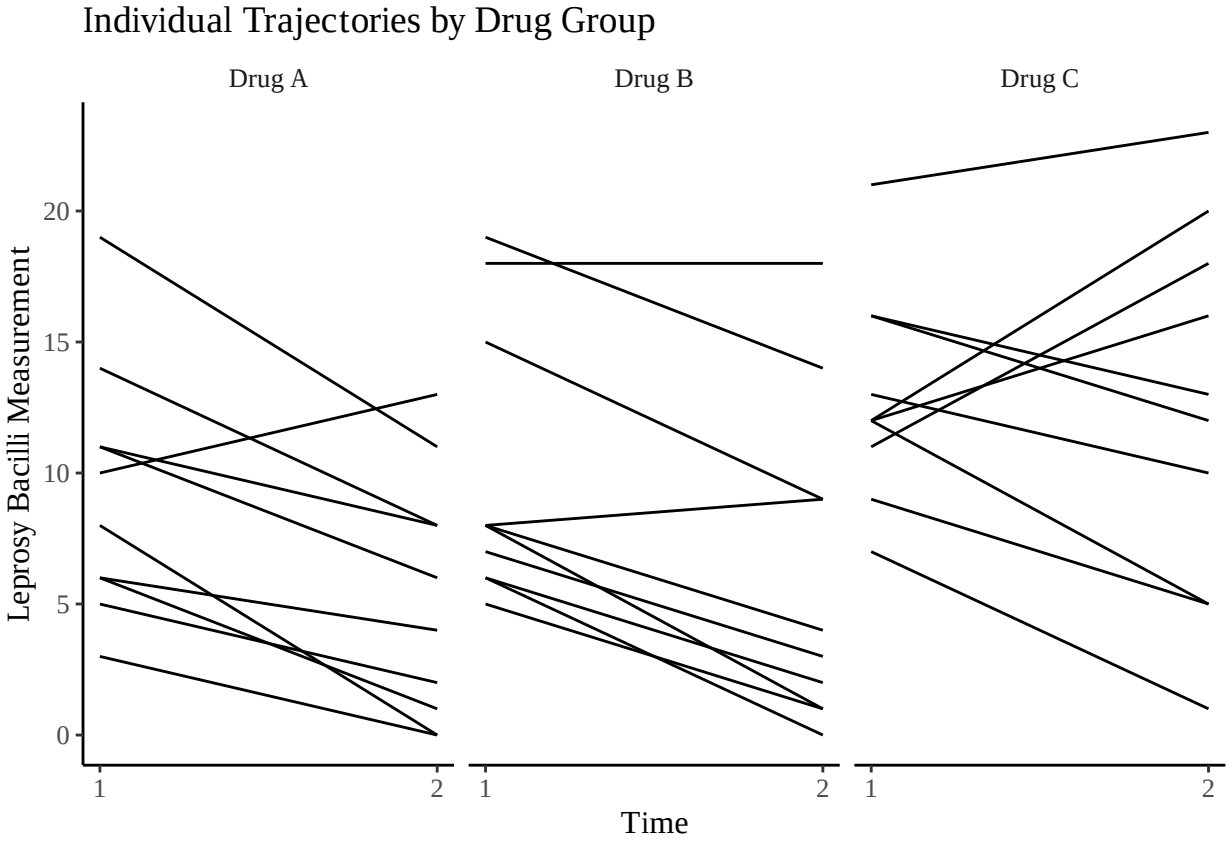
Blain Morin

October 27, 2018

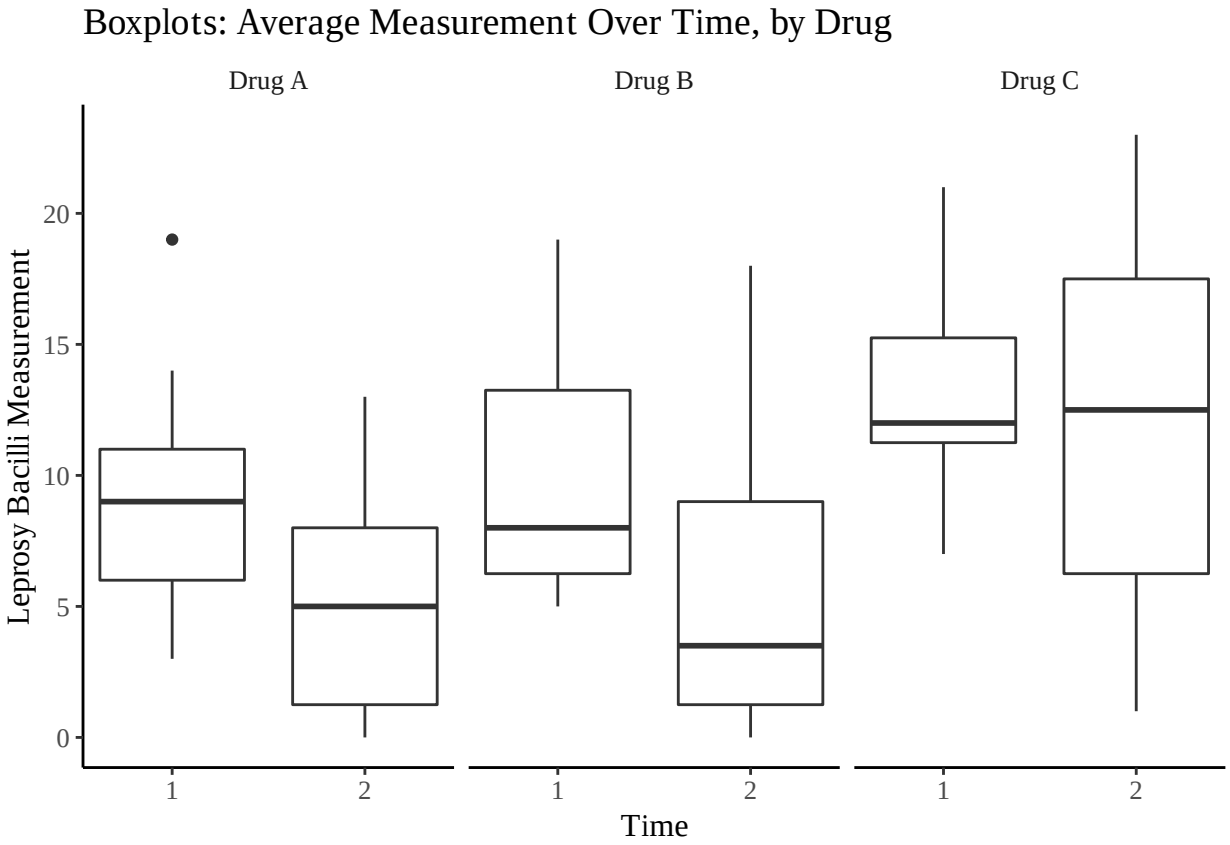
Part 1:

Question 1: Create the following plots:

i. Line ("spaghetti") plots to present the individual trajectories of the primary outcome by treatment group



ii. Plots to present the overall change in the mean primary outcome by treatment group



Question 2: Choose an appropriate regression model to describe the effect of treatment on the change in the primary outcome over time. Fit the model to the data and answer the following questions:

Table 1: Mixed Model Regressions

	Dependent variable:		
	Leprosy Bacilli Measurement		
	Random Intercept Model	Random Slope Model	Fixed Effects Model
	(1)	(2)	(3)
Drug B	0.600	0.600	0.600
	(3.473)	(2.756)	(3.473)
Drug C	0.200	0.200	0.200
	(3.473)	(2.756)	(3.473)
Time (Continuous)	-4.000^{***}	-4.000^{***}	-4.000^{***}
	(1.243)	(1.547)	(1.243)
Drug B : Time	0.100	0.100	0.100
	(1.758)	(2.187)	(1.758)
Drug C : Time	3.400^{*}	3.400	3.400^{*}
	(1.758)	(2.187)	(1.758)
Constant	13.300^{***}	13.300^{***}	13.300^{***}
	(2.456)	(1.949)	(2.456)
Observations	60	60	60
Log Likelihood	-164.239	-164.042	-164.239
Akaike Inf. Crit.	344.478	344.084	344.478
Bayesian Inf. Crit.	361.233	360.839	360.390

Note: *p<0.1; **p<0.05; ***p<0.01
Reference Group is Drug A

i. Explain why you chose the particular model.

ii. Is the change in the primary outcome the same for all three treatments according to this model?

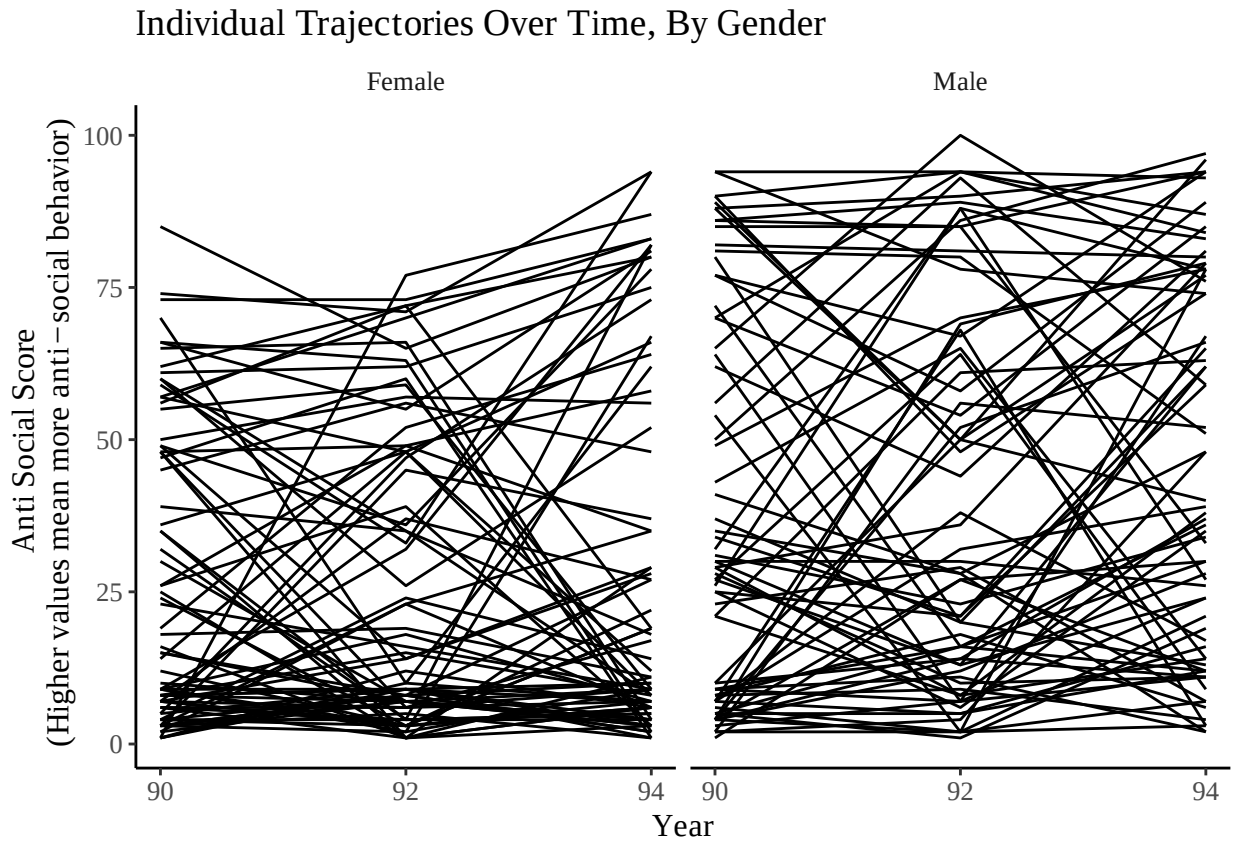
iii. Interpret the regression coefficients.

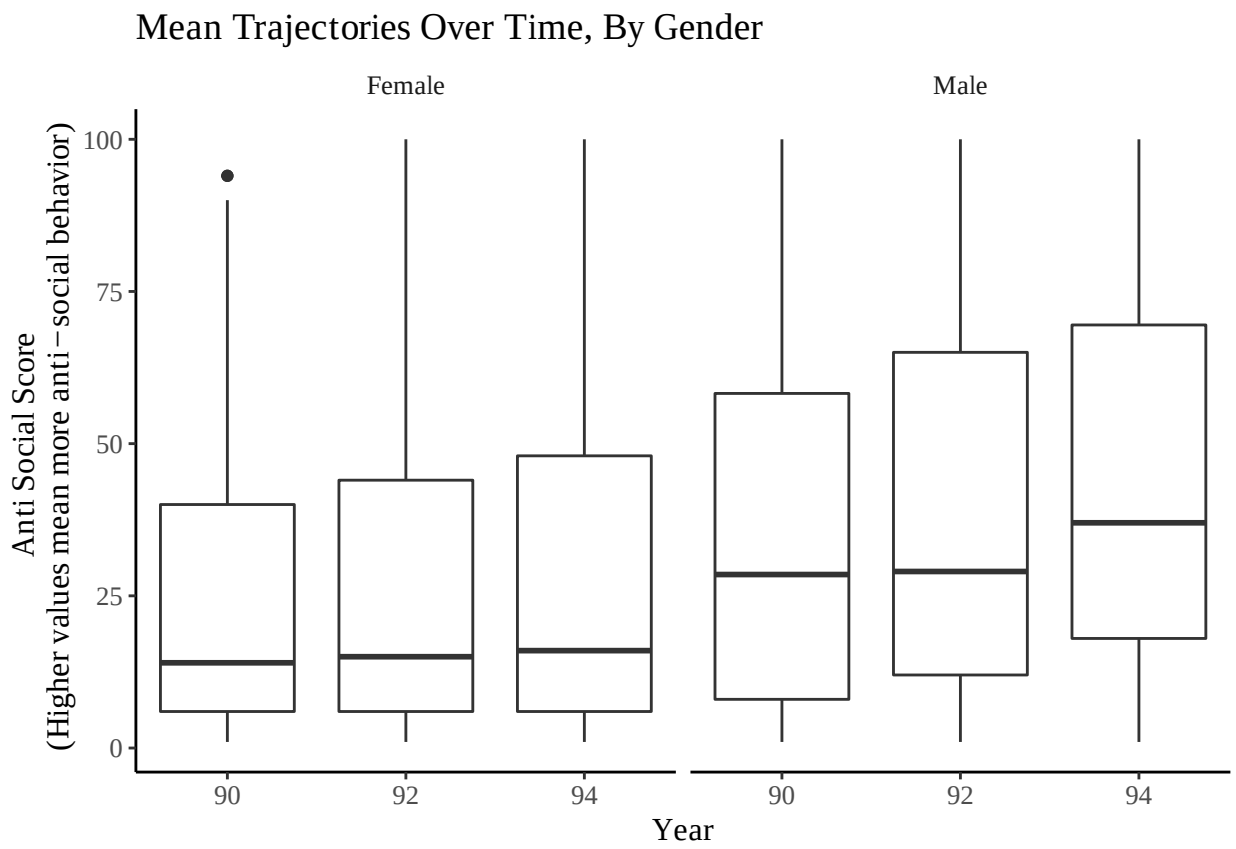
Part 2:

Question 3: Explore the data. Provide descriptive statistics and graphs to present the information included in this dataset in a meaningful and comprehensive way to your collaborators.

Table 2: Summary Statistics

Statistic	Mean	St. Dev.	Min	Max	N
Anti Score	33.271	28.964	1	100	1,743
Female	0.504	0.500	0	1	1,743
Poverty	0.329	0.470	0	1	1,743
Mom's Age	20.656	2.188	16	25	1,743
Child's Age	8.944	0.601	8	10	1,743
Hispanic	0.244	0.430	0	1	1,743
Black	0.363	0.481	0	1	1,743
Mom Works	0.336	0.472	0	1	1,743
Mom Married	0.236	0.425	0	1	1,743


Individual Trajectories Over Time, By Gender
Female
Male
Anti Social Score
(Higher values mean more anti−social behavior)
100
75
50
25
0
90
92
94
Year



Question 4: One of the doctors (Doctor 1) is only interested in overall differences in the antisocial behavior scores between baseline and the end of the follow-up period (1994). He is not interested in describing any particular time trend. He also wants to compare these differences between boys and girls.

i. What methodology would you apply in order to answer this doctor's research question? State the form (write the mathematical formula) of the model (M1) that you will fit to your data.

Mean Response

$$M1 : E[Y_i|X_i] = \beta_0 + \beta_1 Female + \beta_2 Time94 + \beta_3 Female * Time94$$

ii. Based on the model results what is your conclusion about the changes in the mean response over time between boys and girls?

Table 3: Regression Results M1

	Dependent variable:
	Anti Social Score
Female	-10.552^{***}
	(2.336)
Year = 94	8.396^{***}
	(1.680)
Female * Time94	-5.416^{**}
	(2.365)
Constant	35.934^{***}
	(1.659)
Observations	1,162
Log Likelihood	−5,440.820
Akaike Inf. Crit.	10,893.640
Bayesian Inf. Crit.	10,923.970
Note:	*p<0.1; **p<0.05; ***p<0.01
	Year is categorical, reference year is 1990

iii. What is the model estimate of the difference in the average antisocial behavior scores between:

a. boys and girls at baseline?

$$/\beta_1/ = 10.552$$

b. boys and girls in 1994?

$$/\beta_1 + \beta_3/ = 15.569$$

c. 1990 and 1994 for boys?

$$/\beta_2/ = 8.396$$

d. 1990 and 1994 for girls?

$$/\beta_2 + \beta_3/ = 2.98$$

e. boys in 1994 and girls in 1990?

$$/\beta_2 - \beta_1/ = 18.948$$

f. boys in 1990 and girls in 1994?

$$/\beta_1 + \beta_2 + \beta_3/ = 7.572$$

iv. List key weaknesses of this methodology

- Not suitable for unbalanced data
- Ignores time ordering
- Hard to include continuous covariates in the model
- The total number of parameters grows rapidly with

Question 5: Another doctor (Doctor 2), who is also interested in overall changes of antisocial behavior score between boys and girls, believes that there must be some time trend in these changes. He wants to describe the trends during the follow-up period, adjusting for other important covariates.

i. What methodology (model M2) would be more appropriate in this case? Why?

-Marginal model, not interested in individual predictions, can add continuous covariates, treats time as continuous

ii. Using this methodology try to answer the primary research question to the best of your ability, by:

a. Choosing an appropriate covariance pattern model.

Table 4: Choosing Best Correlation Structure for Full Model

	Dependent variable:		
	Anti Social Score		
	Unstructured	AR1	Exchangeable
	(1)	(2)	(3)
Year (Continuous)	−21.628	−21.758	−21.020
	(60.531)	(64.027)	(56.563)
Year2	0.129	0.130	0.126
	(0.329)	(0.348)	(0.307)
Mom Age	−0.481	−0.482	−0.402
	(0.453)	(0.452)	(0.434)
Female = 1	129.614	117.913	184.434
	(3,916.701)	(4,146.603)	(3,663.185)
Age	1.881	1.712	1.782
	(1.633)	(1.629)	(1.565)
Hispanic = 1	−3.216	−3.146	−3.701
	(2.476)	(2.470)	(2.374)
Black = 1	3.581	3.877*	2.845
	(2.258)	(2.252)	(2.167)
Mom Works = 1	4.753**	5.056**	4.006**
	(2.058)	(2.054)	(1.977)
Mom Married = 1	0.069	0.190	−0.209
	(2.266)	(2.261)	(2.176)
Poverty = 1	4.134***	3.865**	5.399***
	(1.522)	(1.520)	(1.543)
Year * Female	−1.768	−1.513	−2.966
	(85.244)	(90.167)	(79.656)
Year2 * Female	0.002	0.001	0.009
	(0.464)	(0.490)	(0.433)
Constant	927.300	934.649	898.798
	(2,781.303)	(2,944.573)	(2,601.251)
Observations	1,743	1,743	1,743
Log Likelihood	−8,022.325	−8,030.515	−8,046.573
Akaike Inf. Crit.	16,078.650	16,091.030	16,123.150
Bayesian Inf. Crit.	16,171.400	16,172.870	16,204.990
Note:			*p<0.1; **p<0.05; ***p<0.01

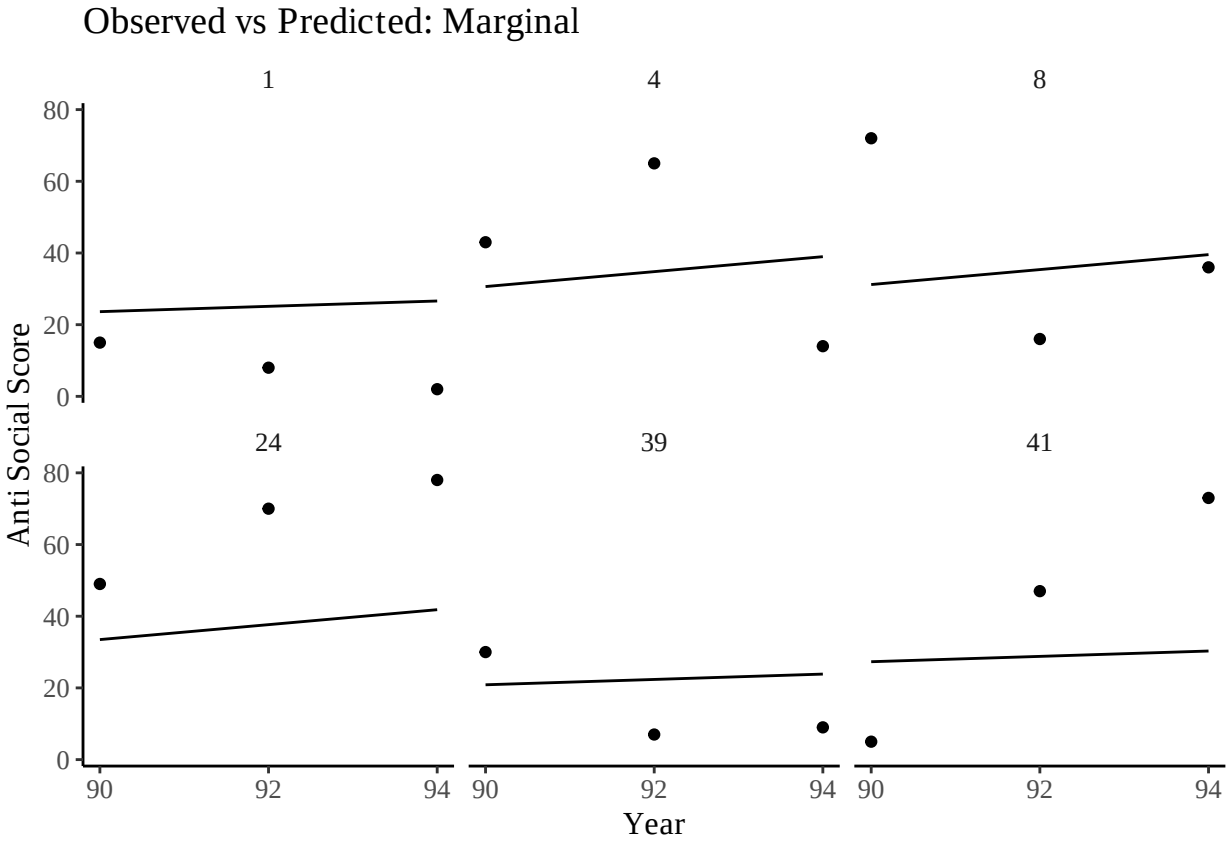
b. Following a model selection procedure to identify important covariates to adjust the results.

Table 5: Backwards Step Regression

	Dependent variable:				
	antiscore				
	(1)	(2)	(3)	(4)	(5)
Year (Continuous)	−21.628	−21.844	−21.844	2.085***	2.085***
	p = 0.721	p = 0.608	p = 0.608	p = 0.00000	p = 0.00000
Year2	0.129	0.130	0.130		
	p = 0.696	p = 0.575	p = 0.575		
Mom Age	−0.481	−0.481	−0.481	−0.481	
	p = 0.289	p = 0.289	p = 0.288	p = 0.288	
Female = 1	129.614	109.965**	109.963**	109.963**	109.904**
	p = 0.974	p = 0.043	p = 0.043	p = 0.043	p = 0.043
Age	1.881	1.881	1.881	1.881	2.284
	p = 0.250	p = 0.250	p = 0.249	p = 0.249	p = 0.150
Hispanic = 1	−3.216	−3.217	−3.215	−3.216	−2.971
	p = 0.194	p = 0.194	p = 0.194	p = 0.194	p = 0.228
Black = 1	3.581	3.580	3.579	3.578	3.842*
	p = 0.113	p = 0.113	p = 0.113	p = 0.113	p = 0.087
Mom Works = 1	4.753**	4.751**	4.745**	4.744**	4.722**
	p = 0.021	p = 0.021	p = 0.021	p = 0.021	p = 0.022
Mom Married = 1	0.069	0.069			
	p = 0.976	p = 0.976			
Poverty = 1	4.134***	4.136***	4.144***	4.145***	4.278***
	p = 0.007	p = 0.007	p = 0.007	p = 0.007	p = 0.005
Year * Female	−1.768	−1.340**	−1.340**	−1.340**	−1.340**
	p = 0.984	p = 0.023	p = 0.023	p = 0.023	p = 0.023
Year2 * Female	0.002				
	p = 0.996				
Constant	927.300	937.208	937.233	−162.133***	−175.888***
	p = 0.739	p = 0.632	p = 0.632	p = 0.0002	p = 0.00002
Observations	1,743	1,743	1,743	1,743	1,743
Log Likelihood	−8,022.325	−8,022.475	−8,024.212	−8,023.826	−8,024.518
Akaike Inf. Crit.	16,078.650	16,076.950	16,078.420	16,075.650	16,075.030
Bayesian Inf. Crit.	16,171.400	16,164.250	16,160.280	16,152.060	16,145.990
Note:					*p<0.1; **p<0.05; ***p<0.01

iii. State the form of the "best" model M2 that you fit to the data. Based on the results from fitting this model, what do you think about the changes in the mean response over time between boys and girls? Interpret the respective model coefficients.

iv. The doctor wishes to know how well the model predicts for individuals with id = 1, 4, 8, 24, 39, and 41. Provide him with the actual values (observed and predicted), and a plot to answer his question. Comment on the results.



v. Adjust your results (M2) for poverty status (if you have not done so already). Interpret the effect of this covariate on the outcome of interest.

Question 6: Another doctor (Doctor 3) is more interested in being able to give advice to parents regarding worrisome social behavior of their child based on the specific characteristics of each case.

i. What methodology would you recommend in this case? Explain.

-mixed model, can make preds for individuals

ii. Perform a detailed model selection procedure to find the model (M3) that best fits the data. Describe the steps you followed and provide results for M3. State the form of model M3 and describe the main model assumptions.

Table 6: Choosing Random Effects

	Dependent variable:		
	antiscore		
	Random Int and Slope	Random Int	Random Slope
	(1)	(2)	(3)
Time	1.591	1.587	1.620
	(1.385)	(1.449)	(1.836)
Time2	0.129	0.130	0.123
	(0.330)	(0.348)	(0.435)
Mom's Age	−0.454	−0.482	−0.342
	(0.451)	(0.452)	(0.329)
Female = 1	−10.724***	−10.688***	−10.865***
	(2.278)	(2.308)	(2.003)
Age	1.509	1.712	0.980
	(1.626)	(1.629)	(1.180)
Hispanic = 1	−3.251	−3.146	−3.694**
	(2.466)	(2.470)	(1.795)
Black = 1	3.883*	3.877*	3.451**
	(2.249)	(2.252)	(1.664)
Mom Works = 1	5.083**	5.056**	4.766***
	(2.051)	(2.054)	(1.515)
Mom Married = 1	0.187	0.190	−0.153
	(2.259)	(2.261)	(1.658)
Poverty = 1	4.205***	3.865**	6.573***
	(1.525)	(1.520)	(1.554)
Time * Female	−1.351	−1.345	−1.392
	(1.950)	(2.040)	(2.585)
Time2 * Female	0.003	0.001	0.015
	(0.464)	(0.490)	(0.612)
Constant	28.115	26.965	30.269**
	(19.306)	(19.339)	(14.047)
Observations	1,743	1,743	1,743
Log Likelihood	−8,026.811	−8,030.515	−8,197.651
Akaike Inf. Crit.	16,087.620	16,091.030	16,425.300
Bayesian Inf. Crit.	16,180.500	16,172.980	16,507.250
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7: Choosing Covariates: Backwards Step Regression

	Dependent variable:				
	antiscore				
	(1)	(2)	(3)	(4)	(5)
Time	1.591	1.586	1.586	2.106***	2.107***
	p = 0.251	p = 0.120	p = 0.120	p = 0.00000	p = 0.00000
Time2	0.129	0.130	0.130		
	p = 0.697	p = 0.575	p = 0.575		
Mom's Age	−0.454	−0.454	−0.455	−0.455	
	p = 0.315	p = 0.316	p = 0.313	p = 0.313	
Female = 1	−10.724***	−10.728***	−10.728***	−10.728***	−10.756***
	p = 0.00001	p = 0.00000	p = 0.00000	p = 0.00000	p = 0.00000
Age	1.509	1.509	1.510	1.510	1.884
	p = 0.354	p = 0.354	p = 0.353	p = 0.353	p = 0.234
Hispanic = 1	−3.251	−3.251	−3.245	−3.245	−3.016
	p = 0.188	p = 0.188	p = 0.188	p = 0.188	p = 0.219
Black = 1	3.883*	3.883*	3.883*	3.883*	4.134*
	p = 0.085	p = 0.085	p = 0.084	p = 0.084	p = 0.065
Mom Works = 1	5.083**	5.082**	5.066**	5.066**	5.048**
	p = 0.014	p = 0.014	p = 0.014	p = 0.014	p = 0.014
Mom Married = 1	0.187	0.187			
	p = 0.934	p = 0.934			
Poverty = 1	4.205***	4.206***	4.225***	4.226***	4.360***
	p = 0.006	p = 0.006	p = 0.006	p = 0.006	p = 0.004
Time * Female	−1.351	−1.340**	−1.340**	−1.340**	−1.339**
	p = 0.489	p = 0.025	p = 0.025	p = 0.025	p = 0.025
Time2 * Female	0.003				
	p = 0.996				
Constant	28.115	28.116	28.188	28.015	15.096
	p = 0.146	p = 0.146	p = 0.144	p = 0.146	p = 0.293
Observations	1,743	1,743	1,743	1,743	1,743
Log Likelihood	−8,026.811	−8,026.963	−8,028.700	−8,028.315	−8,028.946
Akaike Inf. Crit.	16,087.620	16,085.930	16,087.400	16,084.630	16,083.890
Bayesian Inf. Crit.	16,180.500	16,173.340	16,169.350	16,161.120	16,154.920
Note:					*p<0.1; **p<0.05; ***p<0.01

iii. Provide descriptive statistics and plot(s) to compare the estimated (by M3) vs the observed averages by sex at each time point.