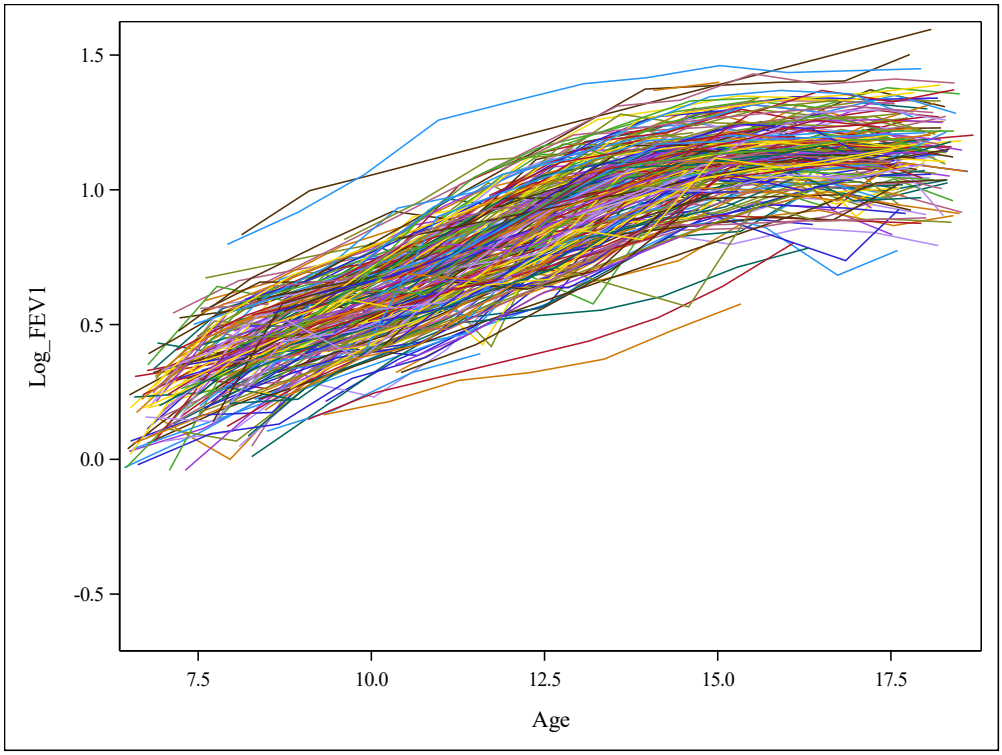
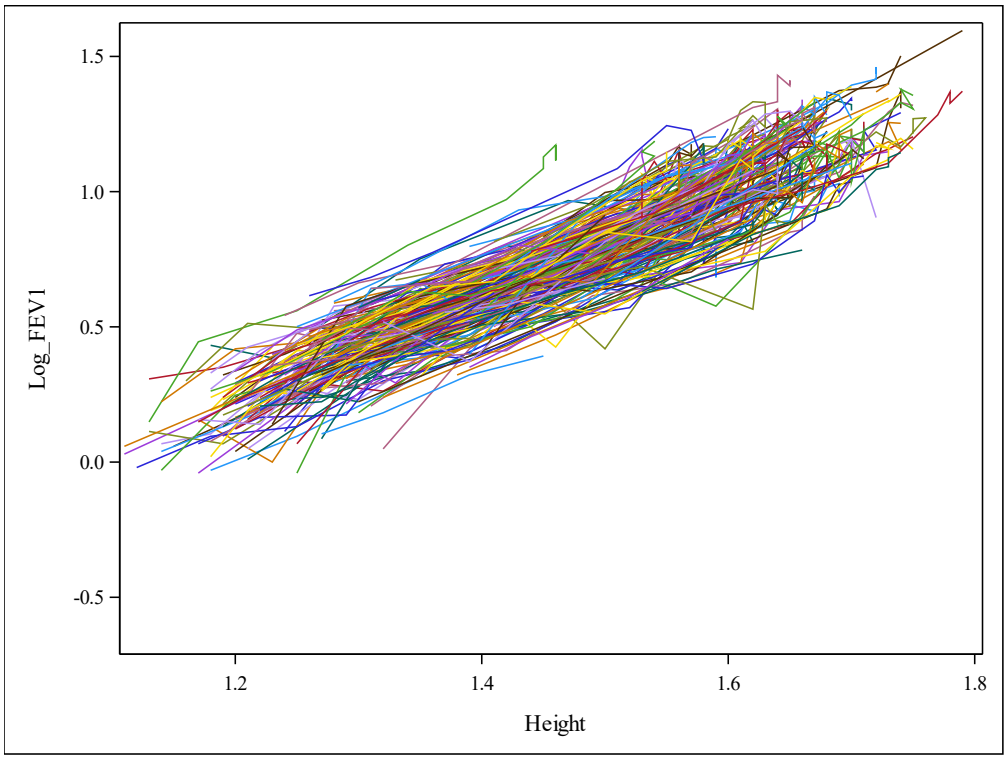
```
*************************************************************
*******       Example Three (Unbalenced data)       *********
*************************************************************;

*The Six Cities Study of Air Pollution and Health was a
longitudinal study designed to characterize lung growth as
measured by changes in pulmonary function in children and
adolescents, and the factors that influence lung function growth.
A cohort of 13,379 children born on or after 1967 was enrolled
in six communities across the U.S.
.;

data air_pol;
input  ID Height Age INI_Height  INI_Age  Log_FEV1;
datalines;
     1     1.20      9.3415       1.20     9.3415     0.21511
     1     1.28     10.3929       1.20     9.3415     0.37156
……..
   300     1.62     17.0075       1.44     11.9617    1.12817
   300     1.63     17.8645       1.44     11.9617    1.16938
;
   run;

Proc SGplot data = air_pol;
series x=Age y=Log_FEV1 / group =ID LineAttrs= (pattern=1 );
run;
```



```
Proc SGplot data = air_pol;
series x=Height y=Log_FEV1 / group =ID LineAttrs= (pattern=1 );
run;
```




```
data air_pol2;
 set air_pol;
lAge = log(Age);
Age_fl = floor(Age);
run;
```

```
proc print data = air_pol2 (obs=10);
run;
```

Obs	ID	Height	Age	INI_Height	INI_Age	Log_FEV1	lAge	Age_fl
1	1	1.20	9.3415	1.20	9.3415	0.21511	2.23447	9
2	1	1.28	10.3929	1.20	9.3415	0.37156	2.34112	10
3	1	1.33	11.4524	1.20	9.3415	0.48858	2.43820	11
4	1	1.42	12.4600	1.20	9.3415	0.75142	2.52252	12
5	1	1.48	13.4182	1.20	9.3415	0.83291	2.59661	13
6	1	1.50	15.4743	1.20	9.3415	0.89200	2.73918	15
7	1	1.52	16.3723	1.20	9.3415	0.87129	2.79559	16
8	2	1.13	6.5873	1.13	6.5873	0.30748	1.88514	6

Obs	ID	Height	Age	INI_Height	INI_Age	Log_FEV1	lAge	Age_fl
9	2	1.19	7.6496	1.13	6.5873	0.35066	2.03465	7
10	2	1.49	12.7392	1.13	6.5873	0.75612	2.54468	12

Our options for covariance matrices for unbalanced data are limited. Here, we'll make the data balanced by using age (floored), but we'll also look at some covariance matrices that use the unbalanced data.

```
proc mixed data=air_pol2;
class ID Age_fl (ref='6');
model Log_FEV1 = Height/solution;
random ID;
repeated Age_fl/ subject=ID type=UN r rcorr;
run;
```

From log:

```
NOTE: An infinite likelihood is assumed in iteration 0 because of a nonpositive definite
      estimated R matrix for ID 6.
NOTE: PROCEDURE MIXED used (Total process time):
      real time          0.09 seconds
      cpu time           0.06 seconds
```

```
proc mixed data=air_pol2;
class ID Age_fl (ref='6');
model Log_FEV1 = Height/solution;
repeated Age_fl / subject=ID type=CS rcorr;
run;
```

Dimensions	
Covariance Parameters	2
Columns in X	2
Columns in Z	0
Subjects	300
Max Obs per Subject	12

Number of Observations	
Number of Observations Read	1994

Number of Observations	
Number of Observations Used	1994
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	-2760.23017893	
1	2	-4278.14652498	0.00030439
2	1	-4279.47432273	0.00000866
3	1	-4279.50940034	0.00000001

Estimated R Correlation Matrix for ID 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	1.0000	0.7136	0.7136	0.7136	0.7136	0.7136	0.7136
2	0.7136	1.0000	0.7136	0.7136	0.7136	0.7136	0.7136
3	0.7136	0.7136	1.0000	0.7136	0.7136	0.7136	0.7136
4	0.7136	0.7136	0.7136	1.0000	0.7136	0.7136	0.7136
5	0.7136	0.7136	0.7136	0.7136	1.0000	0.7136	0.7136
6	0.7136	0.7136	0.7136	0.7136	0.7136	1.0000	0.7136
7	0.7136	0.7136	0.7136	0.7136	0.7136	0.7136	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	ID	0.01140

Covariance Parameter Estimates		
Cov Parm	**Subject**	**Estimate**
Residual		0.004575

Fit Statistics	
-2 Res Log Likelihood	-4279.5
AIC (Smaller is Better)	-4275.5
AICC (Smaller is Better)	-4275.5
BIC (Smaller is Better)	-4268.1

```
proc mixed data=air_pol2;
class ID Age_fl (ref='6');
model Log_FEV1 = Height/solution;
repeated / subject=ID type=sp(exp)(Age) r rcorr;
run;
```

Dimensions	
Covariance Parameters	2
Columns in X	2
Columns in Z	0
Subjects	300
Max Obs per Subject	12

Number of Observations	
Number of Observations Read	1994
Number of Observations Used	1994
Number of Observations Not Used	0

Convergence criteria met.

Estimated R Correlation Matrix for ID 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	1.0000	0.8058	0.6482	0.5271	0.4329	0.2838	0.2360
2	0.8058	1.0000	0.8045	0.6541	0.5372	0.3522	0.2929
3	0.6482	0.8045	1.0000	0.8131	0.6678	0.4378	0.3641
4	0.5271	0.6541	0.8131	1.0000	0.8214	0.5385	0.4478
5	0.4329	0.5372	0.6678	0.8214	1.0000	0.6556	0.5452
6	0.2838	0.3522	0.4378	0.5385	0.6556	1.0000	0.8316
7	0.2360	0.2929	0.3641	0.4478	0.5452	0.8316	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
SP(EXP)	ID	4.8692
Residual		0.01539

Fit Statistics

-2 Res Log Likelihood	-4409.9
AIC (Smaller is Better)	-4405.9
AICC (Smaller is Better)	-4405.9
BIC (Smaller is Better)	-4398.5

```
proc mixed data=air_pol2;
class ID Age_fl (ref='6');
model Log_FEV1 = Height/solution;
repeated / subject=ID type=sp(exp)(lAge) r rcorr;
run;
```

Estimated R Correlation Matrix for ID 1							
Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	1.0000	0.7625	0.5958	0.4809	0.3983	0.2772	0.2402
2	0.7625	1.0000	0.7813	0.6306	0.5224	0.3636	0.3150
3	0.5958	0.7813	1.0000	0.8071	0.6685	0.4653	0.4032
4	0.4809	0.6306	0.8071	1.0000	0.8284	0.5765	0.4995
5	0.3983	0.5224	0.6685	0.8284	1.0000	0.6960	0.6030
6	0.2772	0.3636	0.4653	0.5765	0.6960	1.0000	0.8664
7	0.2402	0.3150	0.4032	0.4995	0.6030	0.8664	1.0000

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
SP(EXP)	ID	0.3934
Residual		0.01516

Fit Statistics	
-2 Res Log Likelihood	-4449.3
AIC (Smaller is Better)	-4445.3
AICC (Smaller is Better)	-4445.3
BIC (Smaller is Better)	-4437.9

Now that we've found a good covariance, we'll move towards looked for the best way to represent the covariates height and age.

```
proc mixed data=air_pol2 method=ml;
class ID Age_fl (ref='6');
model Log_FEV1 = Height/solution;
repeated / subject=ID type=sp(exp)(lAge) rcorr;
run;
```

Fit Statistics

Fit Statistics	
-2 Log Likelihood	-4463.4
AIC (Smaller is Better)	-4455.4
AICC (Smaller is Better)	-4455.4
BIC (Smaller is Better)	-4440.6

```
proc mixed data=air_pol2 method=ml;
class ID Age_fl (ref='6');
model Log_FEV1 = Height lAge/solution;
repeated / subject=ID type=sp(exp)(lAge) rcorr;
run;
```

Fit Statistics	
-2 Log Likelihood	-4527.7
AIC (Smaller is Better)	-4517.7
AICC (Smaller is Better)	-4517.6
BIC (Smaller is Better)	-4499.2

```
proc mixed data=air_pol2 method=ml;
class ID Age_fl (ref='6');
model Log_FEV1 = Height Age/solution;
repeated / subject=ID type=sp(exp)(lAge) rcorr;
run;
```

Fit Statistics	
-2 Log Likelihood	-4522.5
AIC (Smaller is Better)	-4512.5
AICC (Smaller is Better)	-4512.5
BIC (Smaller is Better)	-4494.0

```
proc mixed data=air_pol2 method=ml;
class ID Age_fl (ref='6');
```

```
model Log_FEV1 = Height Age Age*Age/solution;
repeated / subject=ID type=sp(exp)(lAge) rcorr;
run;
```

Fit Statistics	
-2 Log Likelihood	-4533.2
AIC (Smaller is Better)	-4521.2
AICC (Smaller is Better)	-4521.2
BIC (Smaller is Better)	-4499.0

```
proc mixed data=air_pol2 method=ml;
class ID Age_fl (ref='6');
model Log_FEV1 = Height Age Height*Height Age*Age/solution;
repeated / subject=ID type=sp(exp)(lAge) rcorr;
run;
```

Fit Statistics	
-2 Log Likelihood	-4540.5
AIC (Smaller is Better)	-4526.5
AICC (Smaller is Better)	-4526.4
BIC (Smaller is Better)	-4500.5

```
proc mixed data=air_pol2 method=ml;
class ID Age_fl (ref='6');
model Log_FEV1 = Height Age_fl/solution;
repeated / subject=ID type=sp(exp)(lAge) rcorr;
run;
```

Fit Statistics	
-2 Log Likelihood	-4594.6
AIC (Smaller is Better)	-4562.6
AICC (Smaller is Better)	-4562.3

Fit Statistics	
BIC (Smaller is Better)	-4503.3

Now we'll go back to the covariance model fit:

```
proc mixed data=air_pol2;
class ID Age_fl (ref='6');
model Log_FEV1 = Height Age_fl/solution;
repeated / subject=ID type=sp(exp)(lAge) rcorr;
run;
```

Fit Statistics	
-2 Res Log Likelihood	-4482.4
AIC (Smaller is Better)	-4478.4
AICC (Smaller is Better)	-4478.4
BIC (Smaller is Better)	-4471.0

```
proc mixed data=air_pol2;
class ID Age_fl (ref='6');
model Log_FEV1 = Height Age_fl/solution;
repeated / subject=ID type=sp(exp)(Age) rcorr;
run;
```

Fit Statistics	
-2 Res Log Likelihood	-4446.0
AIC (Smaller is Better)	-4442.0
AICC (Smaller is Better)	-4442.0
BIC (Smaller is Better)	-4434.6

```
proc mixed data=air_pol2;
class ID Age_fl (ref='6');
model Log_FEV1 = Height Age_fl/solution;
repeated / subject=ID type=CS rcorr;
run;
```

Fit Statistics	
-2 Res Log Likelihood	-4486.4
AIC (Smaller is Better)	-4482.4
AICC (Smaller is Better)	-4482.4
BIC (Smaller is Better)	-4475.0

Oh my!! Now we need to go back to the mean model fit and check those out again! I'll save you the suspense, and tell you the same model was the best. Let's take a look at the final results.

```
proc mixed data=air_pol2;
class ID Age_fl (ref='6');
model Log_FEV1 = Height Age_fl/solution;
repeated / subject=ID type=CS rcorr;
run;
```

The Mixed Procedure

Model Information	
Data Set	WORK.AIR_POL2
Dependent Variable	Log_FEV1
Covariance Structure	Compound Symmetry
Subject Effect	ID
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Between-Within

Class Level Information		
Class	Levels	Values
ID	300	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170

Class Level Information		
Class	**Levels**	**Values**
		171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
Age_fl	13	7 8 9 10 11 12 13 14 15 16 17 18 6

Dimensions	
Covariance Parameters	2
Columns in X	15
Columns in Z	0
Subjects	300
Max Obs per Subject	12

Number of Observations	
Number of Observations Read	1994
Number of Observations Used	1994
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	-2851.77393747	
1	2	-4484.01011759	0.00050417
2	1	-4486.30601741	0.00002294
3	1	-4486.40220546	0.00000006
4	1	-4486.40243953	0.00000000

Convergence criteria met.

Estimated R Correlation Matrix for ID 1

Row	Col1	Col2	Col3	Col4	Col5	Col6	Col7
1	1.0000	0.7449	0.7449	0.7449	0.7449	0.7449	0.7449
2	0.7449	1.0000	0.7449	0.7449	0.7449	0.7449	0.7449
3	0.7449	0.7449	1.0000	0.7449	0.7449	0.7449	0.7449
4	0.7449	0.7449	0.7449	1.0000	0.7449	0.7449	0.7449
5	0.7449	0.7449	0.7449	0.7449	1.0000	0.7449	0.7449
6	0.7449	0.7449	0.7449	0.7449	0.7449	1.0000	0.7449
7	0.7449	0.7449	0.7449	0.7449	0.7449	0.7449	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	ID	0.01129
Residual		0.003866

Fit Statistics

-2 Res Log Likelihood	-4486.4
AIC (Smaller is Better)	-4482.4
AICC (Smaller is Better)	-4482.4
BIC (Smaller is Better)	-4475.0

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	1634.63	<.0001

Solution for Fixed Effects

Effect	Age_fl	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept		-1.6985	0.06282	299	-27.04	<.0001

Solution for Fixed Effects

Effect	Age_fl	Estimate	Standard Error	DF	t Value	Pr > |t|
Height		1.5699	0.05175	1681	30.34	<.0001
Age_fl	7	0.05777	0.01129	1643	5.12	<.0001
Age_fl	8	0.08168	0.01203	1643	6.79	<.0001
Age_fl	9	0.09790	0.01378	1643	7.10	<.0001
Age_fl	10	0.1119	0.01589	1643	7.04	<.0001
Age_fl	11	0.1203	0.01848	1643	6.51	<.0001
Age_fl	12	0.1490	0.02124	1643	7.01	<.0001
Age_fl	13	0.1791	0.02324	1643	7.70	<.0001
Age_fl	14	0.2205	0.02437	1643	9.05	<.0001
Age_fl	15	0.2463	0.02491	1643	9.88	<.0001
Age_fl	16	0.2535	0.02512	1643	10.09	<.0001
Age_fl	17	0.2581	0.02538	1643	10.17	<.0001
Age_fl	18	0.2527	0.02665	1643	9.48	<.0001
Age_fl	6	0	.	.	.	.

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
Height	1	1681	920.28	<.0001
Age_fl	12	1643	27.48	<.0001