```
*The 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 examined changes in
percent body fat before and after menarche. The data represent a subset of the study
materials and should not be used to draw substantive conclusions.

data menarche;
input  ID Age  Age_at_Men Time_r_men Perc_BF;
datalines;
               1      9.32    13.19    -3.87     7.94
               1     10.33    13.19    -2.86    15.65
               1     11.24    13.19    -1.95    13.51
……..
             162     16.76    12.63     4.13    18.53
;


ods rtf file="C:\Users\mclaina\Dropbox\Classes\755_Spring_2018\Examples\14 - Model
Diagnostics.rtf";


ods html;
ods graphics on;

proc mixed data = menarche2;
class ID;
model Perc_BF =  some model/ solution vciry residual outp=BF_pred outpm=BF_pred2
influence(effect=ID est);
random some random effects;
run;
```

Model Information	
Data Set	WORK.MENARCHE2
Dependent Variable	Perc_BF
Covariance Structure	Unstructured
Subject Effect	ID
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information		
Class	**Levels**	**Values**
ID	162	1 2 … 162

Dimensions	
Covariance Parameters	7
Columns in X	3
Columns in Z per Subject	3
Subjects	162
Max Obs per Subject	10

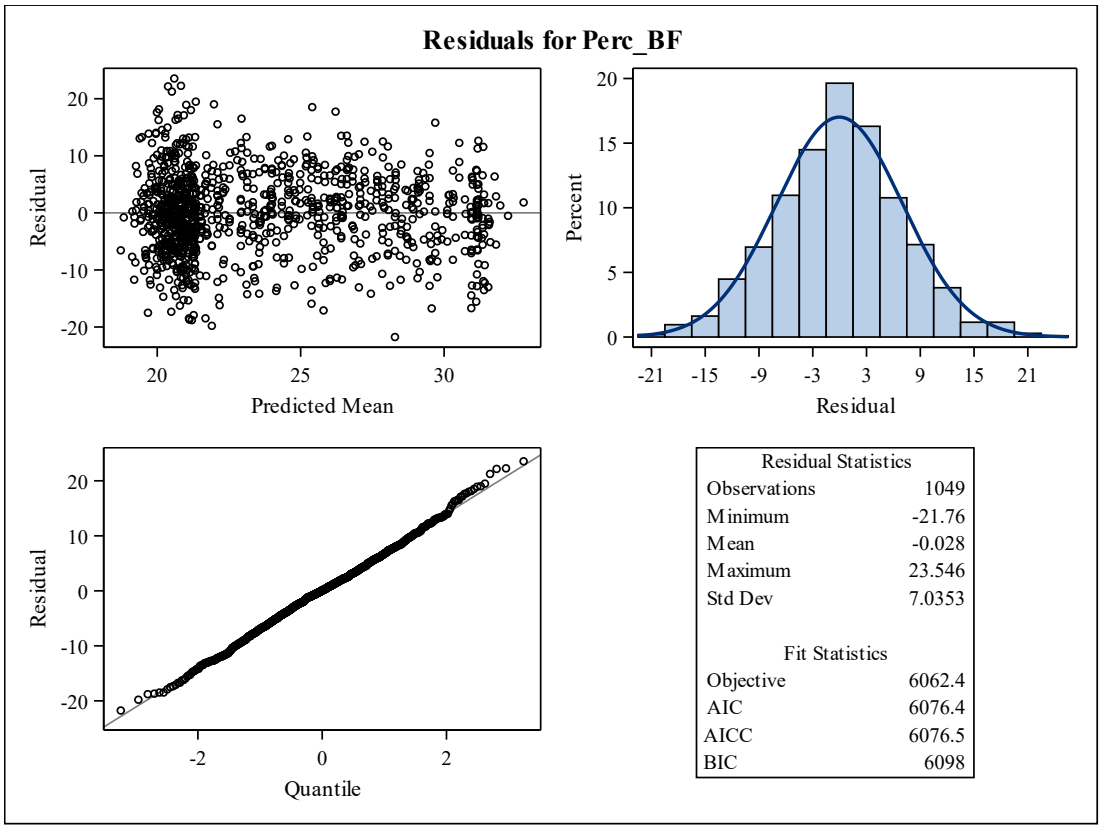
Number of Observations	
Number of Observations Read	1049
Number of Observations Used	1049
Number of Observations Not Used	0

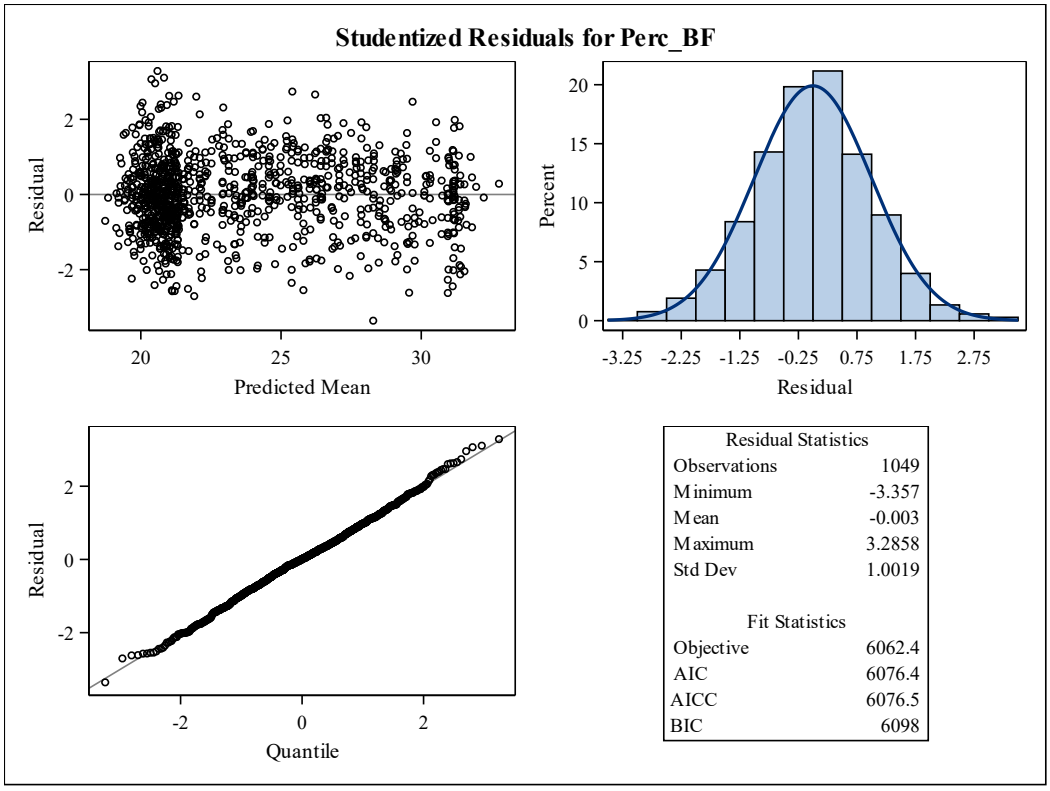
Iteration History			
Iteration	**Evaluations**	**-2 Res Log Like**	**Criterion**
0	1	7073.41750695	
1	2	6063.49976979	0.00045902
2	1	6062.44105637	0.00001857
3	1	6062.40145176	0.00000004
4	1	6062.40136458	0.00000000

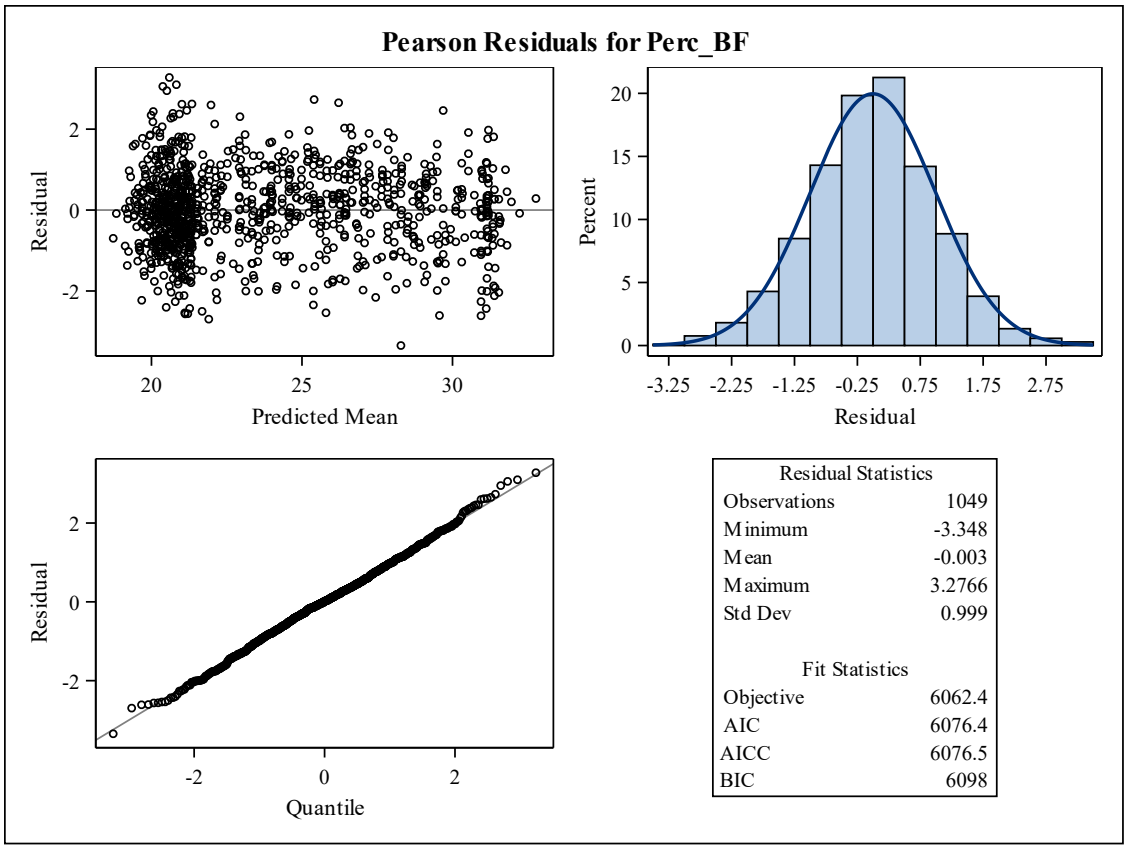
Convergence criteria met.

Estimated V Matrix for ID 1						
Row	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**	**Col6**
1	60.2897	46.9926	43.5473	39.9506	36.0086	32.8875
2	46.9926	54.3058	42.8865	40.8548	38.5541	35.3123
3	43.5473	42.8865	51.7642	41.6694	40.8475	37.4971
4	39.9506	40.8548	41.6694	51.9930	43.2418	39.7779
5	36.0086	38.5541	40.8475	43.2418	55.0584	42.0457
6	32.8875	35.3123	37.4971	39.7779	42.0457	48.8593

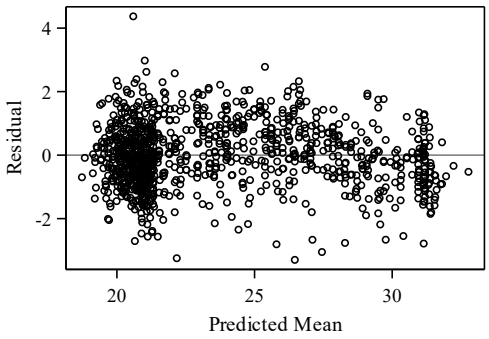
Estimated V Correlation Matrix for ID 1						
Row	**Col1**	**Col2**	**Col3**	**Col4**	**Col5**	**Col6**
1	1.0000	0.8213	0.7795	0.7136	0.6250	0.6059
2	0.8213	1.0000	0.8089	0.7689	0.7051	0.6855
3	0.7795	0.8089	1.0000	0.8032	0.7651	0.7456
4	0.7136	0.7689	0.8032	1.0000	0.8082	0.7892
5	0.6250	0.7051	0.7651	0.8082	1.0000	0.8107
6	0.6059	0.6855	0.7456	0.7892	0.8107	1.0000

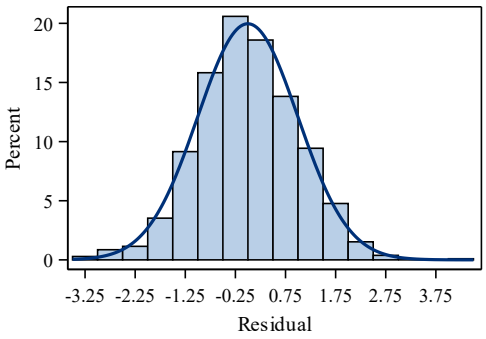


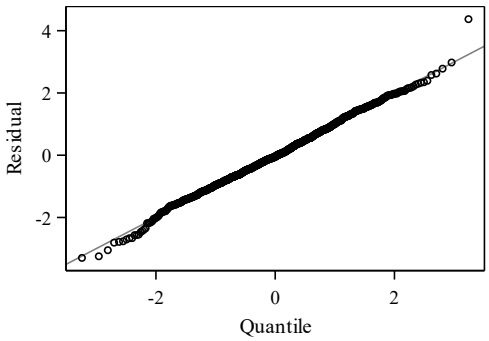

Studentized Residuals for Perc_BF
Residual
Predicted Mean
Percent
Residual
Residual
Quantile
Residual Statistics
Observations 1049
Minimum -3.357
Mean -0.003
Maximum 3.2858
Std Dev 1.0019
Fit Statistics
Objective 6062.4
AIC 6076.4
AICC 6076.5
BIC 6098


Pearson Residuals for Perc_BF
Residual
Predicted Mean
Percent
Residual
Residual
Quantile
Residual Statistics
Observations 1049
Minimum -3.348
Mean -0.003
Maximum 3.2766
Std Dev 0.999
Fit Statistics
Objective 6062.4
AIC 6076.4
AICC 6076.5
BIC 6098

Scaled Residuals for Perc_BF



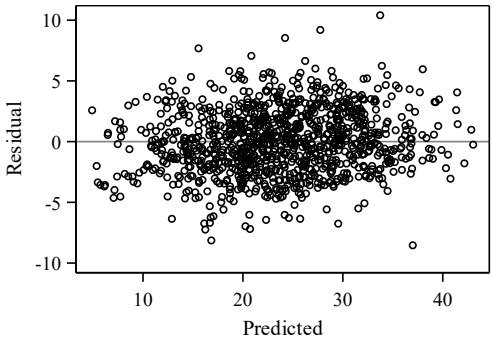


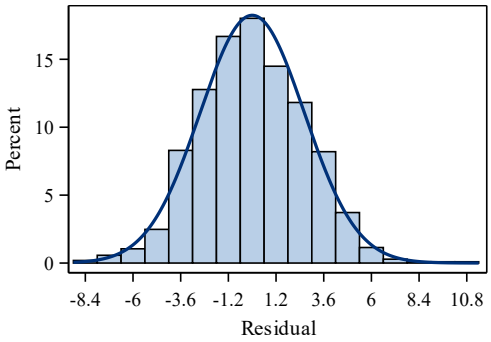


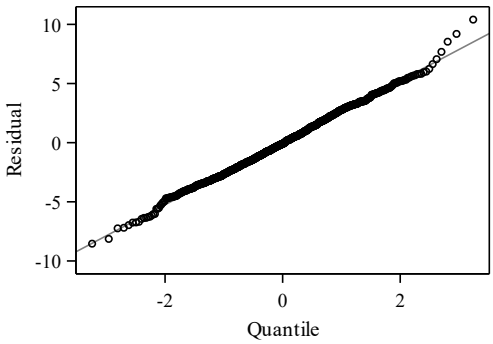
Residual Statistics	
Observations	1049
Minimum	-3.299
Mean	-0.002
Maximum	4.3704
Std Dev	0.999

Fit Statistics	
Objective	6062.4
AIC	6076.4
AICC	6076.5
BIC	6098

Conditional Residuals for Perc_BF

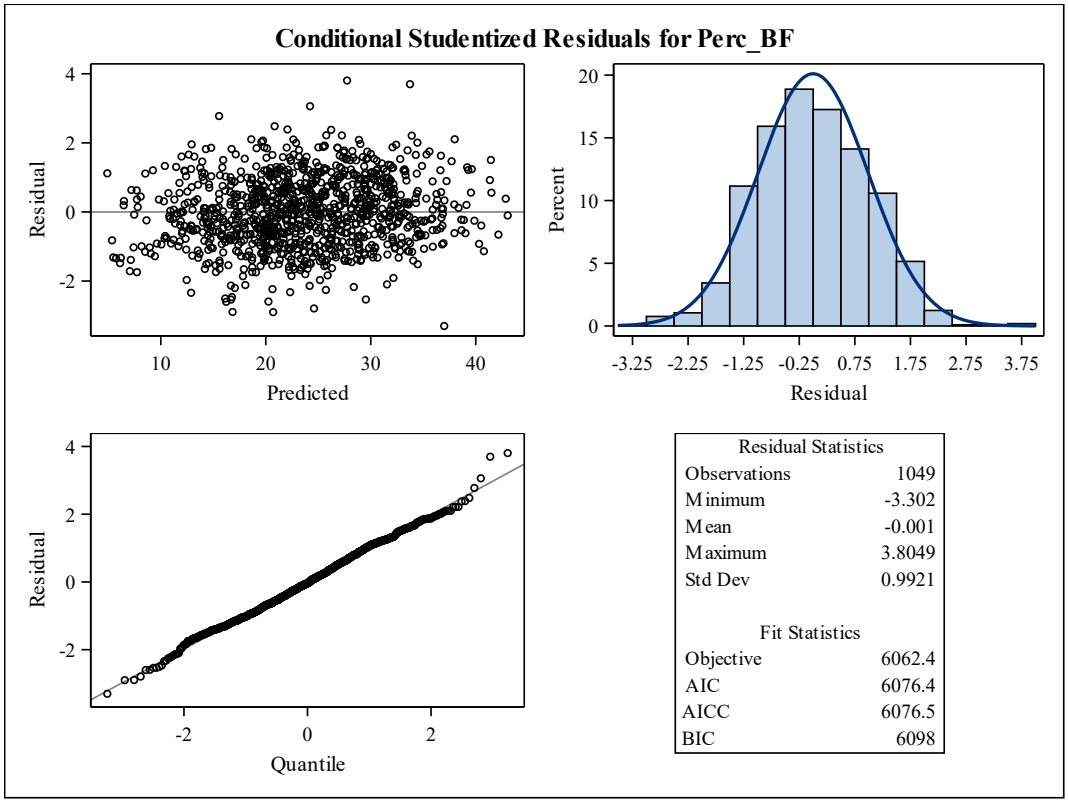


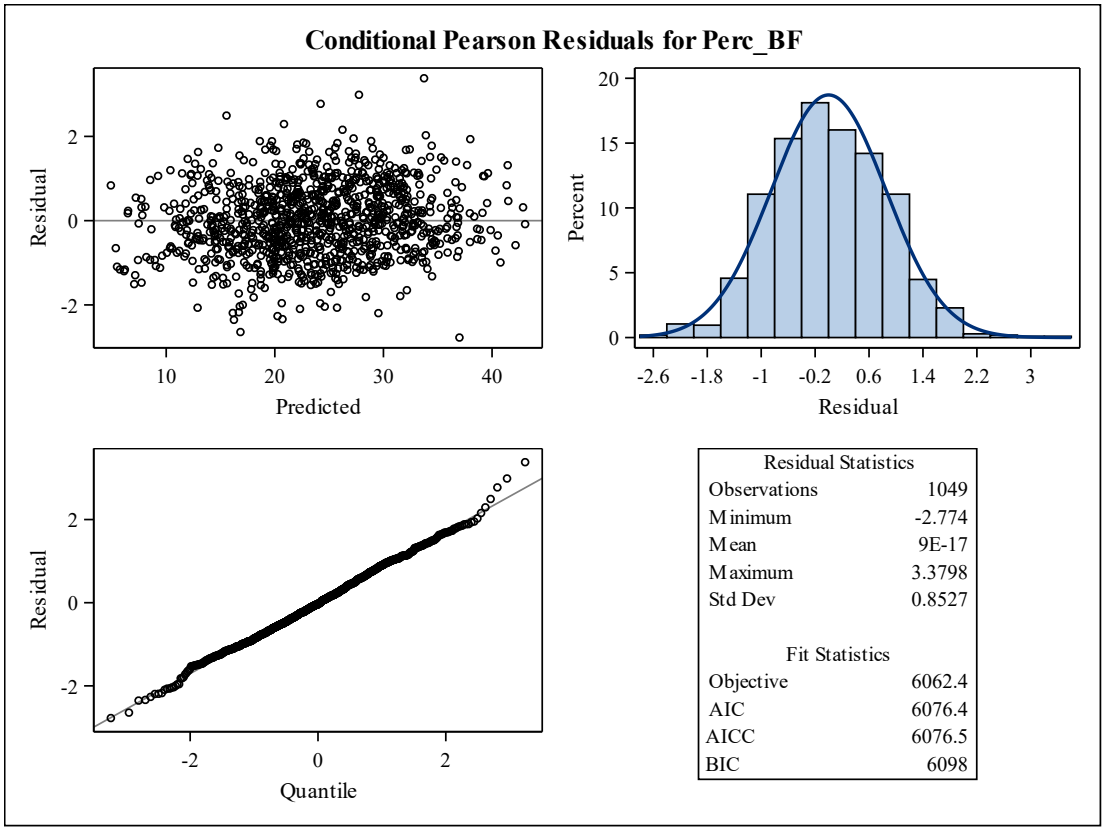




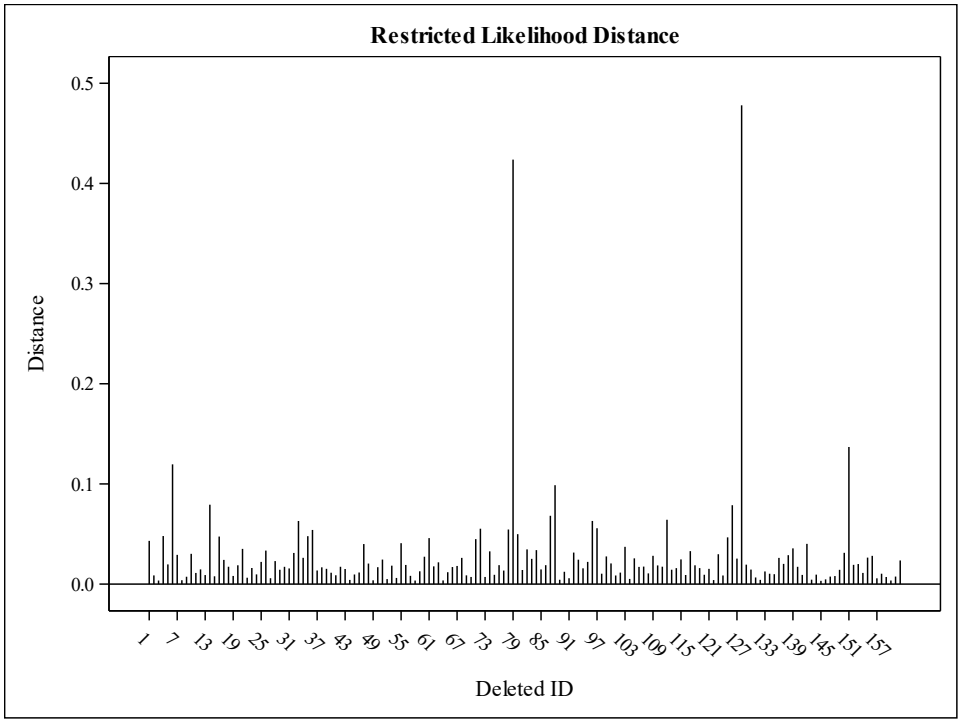
Residual Statistics	
Observations	1049
Minimum	-8.539
Mean	28E-17
Maximum	10.402
Std Dev	2.6245

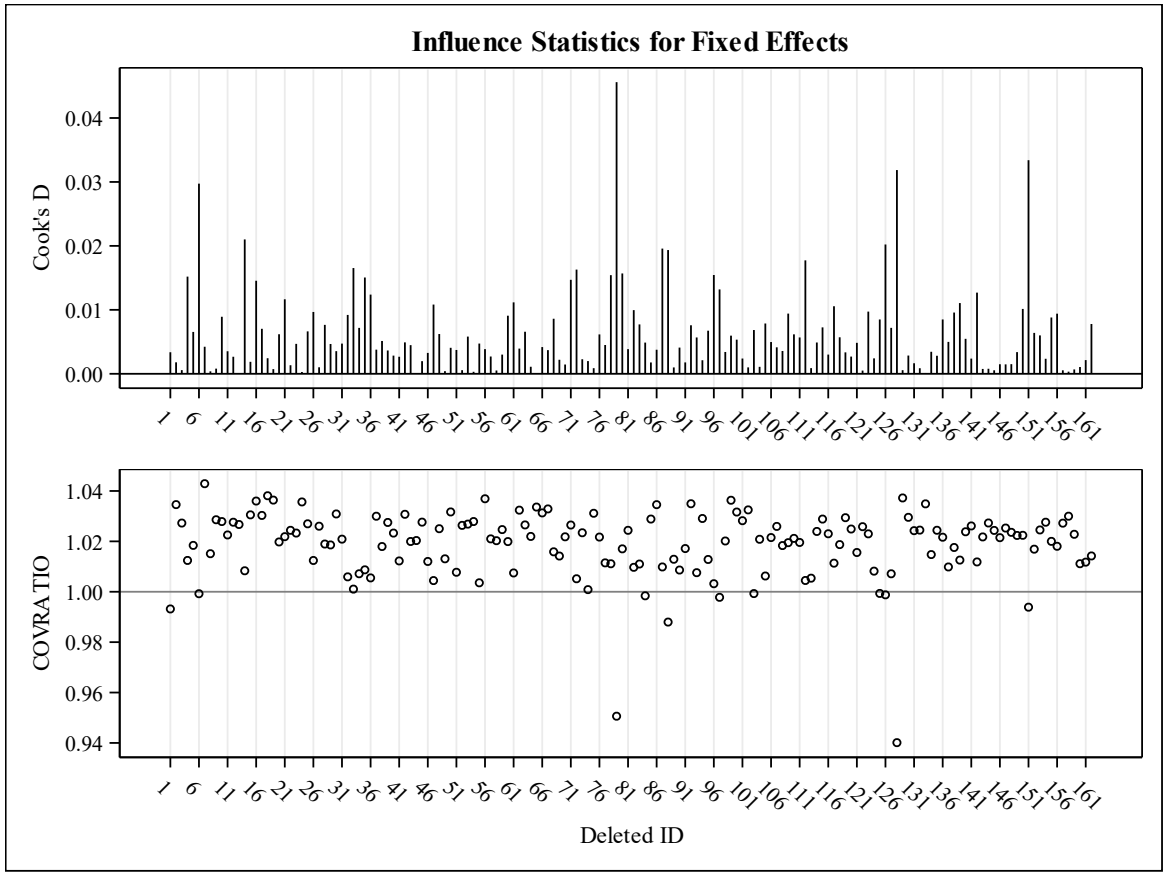
Fit Statistics	
Objective	6062.4
AIC	6076.4
AICC	6076.5
BIC	6098

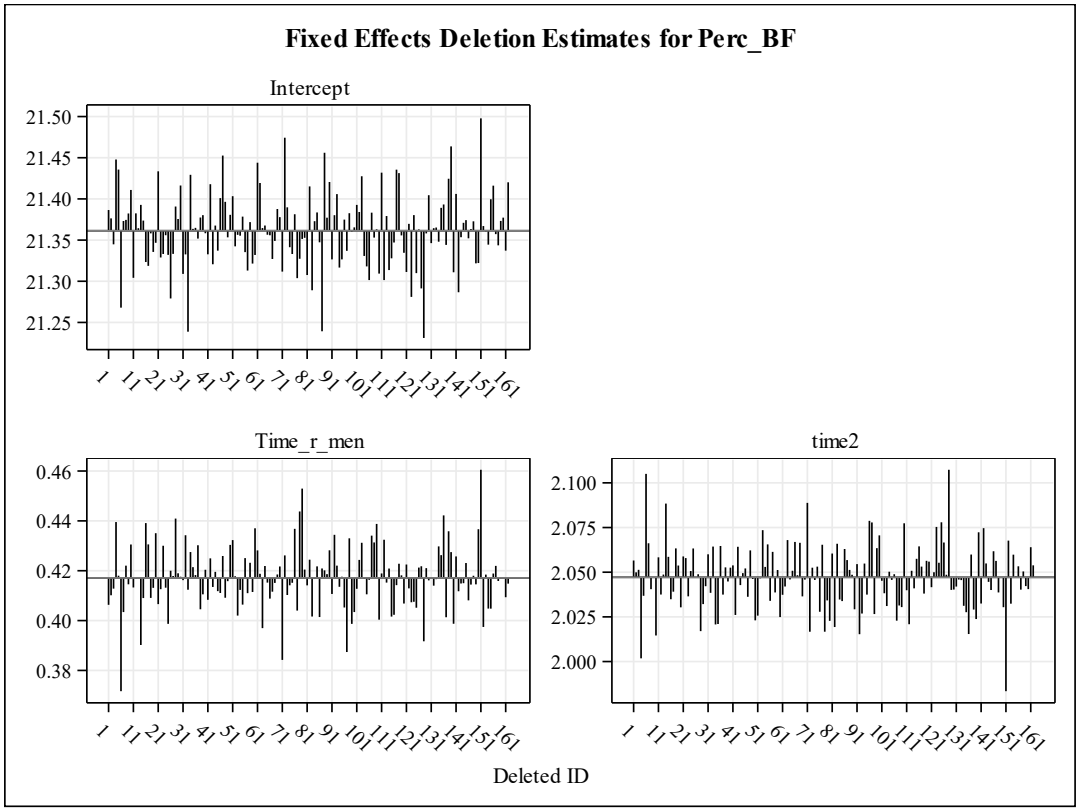

Conditional Studentized Residuals for Perc_BF
Residual
Predicted
Percent
Residual
Residual
Quantile
Residual Statistics
Observations 1049
Minimum -3.302
Mean -0.001
Maximum 3.8049
Std Dev 0.9921
Fit Statistics
Objective 6062.4
AIC 6076.4
AICC 6076.5
BIC 6098

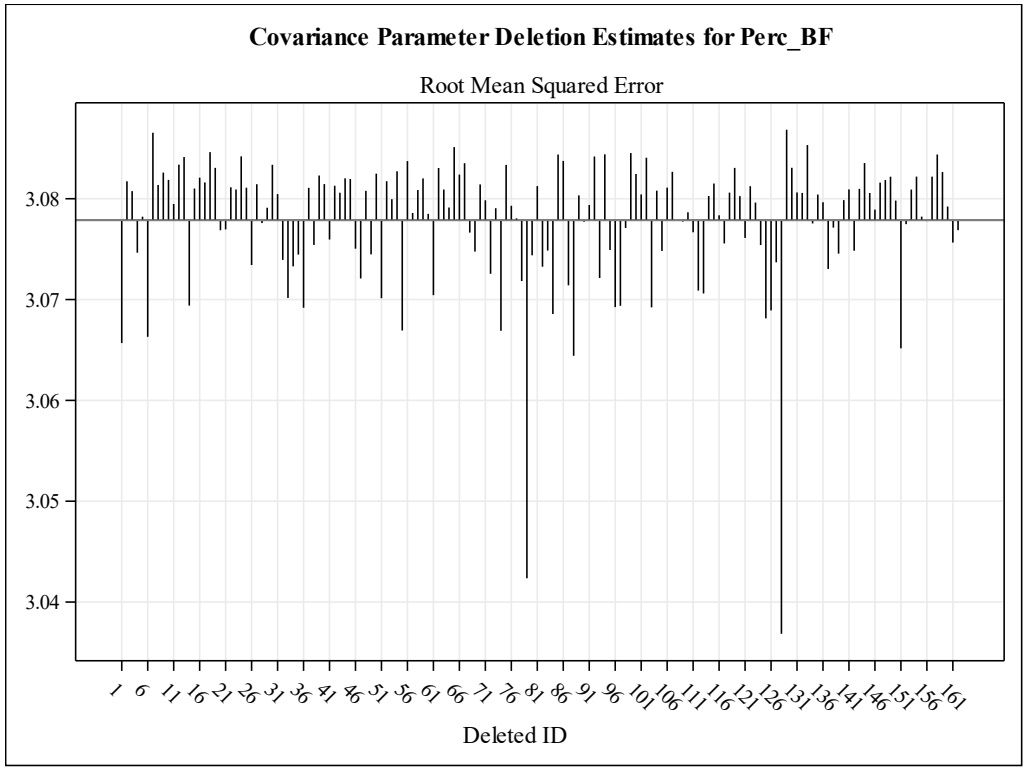

Conditional Pearson Residuals for Perc_BF
Residual
Predicted
Percent
Residual
Residual
Quantile
Residual Statistics
Observations 1049
Minimum -2.774
Mean 9E-17
Maximum 3.3798
Std Dev 0.8527
Fit Statistics
Objective 6062.4
AIC 6076.4
AICC 6076.5
BIC 6098

Influence Diagnostics for Levels of ID								
ID	Number of Observations in Level	PRESS Statistic	Cook's D	MDFFITS	COVRATIO	COVTRACE	RMSE without deleted level	Restricted Likelihood Distance
1	6	350.35	0.00337	0.00337	0.9932	0.0068	3.06568	0.0432
2	9	144.73	0.00179	0.00177	1.0346	0.0342	3.08173	0.0087
3	8	87.53	0.00057	0.00057	1.0273	0.0270	3.08075	0.0036
4	6	474.26	0.01520	0.01513	1.0124	0.0124	3.07465	0.0480
5	7	687.09	0.00655	0.00650	1.0184	0.0183	3.07822	0.0197
161	5	76.23	0.00214	0.00213	1.0118	0.0117	3.07565	0.0075
162	5	569.33	0.00779	0.00774	1.0142	0.0142	3.07688	0.0236


Restricted Likelihood Distance
Distance
0.0
0.1
0.2
0.3
0.4
0.5
1 7 13 19 25 31 37 43 49 55 61 67 73 79 85 91 97 103 109 115 121 127 133 139 145 151 157
Deleted ID


Influence Statistics for Fixed Effects
Cook's D
0.00
0.01
0.02
0.03
0.04
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111 116 121 126 131 136 141 146 151 156 161
COVRATIO
0.94
0.96
0.98
1.00
1.02
1.04
1 6 11 16 21 26 31 36 41 46 51 56 61 66 71 76 81 86 91 96 101 106 111 116 121 126 131 136 141 146 151 156 161
Deleted ID


Fixed Effects Deletion Estimates for Perc_BF
Intercept
21.50
21.45
21.40
21.35
21.30
21.25
Time_r_men
0.46
0.44
0.42
0.40
0.38
time2
2.100
2.075
2.050
2.025
2.000
Deleted ID


Covariance Parameter Deletion Estimates for Perc_BF
Root Mean Squared Error
3.08
3.07
3.06
3.05
3.04
Deleted ID

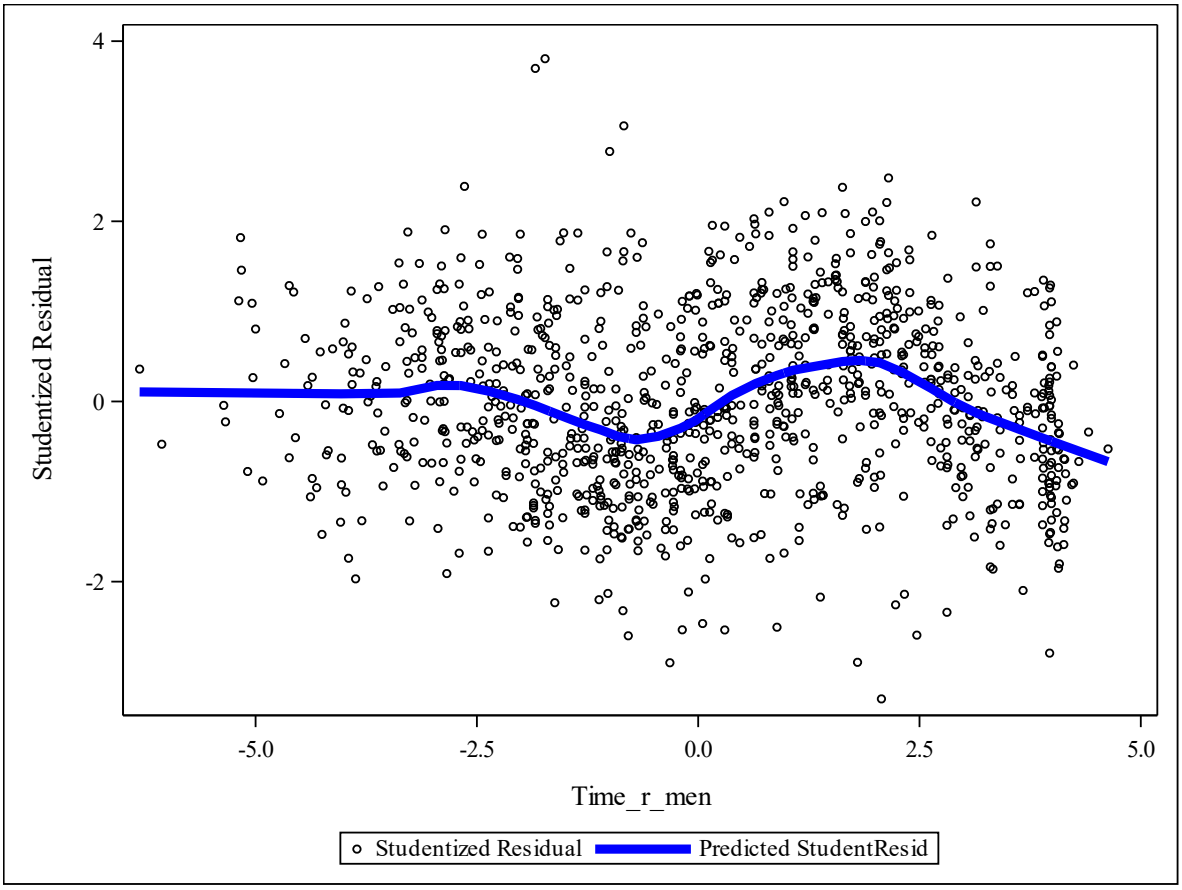
```
*Let's take a look at some on the influential diagnostic datasets;
proc print data = BF_pred (obs=5);
run;
proc print data = BF_pred2 (obs=5);
run;
```

Obs	ID	Pred	StdErrPred	DF	Lower	Upper	Resid	StudentResid	PearsonResid	Alpha
1	1	12.4633	2.04879	565	8.4392	16.4875	-4.52333	-1.96933	-1.46964	0.05
2	1	13.5505	1.53271	565	10.5400	16.5610	2.09953	0.78661	0.68214	0.05
3	1	14.5300	1.27459	565	12.0265	17.0335	-1.01998	-0.36408	-0.33139	0.05
4	1	15.5525	1.34887	565	12.9031	18.2020	7.67746	2.77511	2.49442	0.05
5	1	16.7745	1.74260	565	13.3517	20.1973	-6.25451	-2.46528	-2.03210	0.05

Obs	ID	Pred	StdErrPred	DF	Alpha	Lower	Upper	Resid	ScaledResid	ScaledDep	StudentResid	PearsonResid
1	1	19.7472	0.66886	565	0.05	18.4334	21.0609	-11.8072	-1.52063	1.02258	-1.52630	-1.52063
2	1	20.1684	0.58520	565	0.05	19.0190	21.3179	-4.5184	1.11420	2.25027	-0.61509	-0.61315
3	1	20.5480	0.53952	565	0.05	19.4883	21.6077	-7.0380	-0.22144	0.75212	-0.98098	-0.97822
4	1	20.9443	0.53017	565	0.05	19.9029	21.9856	2.2857	1.97969	2.79848	0.31786	0.31699
5	1	21.4846	0.56147	565	0.05	20.3818	22.5874	-10.9646	-2.56764	-1.90242	-1.48193	-1.47768

```
*Now let's plot the studentized residuals by time relative to menarche to view any
patterns;
*First, we'll create a loess smooth plot;
proc loess data=BF_pred plots=none;
 ods output outputstatistics=out_BF_pred;
 model StudentResid=Time_r_men;
run;

proc sort data=out_BF_pred;
 by Time_r_men DepVar;
run;
*Now we'll view the loess smooth line by the actual values;
proc sgplot data=out_BF_pred;
 scatter x=Time_r_men y=DepVar/ markerattrs=(color=black size=5) ;
 series x=Time_r_men y=pred/ lineattrs=(color=blue thickness=5) ;
 run;
```




```
*Lastly let's check the normality of the residuals using the scaled version;
proc univariate data=BF_pred2 normal;
 var ScaledResid;
 histogram ScaledResid;
 qqplot ScaledResid;
 run;
```

Moments			
N	1049	**Sum Weights**	1049
Mean	-0.0022715	**Sum Observations**	-2.3827516
Std Deviation	0.99904276	**Variance**	0.99808644
Skewness	0.0262867	**Kurtosis**	0.28659264
Uncorrected SS	1046	**Corrected SS**	1045.99459
Coeff Variation	-43982.589	**Std Error Mean**	0.03084582

Basic Statistical Measures			
Location		**Variability**	
Mean	-0.00227	**Std Deviation**	0.99904
Median	-0.04586	**Variance**	0.99809
Mode	.	**Range**	7.66893
		Interquartile Range	1.33801

Tests for Location: Mu0=0				
Test	**Statistic**		**p Value**	
Student's t	**t**	-0.07364	**Pr > |t|**	0.9413
Sign	**M**	-14.5	**Pr >= |M|**	0.3873
Signed Rank	**S**	-2419.5	**Pr >= |S|**	0.8054

Tests for Normality				
Test	**Statistic**		**p Value**	
Shapiro-Wilk	**W**	0.997273	**Pr < W**	0.0730
Kolmogorov-Smirnov	**D**	0.02304	**Pr > D**	>0.1500
Cramer-von Mises	**W-Sq**	0.071804	**Pr > W-Sq**	>0.2500
Anderson-Darling	**A-Sq**	0.501229	**Pr > A-Sq**	0.2153

```
*This dataset we'll look at more form influential observations;

data dental;
input Person Gender $ Y1 Y2 Y3 Y4;
y=y1; Age=8; output;
y=y2; Age=10; output;
y=y3; Age=12; output;
y=y4; Age=14; output;
datalines;
           1        F       21.0    20.0    21.5    23.0
           2        F       21.0    21.5    24.0    25.5
           3        F       20.5    24.0    24.5    26.0
           4        F       23.5    24.5    25.0    26.5
           5        F       21.5    23.0    22.5    23.5
           6        F       20.0    21.0    21.0    22.5
           7        F       21.5    22.5    23.0    25.0
           8        F       23.0    23.0    23.5    24.0
           9        F       20.0    21.0    22.0    21.5
          10        F       16.5    19.0    19.0    19.5
          11        F       24.5    25.0    28.0    28.0
          12        M       26.0    25.0    29.0    31.0
          13        M       21.5    22.5    23.0    26.5
          14        M       23.0    22.5    24.0    27.5
          15        M       25.5    27.5    26.5    27.0
          16        M       20.0    23.5    22.5    26.0
          17        M       24.5    25.5    27.0    28.5
          18        M       22.0    22.0    24.5    26.5
          19        M       24.0    21.5    24.5    25.5
          20        M       23.0    20.5    31.0    26.0
          21        M       27.5    28.0    31.0    31.5
          22        M       23.0    23.0    23.5    25.0
          23        M       21.5    23.5    24.0    28.0
          24        M       17.0    24.5    26.0    29.5
          25        M       22.5    25.5    25.5    26.0
          26        M       23.0    24.5    26.0    30.0
          27        M       22.0    21.5    23.5    25.0
              ;
run;

*A quick note on the "iter" command.  It is the maximum number of additional iterations
PROC MIXED performs to update the fixed effects and covariance parameter estimates
following data point removal. When n > 0, the influence analysis is iterative. The
default is n = 0. If n > 0 and METHOD=REML (default) or METHOD=ML, the procedure updates
fixed effects and variance-covariance parameters after removing the selected observations
with additional Newton-Raphson iterations, starting from the converged estimates for the
entire data. The process stops for each observation or set of observations if the
convergence criterion is satisfied or the number of further iterations exceeds n.;

ods html;
ods graphics on;
proc mixed data=dental;
class Person Gender;
model y = Gender Age Gender*Age / influence(iter=5 effect=Person est);
repeated / type=ar(1) subject=Person;
run;
ods graphics off;
ods html close;
```

Model Information	
Data Set	WORK.DENTAL
Dependent Variable	y
Covariance Structure	Autoregressive
Subject Effect	Person
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**
Person	27	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
Gender	2	F M

Dimensions	
Covariance Parameters	2
Columns in X	6
Columns in Z	0
Subjects	27
Max Obs per Subject	4

Number of Observations	
Number of Observations Read	108
Number of Observations Used	108
Number of Observations Not Used	0

Covariance Parameter Estimates		
Cov Parm	**Subject**	**Estimate**
AR(1)	Person	0.6245
Residual		5.2145

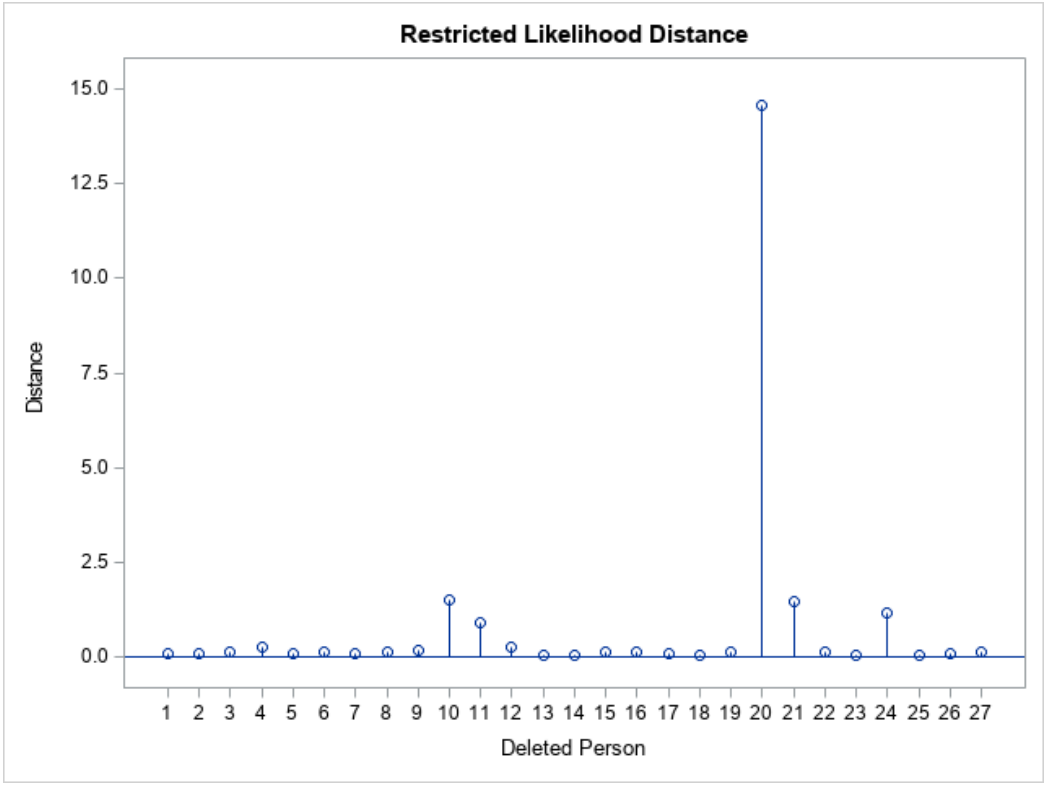
Fit Statistics	
-2 Res Log Likelihood	444.6
AIC (Smaller is Better)	448.6
AICC (Smaller is Better)	448.7
BIC (Smaller is Better)	451.2

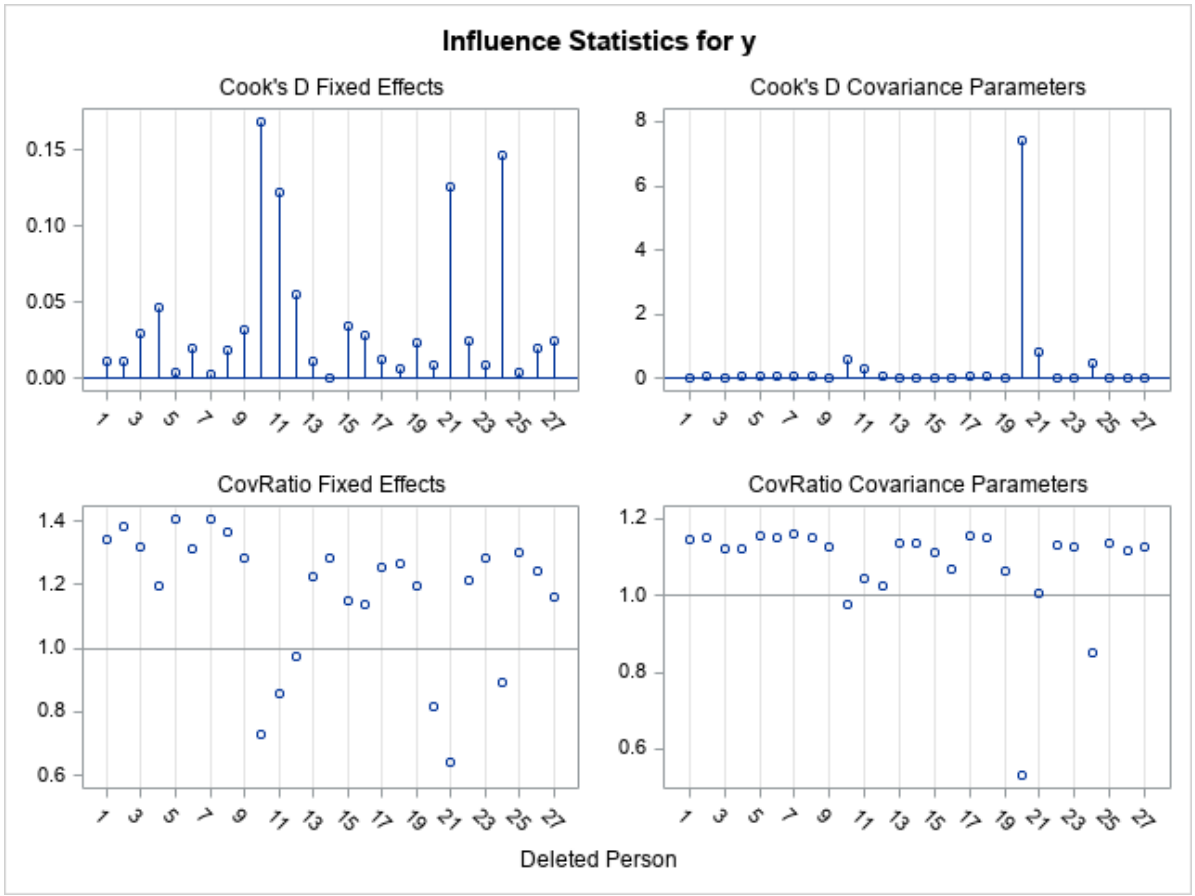
Null Model Likelihood Ratio Test		
DF	**Chi-Square**	**Pr > ChiSq**
1	38.97	<.0001

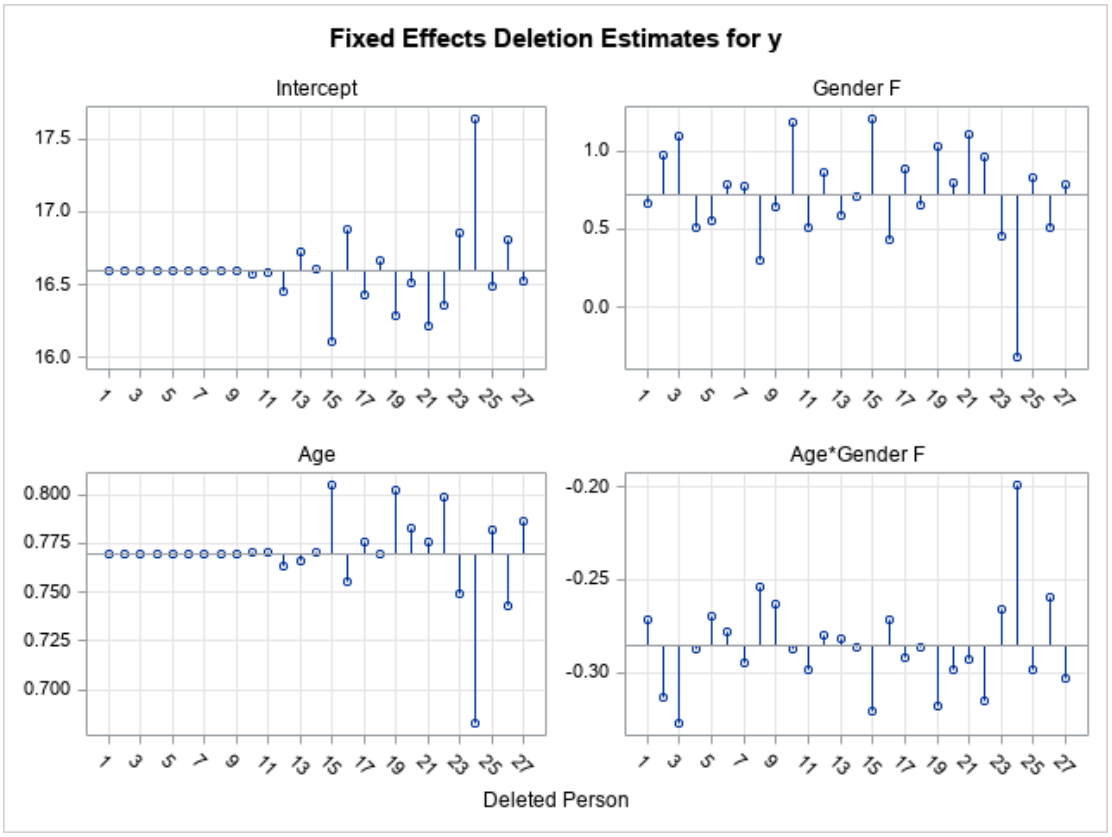
Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
Gender	1	25	0.11	0.7376
Age	1	79	46.77	<.0001
Age*Gender	1	79	2.43	0.1233

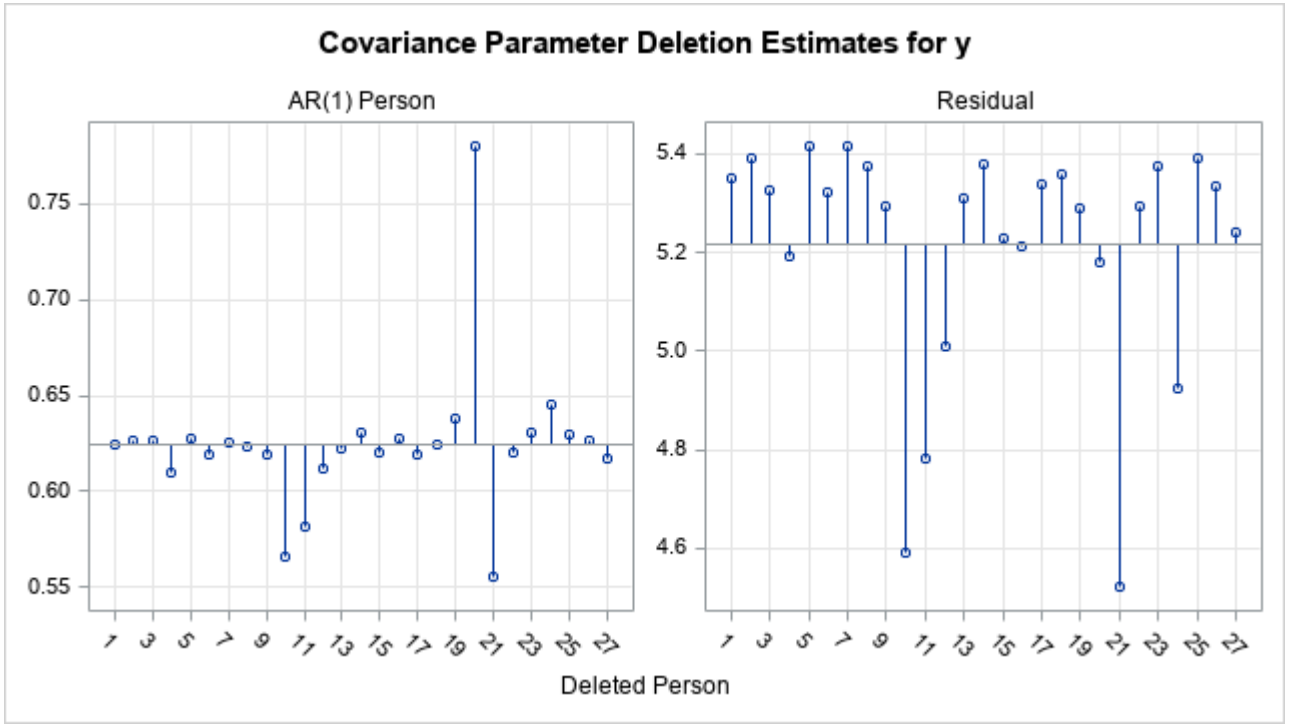
Influence Diagnostics for Levels of Person									
Person	Number of Observations in Level	Iterations	PRESS Statistic	Cook's D	MDFFITS	COVRATIO	COVTRACE	Cook's D CovParms	MDFFITS CovParms
1	4	1	9.6412	0.01119	0.00992	1.3415	0.3098	0.03565	0.0326
2	4	1	3.7181	0.01154	0.01016	1.3815	0.3415	0.05100	0.0456
3	4	1	10.8479	0.02907	0.02591	1.3174	0.2903	0.01812	0.0165
4	4	2	24.4013	0.04667	0.04347	1.1961	0.1888	0.05982	0.0534
5	4	1	1.6900	0.00334	0.00294	1.4051	0.3600	0.06207	0.0548
6	4	1	11.7185	0.01981	0.01776	1.3158	0.2891	0.05000	0.0463
7	4	1	1.2187	0.00307	0.00269	1.4069	0.3613	0.06890	0.0611
8	4	1	5.0250	0.01807	0.01592	1.3675	0.3305	0.05455	0.0495
9	4	1	13.1216	0.03196	0.02875	1.2864	0.2649	0.03419	0.0319
10	4	2	85.2778	0.16899	0.18880	0.7277	0.2899	0.61657	0.6538
11	4	2	69.6436	0.12270	0.12869	0.8583	0.1389	0.32291	0.3103
12	4	2	38.8782	0.05476	0.05435	0.9723	0.0252	0.04632	0.0488
13	4	1	14.8373	0.01148	0.01060	1.2248	0.2102	0.02479	0.0231
14	4	1	6.8024	0.00082	0.00074	1.2851	0.2613	0.03085	0.0272
15	4	1	20.5090	0.03478	0.03244	1.1510	0.1454	0.00823	0.0077
16	4	1	23.5720	0.02813	0.02629	1.1361	0.1319	0.00310	0.0030
17	4	1	8.1930	0.01197	0.01103	1.2538	0.2350	0.05914	0.0545
18	4	1	9.1204	0.00677	0.00619	1.2680	0.2468	0.04130	0.0377
19	4	2	15.6824	0.02364	0.02219	1.1953	0.1854	0.03370	0.0347
20	4	4	42.9162	0.00831	0.00825	0.8175	0.0354	7.42313	19.6236
21	4	2	91.5142	0.12647	0.14980	0.6399	0.4057	0.82320	0.8390
22	4	1	14.2947	0.02470	0.02276	1.2119	0.1991	0.02903	0.0271
23	4	1	6.4046	0.00940	0.00859	1.2817	0.2584	0.02976	0.0263
24	4	2	42.5667	0.14725	0.15055	0.8947	0.1059	0.48613	0.5319
25	4	1	3.6903	0.00450	0.00409	1.2990	0.2727	0.03991	0.0351
26	4	1	7.9353	0.01924	0.01766	1.2430	0.2258	0.01989	0.0181
27	4	2	21.6103	0.02450	0.02298	1.1631	0.1564	0.02558	0.0237

Influence Diagnostics for Levels of Person				
Person	COVRATIO CovParms	COVTRACE CovParms	RMSE without deleted level	Restricted Likelihood Distance
1	1.1489	0.1448	2.31316	0.078
2	1.1493	0.1469	2.32185	0.093
3	1.1232	0.1211	2.30818	0.132
4	1.1237	0.1234	2.27905	0.248
5	1.1579	0.1560	2.32689	0.072
6	1.1510	0.1458	2.30704	0.127
7	1.1609	0.1583	2.32712	0.077
8	1.1499	0.1464	2.31862	0.122
9	1.1269	0.1232	2.30052	0.160
10	0.9747	0.0896	2.14317	1.492
11	1.0449	0.1108	2.18660	0.899
12	1.0277	0.0361	2.23808	0.275
13	1.1387	0.1345	2.30444	0.070
14	1.1352	0.1351	2.31932	0.032
15	1.1126	0.1098	2.28639	0.145
16	1.0687	0.0677	2.28347	0.114
17	1.1559	0.1505	2.31089	0.103
18	1.1495	0.1455	2.31460	0.066
19	1.0665	0.0702	2.29951	0.129
20	0.5307	0.1129	2.27626	14.546
21	1.0053	0.1676	2.12644	1.463
22	1.1312	0.1273	2.30118	0.126
23	1.1284	0.1285	2.31849	0.064
24	0.8505	0.1552	2.21905	1.170
25	1.1378	0.1376	2.32227	0.055
26	1.1154	0.1140	2.30915	0.095
27	1.1264	0.1232	2.28908	0.125


Restricted Likelihood Distance
Distance
15.0
12.5
10.0
7.5
5.0
2.5
0.0
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
Deleted Person


Influence Statistics for y
Cook's D Fixed Effects
Cook's D Covariance Parameters
CovRatio Fixed Effects
CovRatio Covariance Parameters
Deleted Person


Fixed Effects Deletion Estimates for y
Intercept
Gender F
Age
Age*Gender F
Deleted Person


Covariance Parameter Deletion Estimates for y
AR(1) Person
Residual
Deleted Person