BioStatistics Assignment 2

Matthew Parfitt

Student Number: 21679495

Table Of Contents

The aim of which is to study the evolution of the head circumference of infants born to HIV positive mothers, and to study the effects of various explanatory variables on their evolution.

Introduction:

The data set considered in this report results from a longitudinal observational study in paediatric infectious disease. The aim of which is to study the evolution of the head circumference of infants born to HIV positive mothers, and to study the effects of various explanatory variables on their evolution. There will be summary statistics explaining the effect of the explanatory variables on the response variable which will be head circumference. The sole focus will be on patients with 6 or more measurements taken, which will result in our study comprising of 306 patients in total. The report will be final summary of the main important factors that play a role and will be explained, this is after the data description and summary statistics which will identify the random effects model that will be needed, most of this will be shown in the appendix. The experiment was comprised of many explanatory variables such as premature birth, whether the mother started on ARV treatment within 4 weeks of delivery, maternal viral load group, whether the mother was symptomatic at the time of giving birth and whether the baby has low body weight, these will be narrowed down and explanations will be given. With regards to the response variables, which are weight, length and head circumference, we will be solely focusing on head circumference as it is the response variable we have been assigned to test.

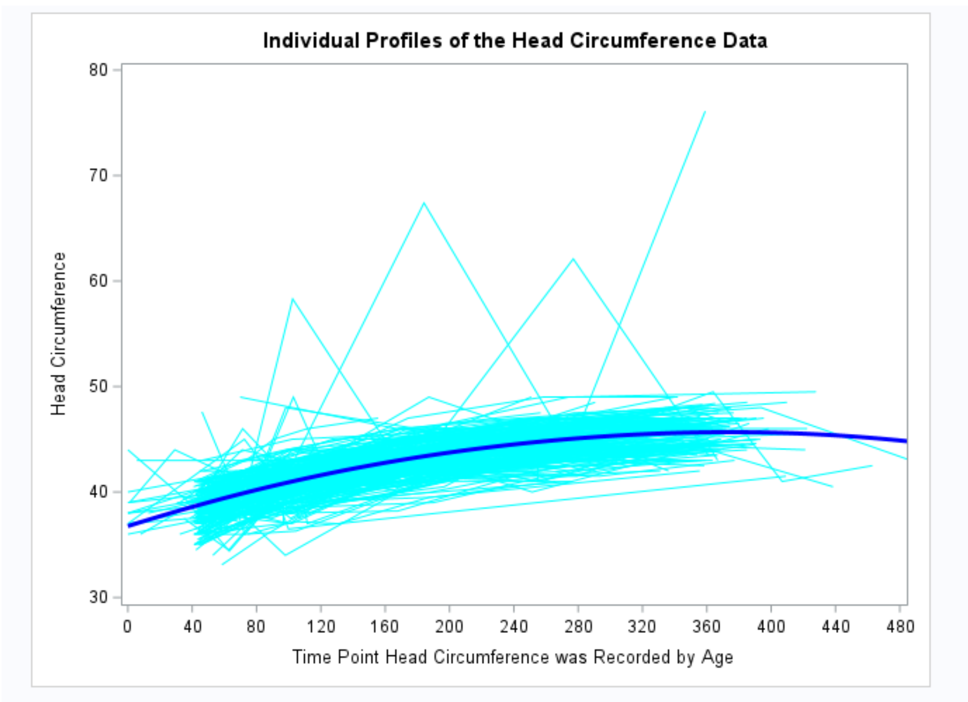
Explanation of Variables:

- ID - Patients ID.
- Premature - Was the infant born prematurely (0 = No; 1 = Yes).
- TreatLess4Weeks - Was the mother started on ARV treatment within 4 weeks of delivery (0 = No; 1 = Yes).
- VLGroup - Maternal viral load group (1 = Undetectable, 2 = (0 – 1000], 3 = "> 1000").
- Symptomatic - Was the mother symptomatic at the time of the infants birth (0 = No; 1 = Yes).
- LowBWeight - When born, did the infant have a lower than normal birth weight (0 = No; 1 = Yes).
- Gender - 0 = Boys and 1 = Girl.
- HeadCir - Head Circumference (cm).
- Age - The age of the infant at each measurement (0 = Birth/Baseline).

Report:

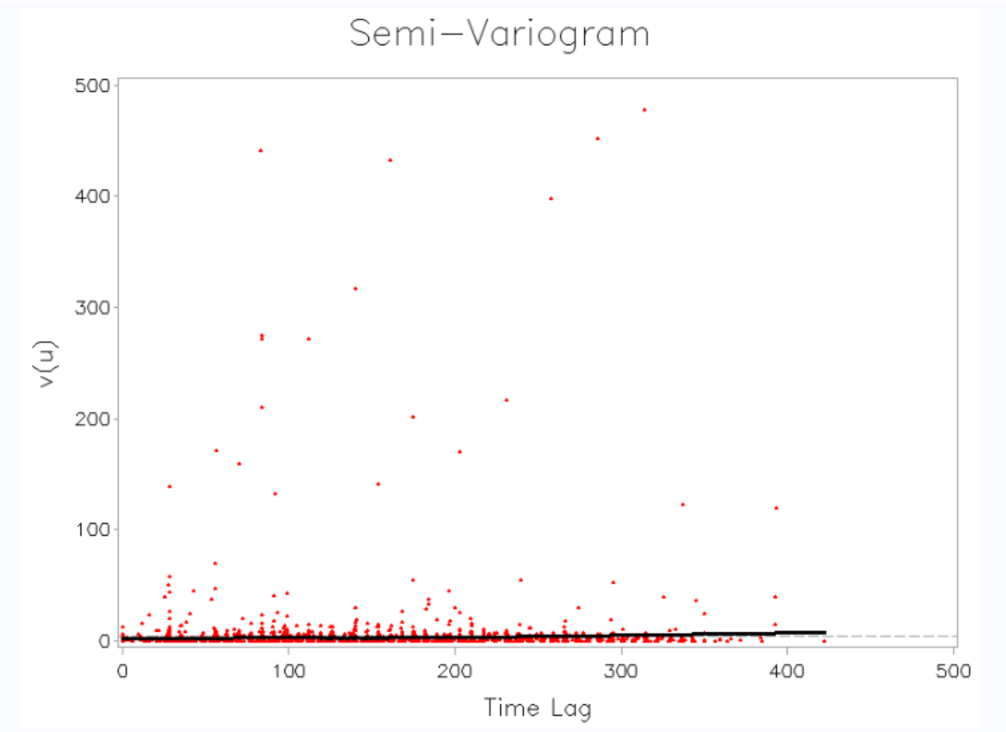
The first step to have an understanding of the different variables is through summary statistics of them. Referring to ***Figure 1.1*** of the appendix, we can see the mean is 42.1cm that the minimum and maximum of the variable Head Circumference is 33.1cm and 76.1cm respectively, which was taken at different ages but helps us to understand the range we are observing. More of the mean, minimums and maximums of the variables were shown in ***Figure 1.2***. After many different summary statistics through SAS, a better understanding of the different variables was now undertaken. After a few frequency distributions were run, many left out the appendix because they were insignificant to our research task, there was one particular that did show a significant relationship. Referring to ***Figure 1.6***, there is a relationship between premature and low body weight evident, this can be seen by observing the frequency distributions as well as studying the table which depicts that 75.56% of the children born were not low body weight and not premature with 10.37% of the babies being premature and born with low body weight, this might seem obvious but it doesn't follow the same pattern as the other frequency plots tested before. The 10.37% is greater than the 7.78% of babies that are born with low body weight but not premature which shows us there is a significant relationship between them which is what we expected.

Analysing ***Figure 1.3*** and ***Figure 1.4*** it can be seen that the distribution of head circumference is between 33cm and 49cm, and with the age distribution heavily weighted between 0 and 240 days. Therefore this can play an important role later in the study where that distribution can be taken to narrow down the range and therefore hopefully lead to more narrow results. Understanding the individual profiles of all the head circumference data can be very important to understanding the pattern of the data when in regards to the summary section of this report.



Therefore analysing the above individual profile plot with a smoothed trend line allows us understand the distribution of the data as a whole, seeing the head circumference increases with age to where around the 380 day mark where the data seems to stop increasing at the gradient it was before, allowing us to understand the growth slightly slow after 300 day mark.

Moving on from summary statistics, a model will have to be decided on and fitted, a longitudinal model is fit where an unstructured covariance structure is disregarded as the time intervals between measurements within the patients is not fixed and vary considerably, a compound symmetry covariance structure is suited the best for this specific data set and range, this is after testing out different spatial covariance structures before. In ***Figure 2.1***, running the longitudinal model previously spoken about, an explanation of residual covariance matrix as well as the correlation matrix based on the covariance matrix is explained in the appendix. There are no higher order terms significant at the 0.05 significance level however the model will be reduced at a later stage to specifically focus on the variables that play a significant role in regards to the variable head circumference. However firstly a look at the smoothed variance function and semi variogram will be undertaken.



The main conclusions from the above semi variogram, is that there is some measurement error in the process, which can come from error in the measurement process itself which is what is expected. To add to this the fitted line does reach the process variance, therefore there is no error due to random effects in the model. Now reducing the model will be the next important step in the process.

Figure 2.4 shows the first step in the process of reducing the model, the model is reduced by taking out the higher-order term with the highest p-value and then the model is rerun until all of the higher order terms end up being removed as they did in this specific test as they were insignificant. ***Figure 2.5*** shows the final model that has been reduced to where it is just the fixed effects remaining, this can be seen within the figure. The information received from the G matrix is that the intercepts and the slopes for age are slightly positively correlated, this allows us to understand that as a child grows older their head circumference increases which is obvious. When analysing the final model there are two explanatory variables that show a significant response which are TreatLess4Weeks and LowBWeight.

Conclusion:

Whether the mother received ARV treatments 4 weeks prior to giving birth seems to show that it plays a significant part in the head circumference of a child when they are born and as they grow for the first 400 days, the estimate seems to show that babies born to mothers that received ARV treatment 4 weeks prior to giving birth are estimated to have a head circumference of 1.5cm larger than a baby that is born to a mother that never received ARV treatment. Low body weight also plays a significant part as it is estimated that if a child is born with low body weight, their head circumference is estimated to be 1.8cm smaller than a baby that is not born with low body weight. The other explanatory variables do not seem to play a massive role in a child's head circumference.

Misspecifications and Recommendations:

Referring to ***Figure 2.6***, there are some outliers that can be seen in the 3 graphs for the ID's that are patients 1112, 788, 2012 and 1591. These 4 can be seen to be outliers and a further investigation into these 4 patients would be highly recommended to understand why they are so different from the rest of the ID's. There is also measurement error that was found in the semi-variogram, this can be improved through more focus on the measurement process itself as in when the tests were undertaken.

Appendix:

Summary Statistics:

First step was to clean and sort the data to the specified needs that we have to use which is observations of 6 or more:

```
proc print data =ass2.paed;
run;

proc freq data=ass2.paed nlevels;
tables id / out=visits nocum nopercent;
run;
proc print data=visits;
run;

/*Create a column where the of cumm visits*/
data visits1;
set visits;
keep count id;
run;
proc print data = visits1;
run;

/* Merging columns back to the full table*/
data visits1;
merge ass2.paed visits1;
by id;
run;
 proc print  data=visits1;
 run;


proc sort data=visits1 out=sorted_data;
where count >=6;
by ID;
run;
proc print data=sorted_data;
run;

data main_data;
 set sorted_data;
 drop count Weight Length;
 proc print data= main_data;
 run;



proc freq data = main_data  nlevels;
tables id;
run;


/* can see max, min, std dev and means etc */
proc means data=main_data;
var HeadCr;
run;
```

Analysis Variable : HeadCr				
N	Mean	Std Dev	Minimum	Maximum
1724	42.1024710	3.1899291	33.1000000	76.1000000

Figure 1.1

Interpretation: As can be seen in the above, the mean of the variable Head Circumference is 42.102cm with a minimum of 33.1cm and a maximum of 76.1cm.

```
proc means data = main_data;
var;
run;
```

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
ID	study_id	2030	514.8965517	309.0514306	9.0000000	1136.00
Premature		271	0.1660517	0.3728155	0	1.0000000
TreatLess4Weeks		258	0.0426357	0.2024270	0	1.0000000
Symptomatic		286	0.1223776	0.3282959	0	1.0000000
LowBWeight		288	0.1770833	0.3824037	0	1.0000000
HeadCr		1724	42.1024710	3.1899291	33.1000000	76.1000000
Age		2030	137.3935205	112.3273522	0	486.7569444

Figure 1.2

```
proc univariate data=main_data;
var headcr;
histogram;
run;

/* Histogram showing the distribution of head circumference and how its
skewed*/
```

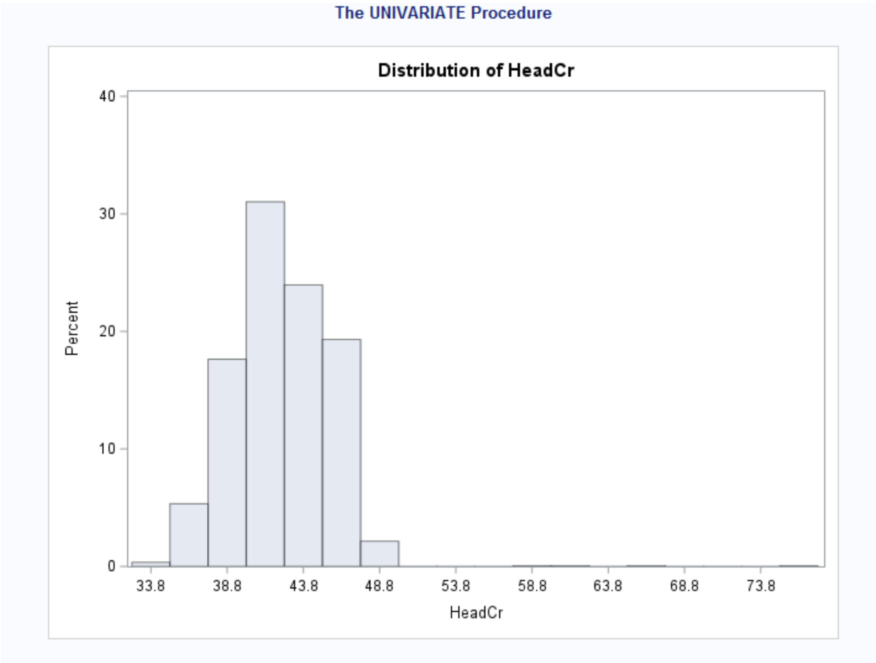


Figure 1.3

```
Title "Summary of Age Variable Distribution";
proc univariate data = main_data noprint;
 var Age;
HISTOGRAM;
 run;
```

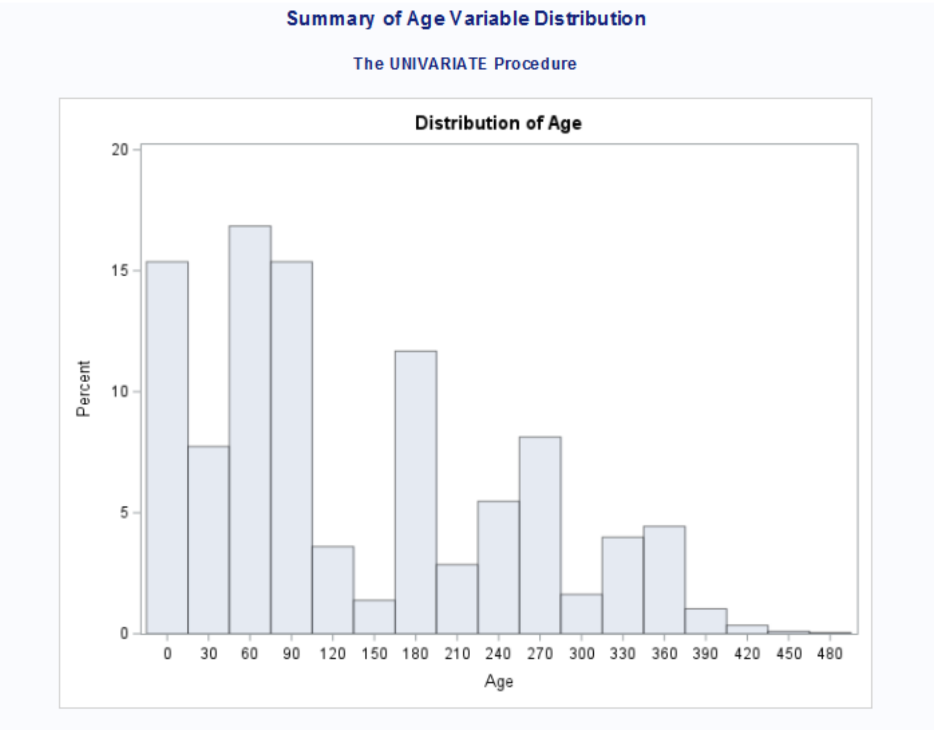


Figure 1.4

```
proc means data=main_data;
var headcr age;
run;
```

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
HeadCr	1724	42.1024710	3.1899291	33.1000000	76.1000000
Age	2030	137.3935205	112.3273522	0	486.7569444

Interpretation: From the above histogram and analysis of variable age we can see with a mean of age 137.40 and a maximum of 486.76 as well as looking at the distribution of the histogram and how it is positively skewed as it is skewed to the right therefore taking an age range between 0 and 240 can be assumed to be sufficient for the purposes of this test later on in our analysis.

```
proc univariate data=main_data;
var HeadCr;
run;
```

Variable: HeadCr

Moments			
N	1724	Sum Weights	1724
Mean	42.102471	Sum Observations	72584.66
Std Deviation	3.18992911	Variance	10.1756478
Skewness	1.12245455	Kurtosis	9.86676639
Uncorrected SS	3073526.18	Corrected SS	17532.6411
Coeff Variation	7.57658408	Std Error Mean	0.07682674

Basic Statistical Measures			
Location		Variability	
Mean	42.10247	Std Deviation	3.18993
Median	42.00000	Variance	10.17565
Mode	41.00000	Range	43.00000
		Interquartile Range	4.40000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	548.0185	Pr > |t|	<.0001
Sign	M	862	Pr >= |M|	<.0001
Signed Rank	S	743475	Pr >= |S|	<.0001

Quantiles (Definition 5)	
Level	Quantile
100% Max	76.1
99%	48.4
95%	47.0
90%	46.0
75% Q3	44.4
50% Median	42.0
25% Q1	40.0
10%	38.0
5%	37.0
1%	36.0
0% Min	33.1

Extreme Observations			
Lowest		Highest	
Value	Obs	Value	Obs
33.1	2026	49.5	810
34.0	1395	58.3	788
34.0	1052	62.1	1196
34.4	240	67.4	1112
34.5	1983	76.1	1444

Missing Values			
Missing Value	Count	Percent Of	
		All Obs	Missing Obs
.	306	15.07	100.00

Interpretation: This shows us that we have the right amount of patients which is 306, now we have cleaned our data set correctly and can carry on.

```
proc sgplot data=main_data nocycleattrs noautolegend;
series y=headcr x=age / group=id lineattrs=(color=cyan pattern=1);
pbspline y=headcr x=age / nomarkers smooth=50 nknots=5
lineattrs=(color=blue pattern=1 thickness=3);
xaxis values=(0 to 480 by 40) label='Time Point Head Circumference was
Recorded by Age';
yaxis values=(30 to 80 by 10) label='Head Circumference';
title 'Individual Profiles of the Head Circumference Data';
run;
```

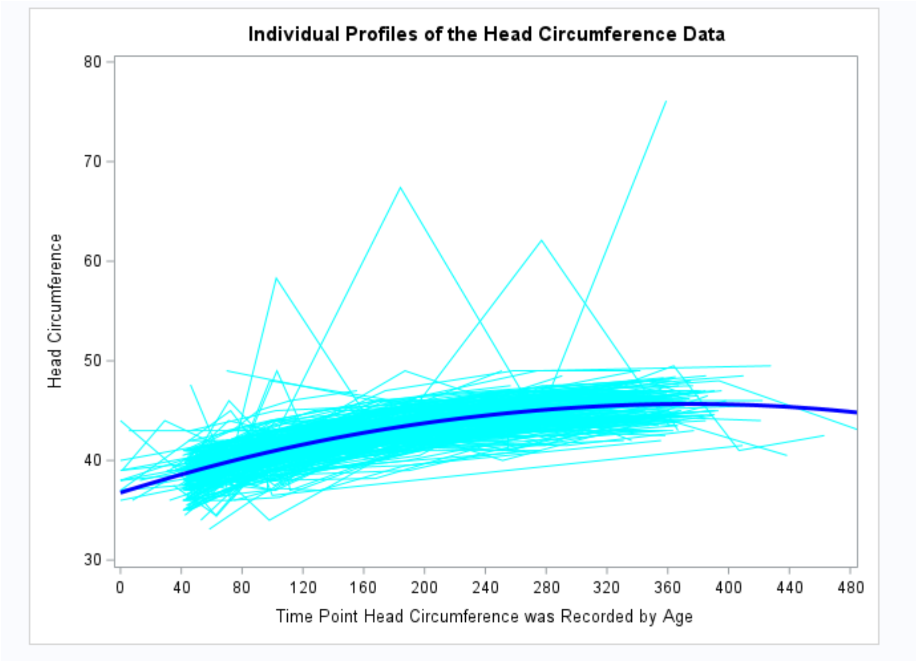


Figure 1.5

The average trend shows that the Head Circumference average is fairly constant around 40cm but rises slightly then seems to straighten and slightly decrease.

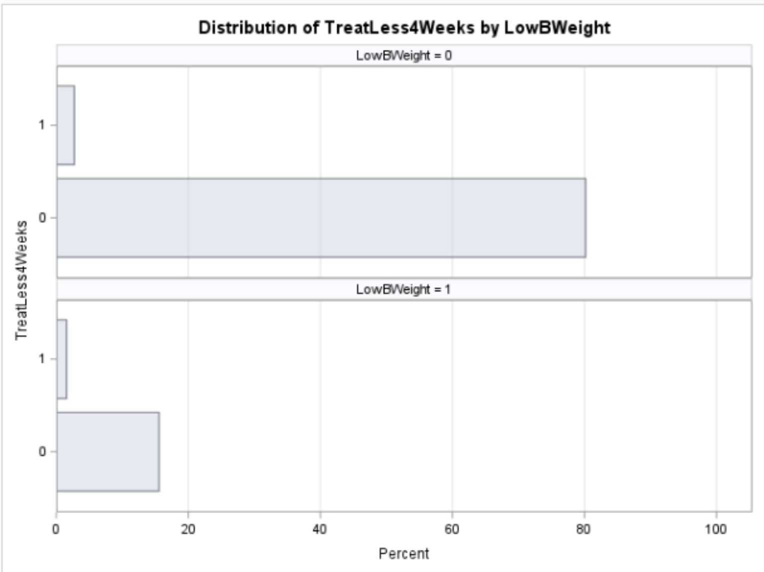
```
ods graphics=on;
   ods noproctitle;
title1"Mothers with ARV treatment within 4 weeks of deivery and whether the
baby has low body weight";
   proc freq data=main_data order=freq nlevels;
   tables TreatLess4Weeks*LowBWeight / nocum
plots=freqplot(orient=horizontal scale=percent);
   run;
```

Mothers with ARV treatment within 4 weeks of deivery and whether the baby has low body weight

Number of Variable Levels			
Variable	Levels	Missing Levels	Nonmissing Levels
TreatLess4Weeks	3	1	2
LowBWeight	3	1	2

Frequency / Percent / Row Pct / Col Pct

Table of TreatLess4Weeks by LowBWeight	LowBWeight		
TreatLess4Weeks	0	1	Total
0	206 80.16 83.74 96.71	40 15.56 16.26 90.91	246 95.72
1	7 2.72 63.64 3.29	4 1.56 36.36 9.09	11 4.28
Total	213 82.88	44 17.12	257 100.00
Frequency Missing = 1773			



Interpretation: With the above frequency plot and test, it can be seen there is not much of a relationship as with the mother not having ARV treatment within 4 weeks of giving birth and the child having normal body weight is 82.88% therefore this seems to not effect each other.

```
ods graphics=on;
   ods noproctitle;
title1"Whether babies were born prematurely and have low Body Weight or
not";
   proc freq data=main_data order=freq nlevels;
   tables Premature*LowBWeight/ nocum plots=freqplot(orient=horizontal
scale=percent);
   run;
```

Whether babies were born prematurely and have low Body Weight or not

Number of Variable Levels			
Variable	Levels	Missing Levels	Nonmissing Levels
Premature	3	1	2
LowBWeight	3	1	2

Frequency Percent Row Pct Col Pct	Table of Premature by LowBWeight			
		LowBWeight		
	Premature	0	1	Total
	0	204 75.56 90.67 92.31	21 7.78 9.33 42.86	225 83.33
	1	17 6.30 37.78 7.69	28 10.37 62.22 57.14	45 16.67
	Total	221 81.85	49 18.15	270 100.00
	Frequency Missing = 1760			

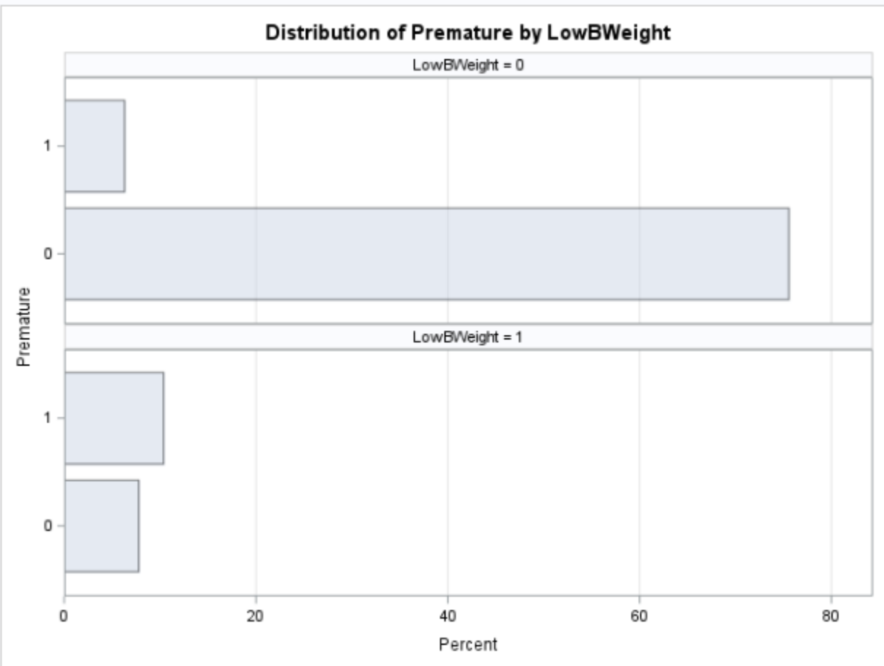
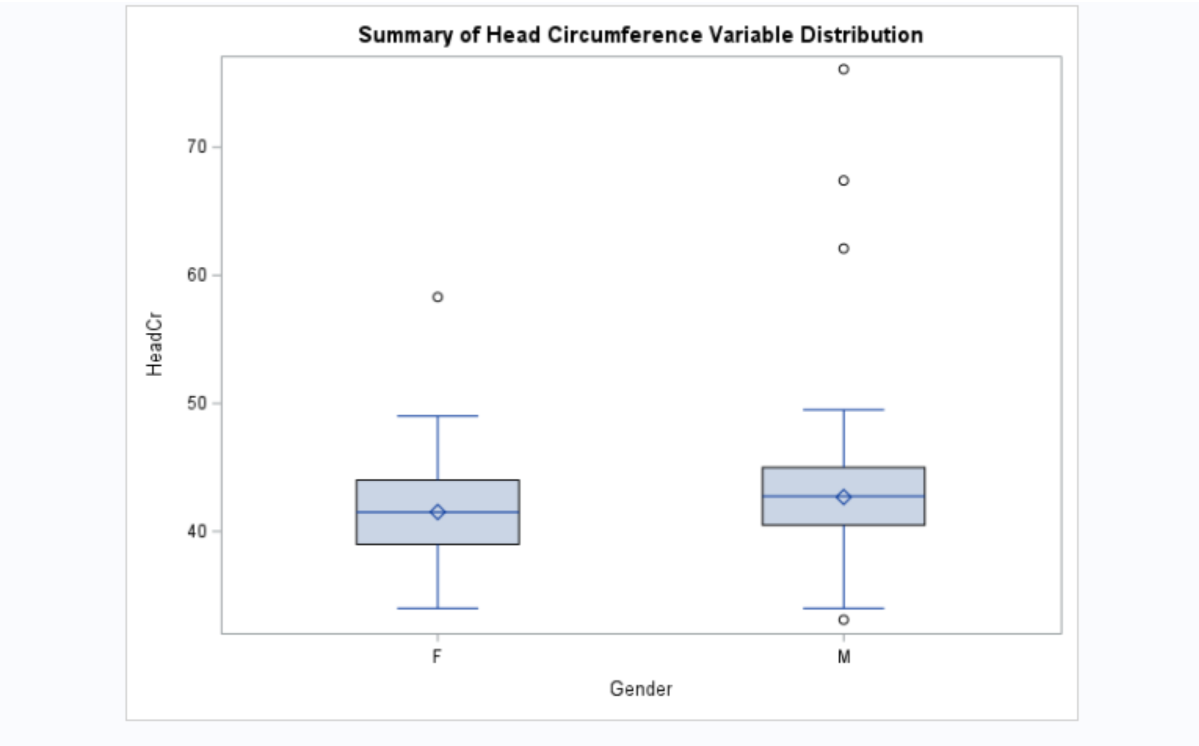


Figure 1.6

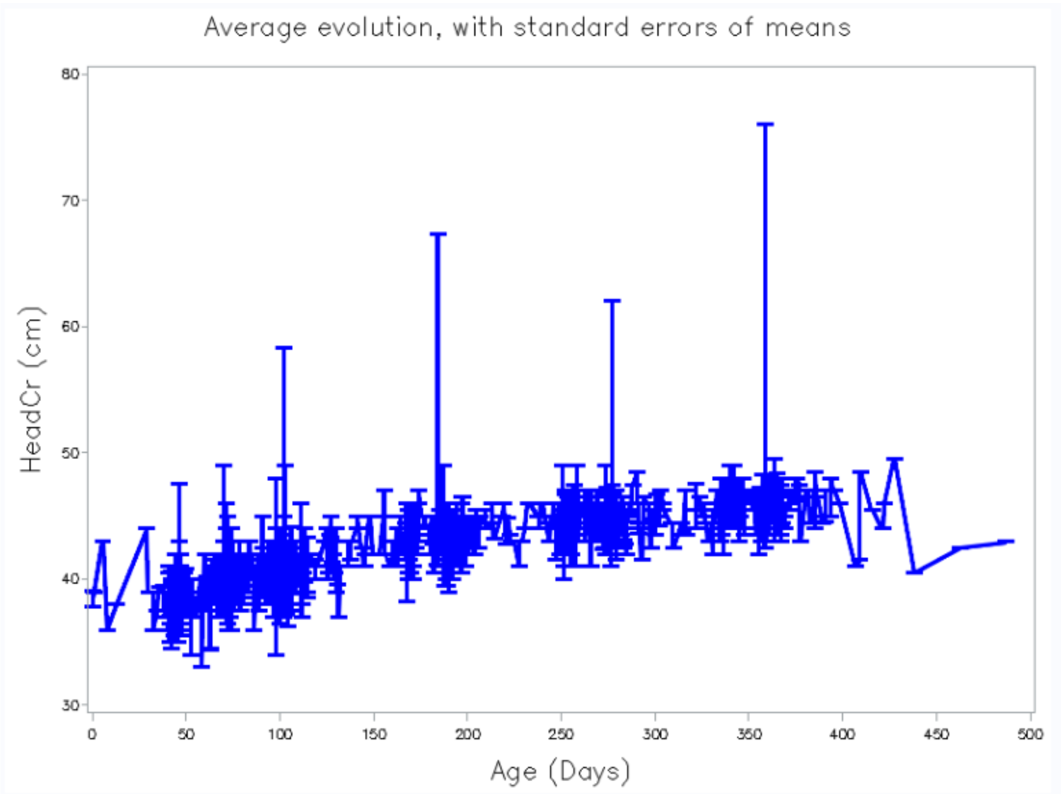
Interpretation: As we can see in the above frequency plot, there is an obvious relationship between Low Body Weight and Premature Babies, as there is a higher number of babies that have low body weight when they are born prematurely. This can be seen in the frequency plot with looking at the distributions of the graphs.

```
/* Getting the Gender column to be filled */
data main_data_new;
set main_data;
retain Gender1;
by ID;
if first.ID then Gender1 = Gender;
else Gender=Gender1;
drop Gender1;
run;

proc print data=main_data_new;
run;


/* Boxplots showing distribution of Head Circumference by Gender */
 PROC SGPLOT  DATA = main_data_new;
   VBOX HeadCr / category = Gender;
RUN;
```



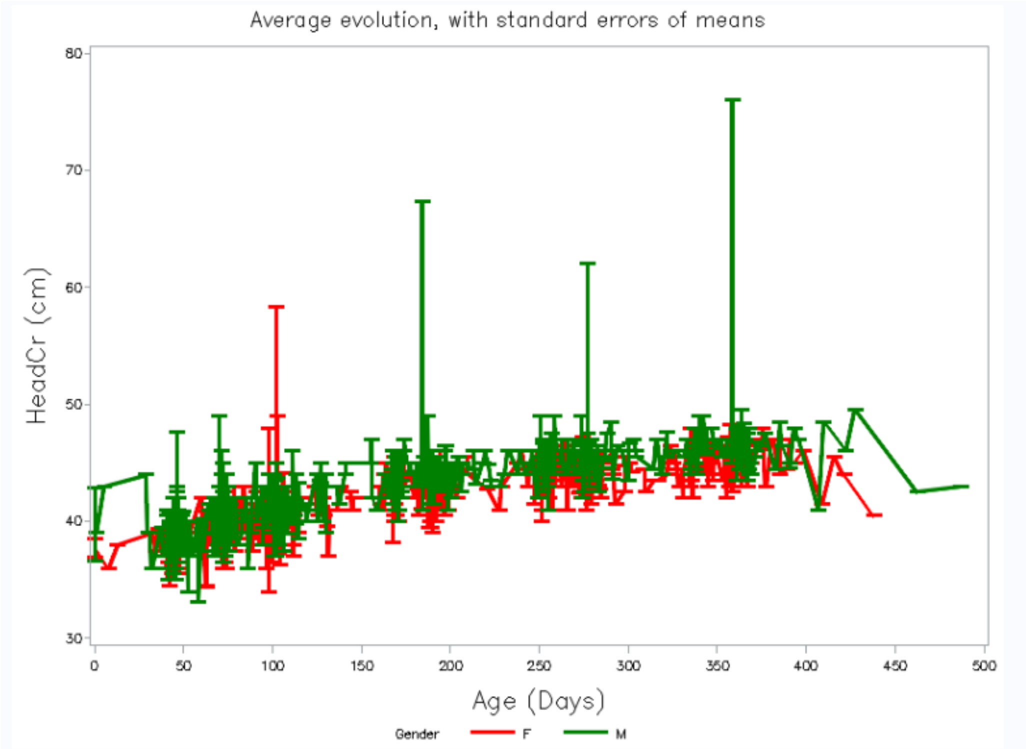
Interpretation: A few outliers can be seen with in the boxplots of male and female, with the male outliers seeming to be more extreme than the females.

```
/* Graph showing headCr and age as a whole */

# Average evolution with standard error bars;
goptions reset=all ftext=simplex rotate=landscape;
proc gplot data=main_data_new;
plot HeadCr*age/ haxis=axis1 vaxis=axis2;
symbol c=blue i =std1mjt w=2 mode=include;
axis1 label=(h=2 'Age (Days)') order=(0 to 500 by 50) minor=none;
axis2 label=(h=2 A=90 'HeadCr (cm)') minor=none;
title h=2 'Average evolution, with standard errors of means';
run; quit;
```



Interpretation: As seen above in the graph, it can be seen that as the age of a child increases their head circumference increases at a steady rate however there are some outliers that will be studied later on in this research task.

```
/*Graph showing male and female seperately*/
# Average evolution with standard error bars per Gender;
goptions reset=all ftext=simplex rotate=landscape;
proc gplot data = main_data_new;
plot HeadCr*age=Gender/haxis=axis1 vaxis=axis2 legend=legend1;
symbol1 c=red i =std2mjt w=2 mode=include;
symbol2 c=green i =std2mjt w=2 mode=include;
axis1 label=(h=2 'Age (Days)') order=(0 to 500 by 50) minor=none;
axis2 label=(h=2 A=90 'HeadCr (cm)') minor=none;
title1 'Average evolution, with standard errors of means';
legend1 label=('Gender');
run; quit;
```



Interpretation: In the above graph it is obvious that the head circumference of males babies are larger and grow at a faster pace than the head circumference of female babies. As well as there seeming to be more outliers in the male portion of the study than the female.

```
/* Setting all the explanatory variables to have filled out columns*/
data main_data_new1;
set main_data;
retain Premature1;
by ID;
if first.ID then Premature1 = Premature;
else Premature=Premature1;
drop Premature1;
run;
proc print data=main_data_new1;
run;
```

```
data main_data_new2;
set main_data_new1;
retain TreatLess4Weeks1;
by ID;
if first.ID then TreatLess4Weeks1 = TreatLess4Weeks;
else TreatLess4Weeks=TreatLess4Weeks1;
drop TreatLess4Weeks1;
run;
proc print data=main_data_new2;
run;

data main_data_new3;
set main_data_new2;
retain Symptomatic1;
by ID;
if first.ID then Symptomatic1 = Symptomatic;
else Symptomatic=Symptomatic1;
drop Symptomatic1;
run;
proc print data=main_data_new3;
run;

data main_data_new4;
set main_data_new3;
retain LowBWeight1;
by ID;
if first.ID then LowBWeight1 = LowBWeight;
else LowBWeight=LowBWeight1;
drop LowBWeight1;
run;
proc print data=main_data_new4;
run;

data main_data_new5;
set main_data_new4;
retain VLGroup1;
by ID;
if first.ID then VLGroup1 = VLGroup;
else VLGroup=VLGroup1;
drop VLGroup1;
run;
proc print data=main_data_new5;
run;

proc sort data=main_data_new4 out=main_age_sorted1;
where Age<240;
by ID;
run;
proc print data= main_age_sorted1;
run;
```

Interpretation: The above code is to fill out the columns of explanatory variables to make certain tests to follow easier to run.

```
proc mixed data=main_data_new5;
model Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight
Age*Premature Age*TreatLess4Weeks Age*VLGroup Age*Symptomatic
Age*LowBWeight
Age*Age Age*Age*Age / solution ddfm=kr(firstorder);
repeated / type=cs local subject=id r rcorr;
title 'Longitudinal Model with Compound Symmetry Covariance Structure';
run;
```

Estimated R Matrix for Subject 1					
Row	Col1	Col2	Col3	Col4	Col5
1	4.0853	1.4898	1.4898	1.4898	1.4898
2	1.4898	4.0853	1.4898	1.4898	1.4898
3	1.4898	1.4898	4.0853	1.4898	1.4898
4	1.4898	1.4898	1.4898	4.0853	1.4898
5	1.4898	1.4898	1.4898	1.4898	4.0853

Estimated R Correlation Matrix for Subject 1					
Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.3647	0.3647	0.3647	0.3647
2	0.3647	1.0000	0.3647	0.3647	0.3647
3	0.3647	0.3647	1.0000	0.3647	0.3647
4	0.3647	0.3647	0.3647	1.0000	0.3647
5	0.3647	0.3647	0.3647	0.3647	1.0000

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
Variance	ID	1.6488
CS	ID	1.4898
Residual		0.9467

Fit Statistics	
-2 Res Log Likelihood	2956.8
AIC (Smaller is Better)	2962.8
AICC (Smaller is Better)	2962.9
BIC (Smaller is Better)	2971.4

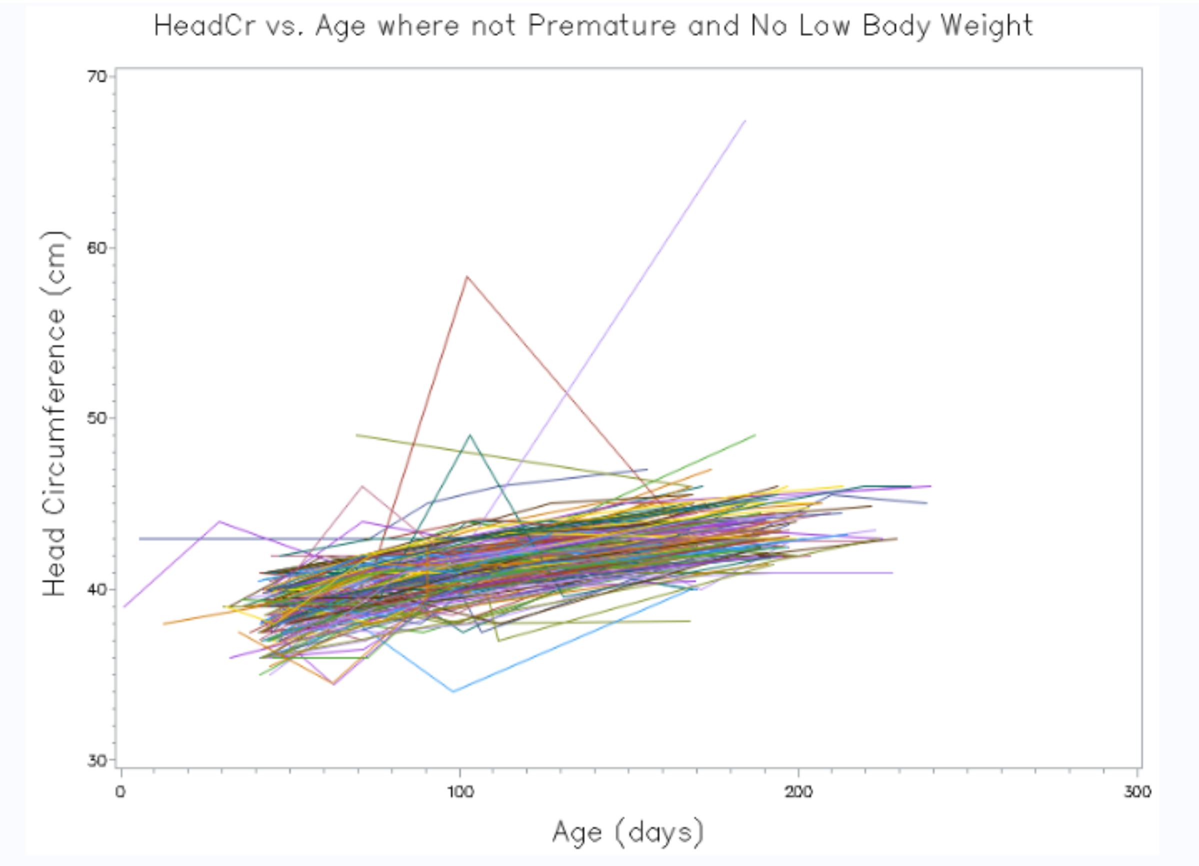
Null Model Likelihood Ratio Test		
DF	Chi-Square	Pr > ChiSq
2	125.60	<.0001

Figure 2.1

Interpretation: An Unstructured covariance Model would not be legitimate for this as data set as the time intervals are not the same across each patient. Therefore a Compound Symmetry Covariance Structure would be more appropriate. The R matrix represents the residual covariance matrix. The value in row 1 and column 1 represents the variance of the first measurement. The value in row 2 and column 2 represents the variance of the second measurement. The value in row 1 and column 2 represents the covariance of the first and second measurements. The R correlation matrix consists of the correlations of the measurements within patient. The correlations seem to stay constant through the study. The null model likelihood ratio test compares the fitted model to a model with an independent covariance structure. The test is significant, which indicates that the Compound Symmetry power covariance structure does a better job modelling the residual error compared to the independent covariance structure. There are no higher-order terms significant at the .05 level.

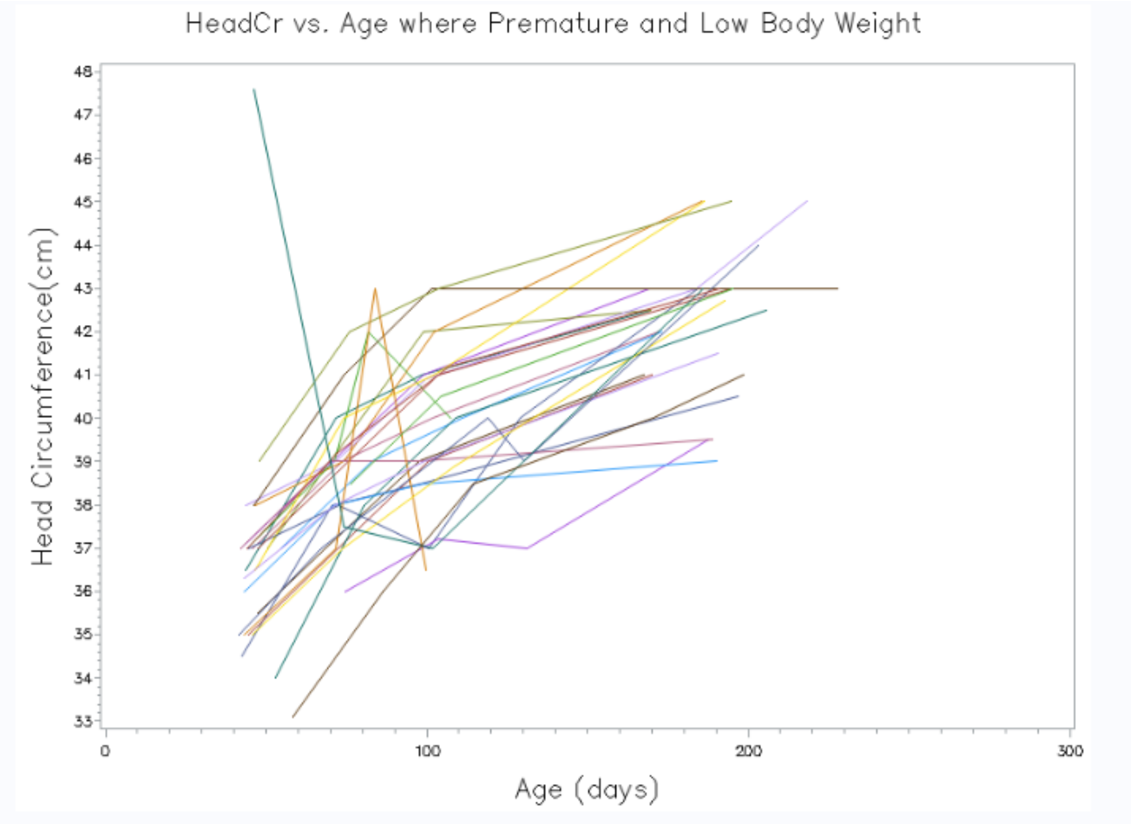
```
data allclear;
set main_head_data;
where Premature = 0 and LowBWeight=0;
run;
proc print data=allclear;
run;

data allnotclear;
set main_head_data;
where Premature = 1 and LowBWeight=1;
run;
proc print data=allnotclear;
run;

data atleastone;
set main_head_data;
where Premature = 1 or LowBWeight=1;
run;
proc print data=atleastone;
run;
```

```
/** Where LowBWeight=0 and Premature=0 */
# Plot individual profiles;
proc sort data = allclear;
by ID;
run;
goptions reset=all ftext=simplex rotate=landscape i=join;
option nobyline;
proc gplot data = allclear(where=(age <= 240 and age >= 0));
plot HeadCr*Age=Id/haxis=axis1 vaxis=axis2 nolegend;
axis1 label=(h=2 'Age (days)');
axis2 label=(h=2 A=90 ' Head Circumference(cm)');
title1 h=2 'HeadCr vs. Age where not Premature and No Low Body Weight';
run; quit;
```



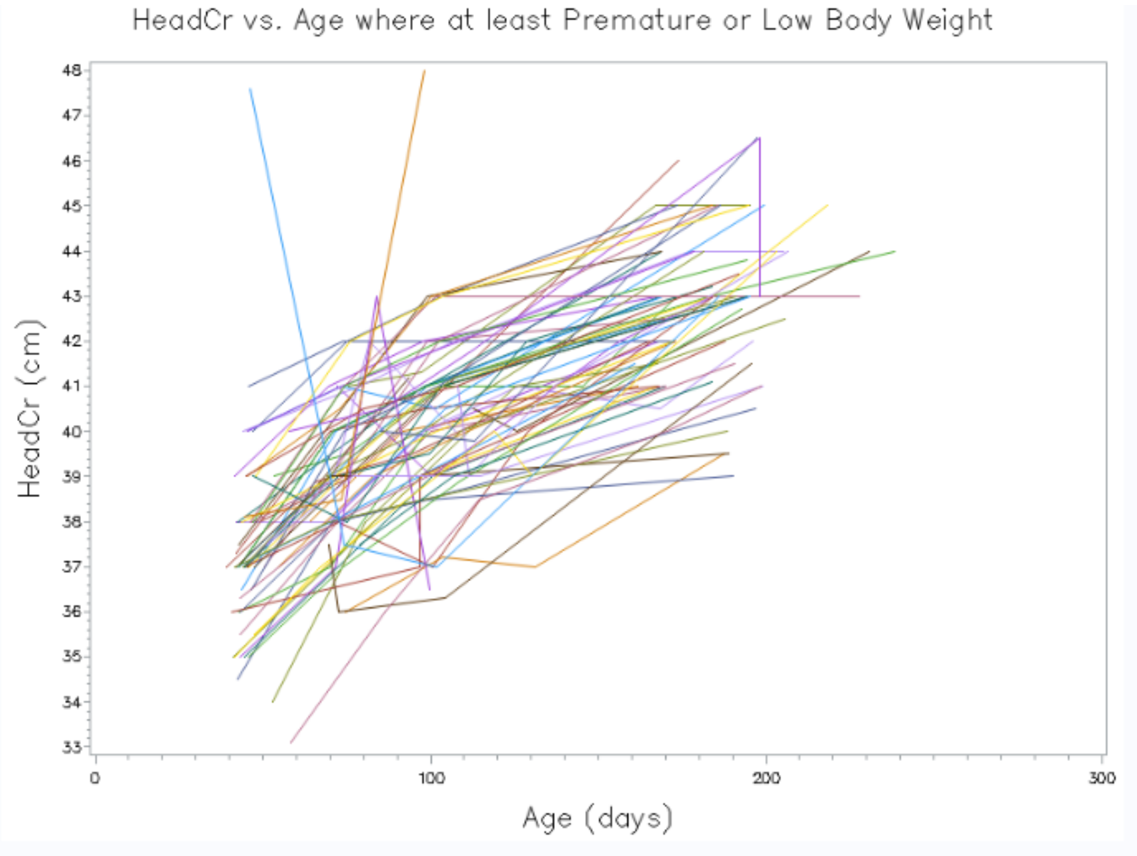
Interpretation: It can be seen in the above individual profile where the babies are born at the usual time where they were expected therefore not premature and also with normal body weight, there are many observations that follow the same pattern which shows normal growth with some outliers of Head Circumference declining after a certain period. The maximum can be seen in quite high ranges such as values of 60cm and nearly up to 70cm with many around the 45cm mark.

```
/* Where Premature = 1 and LowBWeight = 1 */

proc sort data = allnotclear;
by ID;
run;
goptions reset=all ftext=simplex rotate=landscape i=join;
option nobyline;
proc gplot data = allnotclear(where=(age <= 240 and age >= 0));
plot HeadCr*Age=Id/haxis=axis1 vaxis=axis2 nolegend;
axis1 label=(h=2 'Age (days)');
axis2 label=(h=2 A=90 'Head Circumference(cm)');
title1 h=2 'HeadCr vs. Age where Premature and Low Body Weight';
run; quit;
```



Interpretation: In the above profile plot where the babies are born premature and with low body weight, it is obvious there are much fewer observations which than the profile plot where the baby is neither premature nor low body weight. With there being fewer observations, it is clearer to see individual patterns within this profile plot of the babies head circumference growth. There are a few interesting outliers where quite a few of the babies head circumference decreased after a certain period and in this specific profile plot the proportion is much larger than the profile plot before. Also a valuable observation would be the maximum observed in this profile plot can be seen to be around 47cm with many of the larger head circumference observations to be around 44cm and 45cm, this is too a large extent smaller than the profile plot before where the baby was not premature or low body weight

```
/* At least one explanatory variable has occurred*/
proc sort data = atleastone;
by ID;
run;
goptions reset=all ftext=simplex rotate=landscape i=join;
option nobyline;
proc gplot data = atleastone(where=(age <= 240 and age >= 0));
plot HeadCr*Age=Id/haxis=axis1 vaxis=axis2 nolegend;
axis1 label=(h=2 'Age (days)');
axis2 label=(h=2 A=90 'HeadCr (cm)');
title1 h=2 'HeadCr vs. Age where at least Premature or Low Body Weight';
run; quit;
```



Interpretation: This profile plot above is of all the babies born either premature or with low body weight or with both, there is a similar pattern with many of the head circumference maximums still being very low as it is only around 47cm with many around the 44cm mark, therefore similarly to the plot before it is much smaller than the head circumference of babies born not premature and with normal body weight.

```
/* Plot the Smoothed Variance Function */
proc glm data = main_data_new5;
model headcr =age;
output out=test r=resid;
run; quit;
data test;
set test;
r2 = resid**2;
run; quit;
proc loess data = test;
ods output scoreresults=out;
model r2 = age/smooth=0.25;
score data=test;
run;
proc sort data = out; by age; run;
goptions reset = all ftext=simplex rotate=landscape;
proc gplot data = out;
plot r2*age=1 p_r2*age=2 / overlay haxis=axis1 vaxis=axis2;
symbol1 c=red v=dot h=0.2 mode = include;
symbol2 c=black i=join w=2 mode = include;
axis1 label=(h=2 'Age') minor=none;
axis2 label=(h=2 a=90 'Squared Residuals') minor=none;
title h=3 'Smoothed Variance Function';
run; quit;
```

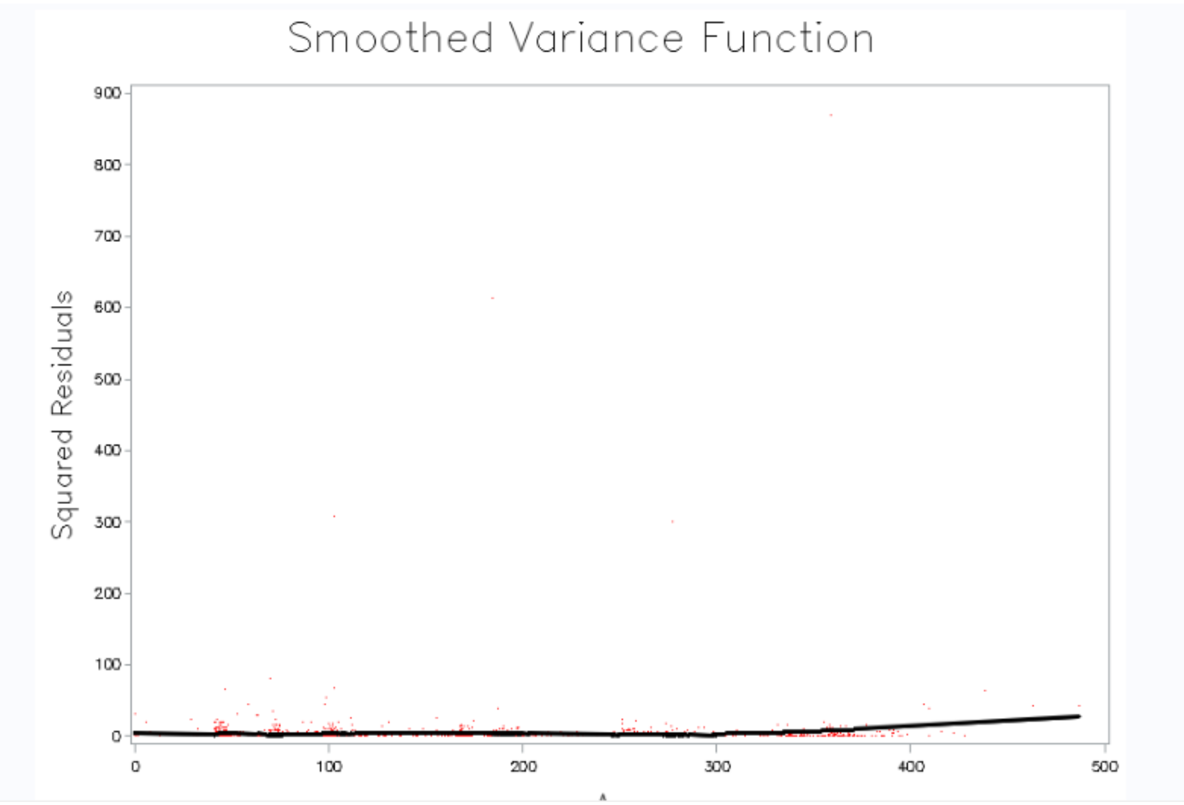


Figure 2.2

```
/* Semi-Variogram; */
/* Calculation of the residuals, linear average trend; */

proc glm data=main_data_new5;
model headcr=age;
output out=out r=residual;
run;

proc variogram data=out outpair=out;
coordinates xc=age yc=id;
compute robust novariogram;
var residual;
run;

data variogram; set out;
if y1=y2; vario=(v1-v2)**2/2; run;
data variance; set out;
if y1 < y2; vario=(v1-v2)**2/2; run;

proc means data=variance mean;
var vario;
run;

# Loess smoothing of the variogram;
proc loess data=variogram;
ods output scoreresults=out;
model vario=distance;
score data=variogram;
run;

proc sort data=out; by distance; run;

goptions reset=all ftext=simplex rotate=landscape;
proc gplot data=out;
plot vario*distance=1 p_vario*distance=2 / overlay haxis=axis1 vaxis=axis2
vref=4.4338 lvref=3;
symbol1 c=red v=dot h=0.5 mode=include;
symbol2 c=black i=join w=2 mode=include;
axis1 label=(h=2 'Time Lag') value=(h=1.5) minor=none;
axis2 label=(h=2 A=90 'v(u)') value=(h=1.5) minor=none;
title h=3 'Semi-Variogram';
run; quit;
```

Semi-Variogram

The MEANS Procedure

Analysis Variable : vario
Mean
4.4337783

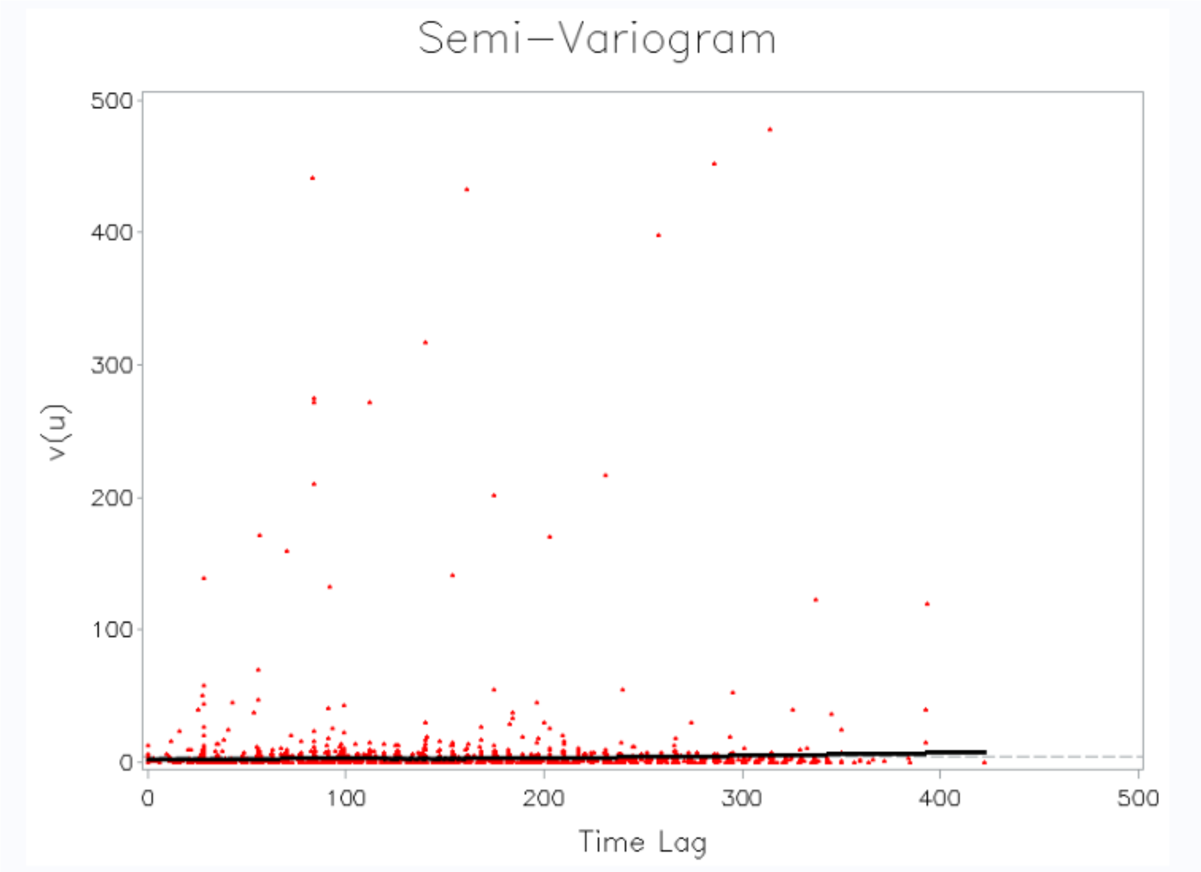


Figure 2.3

Interpretation: It can be seen that the fitted penalized B spline curve does not tend toward zero as the time interval tends to zero, the sample variogram clearly shows that the model has some measurement error (error in the measurement process itself). Furthermore, the fitted line does not have a slope of zero, which indicates that there is serial correlation in the model (Head Circumference measurements vary over time within subject). The serial correlation function appears to be relatively linear. Finally, because the fitted line does reach the process variance, therefore there is no error due to random effects in the model.

Reducing the model

```
proc mixed data=main_data_new5 method =ml;
model Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight
Age*Premature Age*TreatLess4Weeks Age*VLGroup Age*Symptomatic
Age*LowBWeight
Age*Age Age*Age*Age/ solution ddfm=kr(firstorder);
random intercept age /  type = cs subject = id g vcorr;
title 'Longitudinal Model with Compound Symmetry Structure';
run;
```

Fit Statistics	
-2 Log Likelihood	2911.7
AIC (Smaller is Better)	2943.7
AICC (Smaller is Better)	2944.5
BIC (Smaller is Better)	2989.2

Solution for Fixed Effects

Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	36.5936	0.4507	575	81.20	<.0001
Age	0.05175	0.007928	633	6.53	<.0001
Premature	-0.8100	0.4067	577	-1.99	0.0469
TreatLess4Weeks	1.4136	0.8139	557	1.74	0.0830
VLGroup	0.2512	0.1829	592	1.37	0.1701
Symptomatic	0.2435	0.6795	586	0.36	0.7202
LowBWeight	-1.9981	0.5744	587	-3.48	0.0005
Age*Premature	0.002564	0.002649	296	0.97	0.3340
Age*TreatLess4Weeks	0.000758	0.005633	493	0.13	0.8929
Age*VLGroup	-0.00121	0.001189	315	-1.02	0.3100
Age*Symptomatic	0.000741	0.004521	348	0.16	0.8699
Age*LowBWeight	0.001835	0.003754	304	0.49	0.6254
Age*Age	-0.00008	0.000043	631	-1.95	0.0521
Age*Age*Age	2.915E-8	0	0	.	.

Figure 2.4

Interpretation: The higher order term Age*Treatless4Weeks will be dropped as it is has the highest p-value, there is an AICC of 2944.5.

```
proc mixed data=main_data_new5 method =ml;
model Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight
Age*Premature Age*VLGroup Age*Symptomatic Age*LowBWeight
Age*Age Age*Age*Age/ solution ddfm=kr(firstorder);
random intercept age /  type = cs subject = id g vcorr;
title 'Longitudinal Model with Compound Symmetry Structure';
run;
```

Fit Statistics	
-2 Log Likelihood	2911.7
AIC (Smaller is Better)	2941.7
AICC (Smaller is Better)	2942.4
BIC (Smaller is Better)	2984.4

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	36.5976	0.4497	577	81.38	<.0001
Age	0.05170	0.007920	632	6.53	<.0001
Premature	-0.8108	0.4067	577	-1.99	0.0467
TreatLess4Weeks	1.4949	0.5480	376	2.73	0.0067
VLGroup	0.2469	0.1800	606	1.37	0.1708
Symptomatic	0.2509	0.6773	589	0.37	0.7112
LowBWeight	-2.0044	0.5726	590	-3.50	0.0005
Age*Premature	0.002572	0.002649	296	0.97	0.3324
Age*VLGroup	-0.00117	0.001150	323	-1.02	0.3105
Age*Symptomatic	0.000671	0.004491	349	0.15	0.8813
Age*LowBWeight	0.001892	0.003730	305	0.51	0.6124
Age*Age	-0.00008	0.000042	631	-1.94	0.0524
Age*Age*Age	2.874E-8	0	0	.	.

Interpretation: The higher order term Age*Symptomatic will be dropped as it has the highest p-value, with an AICC 2942.2

```
proc mixed data=main_data_new5 method =ml;
model Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight
Age*Premature Age*VLGroup Age*LowBWeight
Age*Age Age*Age*Age/ solution ddfm=kr(firstorder);
random intercept age /  type = cs subject = id g vcorr;
title 'Longitudinal Model with Compound Symmetry Structure';
run;
```

Interpretation: The higher order term Age*LowBWeight was dropped, with an AICC of 2940.4

```
proc mixed data=main_data_new5 method =ml;
model Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight
Age*Premature Age*VLGroup
Age*Age Age*Age*Age/ solution ddfm=kr(firstorder);
random intercept age /  type = cs subject = id g vcorr;
title 'Longitudinal Model with Compound Symmetry Structure';
run;
```

Interpretation: The higher order term Age*VLGroup was dropped, with an AICC of 2938.9

```
proc mixed data=main_data_new5 method =ml;
model Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight
Age*Premature
Age*Age Age*Age*Age/ solution ddfm=kr(firstorder);
random intercept age /  type = cs subject = id g vcorr;
title 'Longitudinal Model with Compound Symmetry Structure';
run;
```

Interpretation: The higher order term Age*Premature was dropped, with an AICC of 2937.9.

```
proc mixed data=main_data_new5 method =ml;
model Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight
Age*Age Age*Age*Age/ solution ddfm=kr(firstorder);
random intercept age /  type =cs subject = id g vcorr;
title 'Longitudinal Model with Compound Symmetry Structure';
run;
```

Interpretation: The cubic effect of age was dropped, with an AICC of 2937.8.

```
proc mixed data=main_data_new5 method =ml;
model Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight
Age*Age / solution ddfm=kr(firstorder);
random intercept age /  type =cs subject = id g vcorr;
title 'Longitudinal Model with Compound Symmetry Structure';
run;
```

Interpretation: The quadratic effect of age was dropped, with an AICC of 2935.9.

```
proc mixed data=main_data_new5 method =ml;
model Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight
/ solution ddfm=kr(firstorder);
random intercept age /  type =cs subject = id g vcorr;
title 'Longitudinal Model with Compound Symmetry Structure';
run;
```

Interpretation: AICC has now increased to the highest value out of all the tests, therefore the main effects will be the final model as all higher order terms are now removed.

Showing Final Model:

```
proc mixed data=main_data_new5;
model Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight /
solution ddfm=kr(firstorder);
random intercept Age / solution type = cs subject = id g vcorr;
title '';
run;
```

Estimated G Matrix

Row	Effect	Subject	Col1	Col2
1	Intercept	1	0.000029	0.000029
2	Age	1	0.000029	0.000029

Estimated V Correlation Matrix for Subject 1

Row	Col1	Col2	Col3	Col4	Col5
1	1.0000	0.003935	0.004781	0.005773	0.007706
2	0.003935	1.0000	0.05329	0.06435	0.08589
3	0.004781	0.05329	1.0000	0.07819	0.1044
4	0.005773	0.06435	0.07819	1.0000	0.1260
5	0.007706	0.08589	0.1044	0.1260	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
Variance	ID	0
CS	ID	0.000029
Residual		3.4669

Fit Statistics

-2 Res Log Likelihood	3003.1
AIC (Smaller is Better)	3007.1
AICC (Smaller is Better)	3007.1
BIC (Smaller is Better)	3012.8

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	42.77	<.0001

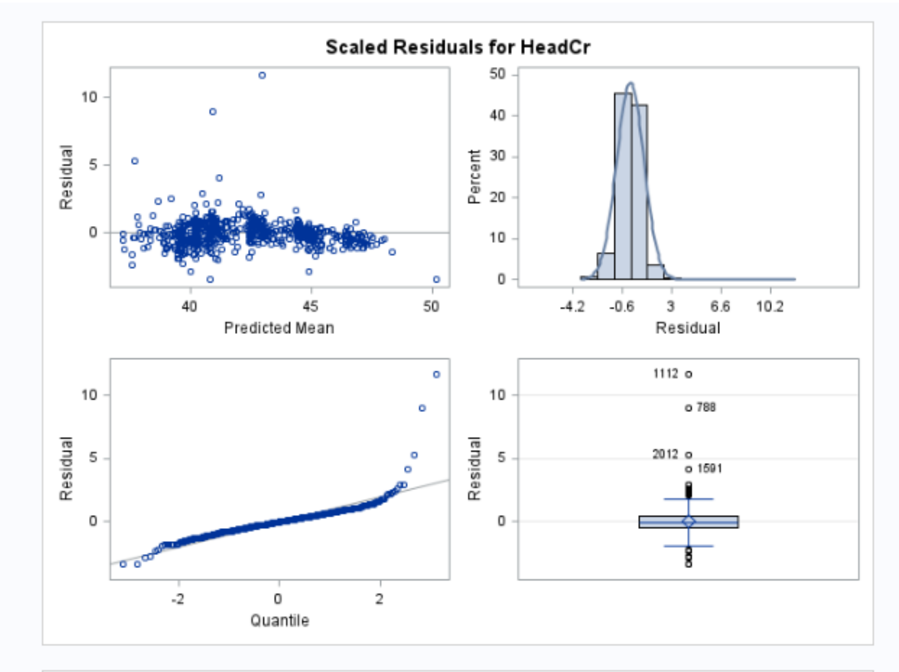
Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	38.3459	0.2346	540	163.45	<.0001
Age	0.02341	0.000866	315	27.04	<.0001
Premature	-0.5055	0.3133	361	-1.61	0.1076
TreatLess4Weeks	1.4479	0.5757	360	2.52	0.0123
VLGroup	0.1472	0.1381	351	1.07	0.2873
Symptomatic	0.3745	0.5072	365	0.74	0.4608
LowBWeight	-1.8001	0.4406	370	-4.09	<.0001

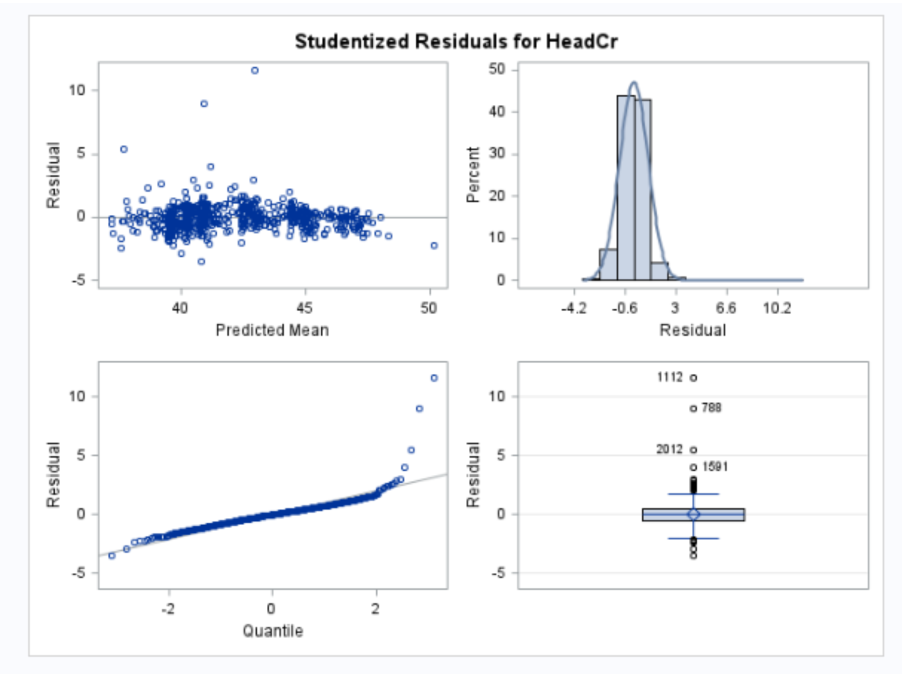
Solution for Random Effects						
Effect	Subject	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Intercept	1	0.01131	0.004497	697	2.52	0.0121
Age	1	0.01131	0.004497	697	2.52	0.0121
Intercept	2	0.001351	0.003959	697	0.34	0.7331
Age	2	0.001351	0.003959	697	0.34	0.7331
Intercept	3	0.002718	0.003840	697	0.71	0.4793
Age	3	0.002718	0.003840	697	0.71	0.4793
Intercept	4	0.004136	0.003928	697	1.05	0.2927
Age	4	0.004136	0.003928	697	1.05	0.2927

Figure 2.5

Interpretation: The G matrix consists of the variances and covariances of the random effects. The value in column 1 and row 1 represents the variance of the intercepts. The value in column 2 and row 2 represents the variance of the slopes for age. The value in column 2 and row 1 represents the covariance of the intercepts and the slopes of age. The information gleaned from the G matrix is that the intercepts and the slopes for age are slightly positvely correlated. When analysing the Solution for Fixed Effects table for the final reduced model, it can be seen that with one unit increase in Age, head circumference increases by 0.02341cm. Therefore there is a positive relationship between the two variables. When analysis the premature variable, it can be seen that if the child is premature it is seen the head circumference decreases by 0.5055cm, therefore a negative relationship exists. VLGroup and Symptomatic have estimates of 0.1472 and 0.3745 respectively, therefore head circumference is expected to increase at those specified amounts with these variables being present. There is significant spikes in the estimates of TreatLess4Weeks and LowBWeight, with LowBWeight having an estimate of -1.8001 therefore a child is estimated to have a 1.8001cm smaller head circumference than a child who doesn't have low body weight. To add to this, TreatLess4weeks also seems to show a significant estimate with an estimate of 1.4479 therefore a child born with a mother that has had ARV treatment in the past 4 weeks is estimated to have a child with 1.4479cm greater head circumference than a child whose mother wasn't on ARV treatment 4 weeks prior to giving birth.

```
ods listing close;
ods ouput solutionr=randomeffects;
proc mixed data=main_data_new5 plots=( distance press studentpanel(marginal
box) pearsonpanel(marginal box) vcirypanel(box));
model Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight /
s influence(effect=id iter=5) vciry;
random intercept age/ type=cs subject=id g v solution;
title 'Kidney: Random Intercept + Slope, Simple';
run;
ods listing;
proc sort data=randomeffects;
   by effect;
run;
```





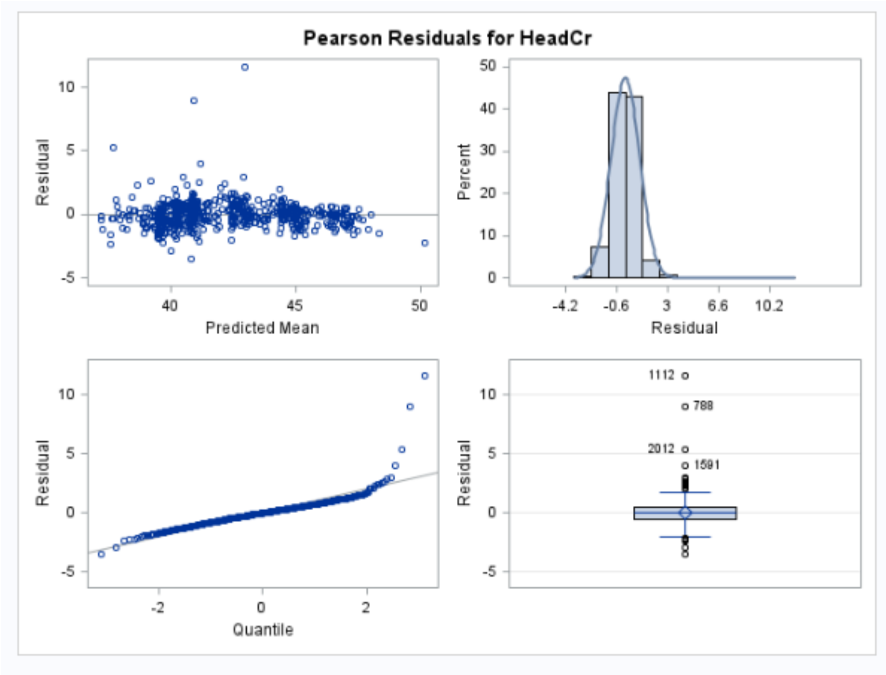


Figure 2.6

```
proc mixed data=main_data_new5 noclprint;
model  Headcr=Age Premature TreatLess4Weeks VLGroup Symptomatic LowBWeight
/ s outp=predicted;
random intercept age/ type=un subject=id g v;
title 'Head Circumference: Random Intercept + Slope, Simple';
run;

proc gplot data=predicted;
   plot resid*(age) / vref=0;
   symbol v=star c=blue;
run;
quit;
```