# NHANES data analysis

No. of observations = 23118

Var. name obs. mean median s.d. min. max.

1 muac 23118 227.29 222 50.35 59 499

2 muac\_left\_arm 1788 229.82 226 57.41 107 485

3 dataset

4 agey 23118 12.29 12.08 5.34 1.98 26.02

5 sex 23118 1.491 1 0.5 1 2

6 muac\_female 11761 226.35 224 48.22 59 499

7 muac\_male 11357 228.26 220 52.45 105 483

8 l\_r\_diff 1788 1.62 1 8.48 -92 110

Converted muac to cm and age to months and dropped if age was less than 2 and greater than 25 years

No. of observations = 22699

Var. name obs. mean median s.d. min. max.

1 sex 22699 1.494 1 0.5 1 2

2 muac 22699 22.62 22.1 4.98 5.9 49.9

3 agem 22699 144.72 143 60.92 24.05 299.99

Gender distribution

Frequency Percent Cum. percent

female 11481 50.6 50.6

male' 11218 49.4 100.0

Total 22699 100.0 100.0

Boys dataset

Var. name obs. mean median s.d. min. max.

1 sex 11218 2 2 0 2 2

2 muac 11218 22.73 21.8 5.18 10.5 48.3

3 agem 11218 140.58 140.25 57.07 24.11 299.93

Girls dataset

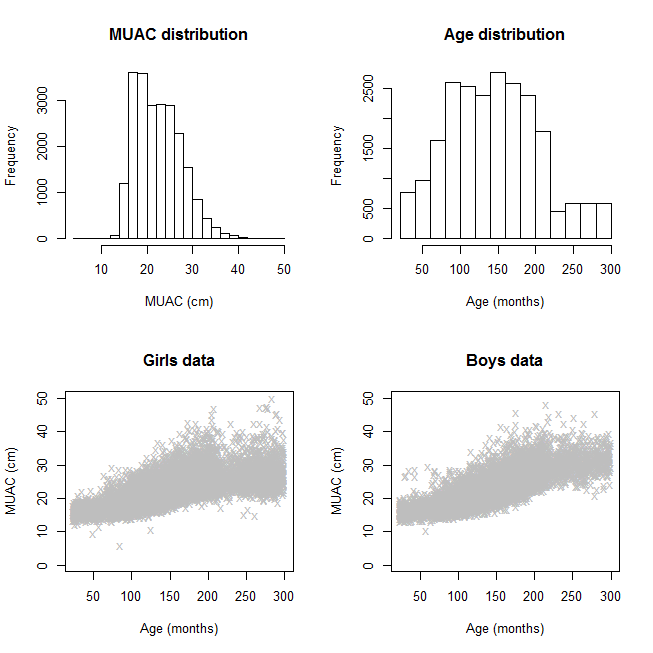
Var. name obs. mean median s.d. min. max.

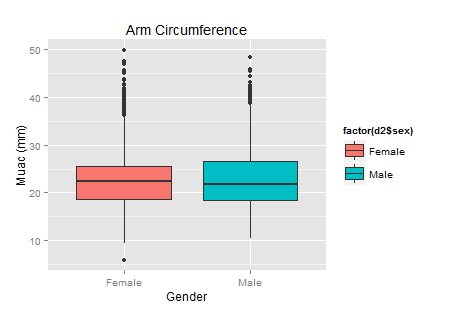
1 sex 11481 1 1 0 1 1

2 muac 11481 22.51 22.3 4.77 5.9 49.9

3 agem 11481 148.76 145.28 64.22 24.05 299.99

Explanatory Data Analysis

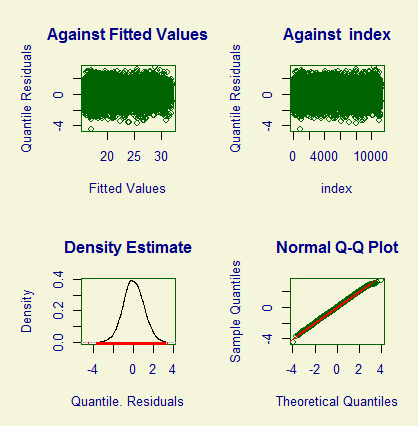


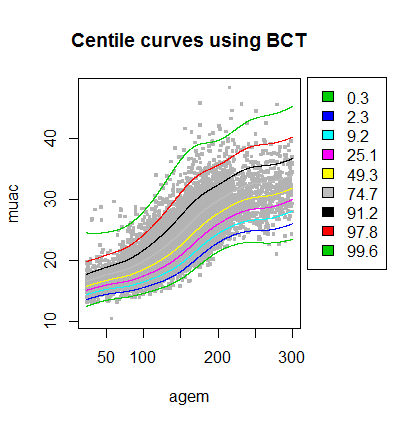


# Model Selection

## Boys model selection using LMS method

Note that the BCT model was selected as the best with global deviance = 51534.6 other than BCCG and BCPE at this stage.





## Selecting Models by fitting the gamlss function and using AIC and GAIC methods.

For both boys and girls datasets, the best models selected were BCT using the AIC and GAIC criterion.

|  |
| --- |
| hf1<-gamlss(muac~cs(agem), sigma.fo=~cs(agem), nu.fo=~cs(agem),tau.fo=~cs(agem),  + data=girlsd, family=BCT)  GAMLSS-RS iteration 1: Global Deviance = 55080.46  GAMLSS-RS iteration 2: Global Deviance = 55026.28  GAMLSS-RS iteration 3: Global Deviance = 55025.3  GAMLSS-RS iteration 4: Global Deviance = 55024.87  GAMLSS-RS iteration 5: Global Deviance = 55024.69  GAMLSS-RS iteration 6: Global Deviance = 55024.61  GAMLSS-RS iteration 7: Global Deviance = 55024.59  GAMLSS-RS iteration 8: Global Deviance = 55024.58  GAMLSS-RS iteration 9: Global Deviance = 55024.57  GAMLSS-RS iteration 10: Global Deviance = 55024.57  GAMLSS-RS iteration 11: Global Deviance = 55024.57  GAMLSS-RS iteration 12: Global Deviance = 55024.57  GAMLSS-RS iteration 13: Global Deviance = 55024.57  > hf0<-gamlss(muac~cs(agem), sigma.fo=~cs(agem), nu.fo=~cs(agem),tau.fo=~cs(agem),  + data=girlsd, family=BCPE)  GAMLSS-RS iteration 1: Global Deviance = 55585.37  GAMLSS-RS iteration 2: Global Deviance = 55123.79  GAMLSS-RS iteration 3: Global Deviance = 55110.92  GAMLSS-RS iteration 4: Global Deviance = 55111.26  GAMLSS-RS iteration 5: Global Deviance = 55111.49  GAMLSS-RS iteration 6: Global Deviance = 55111.55  GAMLSS-RS iteration 7: Global Deviance = 55111.55  GAMLSS-RS iteration 8: Global Deviance = 55111.55  GAMLSS-RS iteration 9: Global Deviance = 55111.55  > AIC(hf1,hf0,k=3)  df AIC  hf1 20.00271 55084.58  hf0 20.00276 55171.56  > GAIC(hf1,hf0,k=3)  df AIC  hf1 20.00271 55084.58  hf0 20.00276 55171.56  > hm1<-gamlss(muac~cs(agem),nu.fo=~cs(agem),tau.fo=~cs(agem),  + sigma.fo=~cs(agem), data=boysd, family=BCT)  GAMLSS-RS iteration 1: Global Deviance = 51658.36  GAMLSS-RS iteration 2: Global Deviance = 51619.28  GAMLSS-RS iteration 3: Global Deviance = 51617.7  GAMLSS-RS iteration 4: Global Deviance = 51617.28  GAMLSS-RS iteration 5: Global Deviance = 51617.15  GAMLSS-RS iteration 6: Global Deviance = 51617.11  GAMLSS-RS iteration 7: Global Deviance = 51617.1  GAMLSS-RS iteration 8: Global Deviance = 51617.09  GAMLSS-RS iteration 9: Global Deviance = 51617.09  GAMLSS-RS iteration 10: Global Deviance = 51617.09  GAMLSS-RS iteration 11: Global Deviance = 51617.1  GAMLSS-RS iteration 12: Global Deviance = 51617.1  GAMLSS-RS iteration 13: Global Deviance = 51617.1  > hm0<-gamlss(muac~cs(agem),nu.fo=~cs(agem),tau.fo=~cs(agem),  + sigma.fo=~cs(agem), data=boysd, family=BCPE)  GAMLSS-RS iteration 1: Global Deviance = 52017.19  GAMLSS-RS iteration 2: Global Deviance = 51658.09  GAMLSS-RS iteration 3: Global Deviance = 51644.59  GAMLSS-RS iteration 4: Global Deviance = 51644.37  GAMLSS-RS iteration 5: Global Deviance = 51644.47  GAMLSS-RS iteration 6: Global Deviance = 51644.53  GAMLSS-RS iteration 7: Global Deviance = 51644.55  GAMLSS-RS iteration 8: Global Deviance = 51644.55  GAMLSS-RS iteration 9: Global Deviance = 51644.55  > AIC(hm1,hm0,k=3)  df AIC  hm1 20.00016 51677.10  hm0 20.00230 51704.55  > GAIC(hm1,hm0,k=3)  df AIC  hm1 20.00016 51677.10  hm0 20.00230 51704.55 |
|  |
|  |
|  |
|  |

# Generating Z-scores for boys and girls separately using the above fitted models

## Boys data

Var. name obs. mean median s.d. min. max.

1 sex 11218 2 2 0 2 2

2 muac 11218 22.73 21.8 5.18 10.5 48.3

3 agem 11218 140.58 140.25 57.07 24.11 299.93

4 males.Zscores 11218 0 -0.01 1 -4.3 3.3

## Girls data

Var. name obs. mean median s.d. min. max.

1 sex 11481 1 1 0 1 1

2 muac 11481 22.51 22.3 4.77 5.9 49.9

3 agem 11481 148.76 145.28 64.22 24.05 299.99

4 females.Zscores 11481 0 -0.02 1 -6.37 3.19

Then, we removed individulas with Z scores >+3SD or <-3SD.

Boys data

Var. name obs. mean median s.d. min. max.

1 sex 11189 2 2 0 2 2

2 muac 11189 22.73 21.8 5.16 13 45.5

3 agem 11189 140.61 140.35 57.05 24.11 299.93

4 males.Zscores 11189 0 -0.01 0.99 -3 2.99

Girls data

Var. name obs. mean median s.d. min. max.

1 sex 11466 1 1 0 1 1

2 muac 11466 22.52 22.3 4.76 12.3 49.9

3 agem 11466 148.76 145.28 64.21 24.05 299.99

4 females.Zscores 11466 0.01 -0.02 0.99 -2.94 2.96

Bringing in simulated data

Frequency Percent Cum. percent

Male 5969 49.7 49.7

Female 6031 50.3 100.0

Total 12000 100.0 100.0

Simulated.boys data

No. of observations = 5969

Var. name obs. mean median s.d. min. max.

1 sex 5969 1 1 0 1 1

2 muac 5969 16.13 16.04 1.43 11.64 23.76

3 agem 5969 46.93 46.88 13.32 24.03 70.03

4 males.Zscores 5969 -0.02 -0.02 1.01 -3.53 3.78

Simulated.girls data

No. of observations = 6031

Var. name obs. mean median s.d. min. max.

1 sex 6031 2 2 0 2 2

2 muac 6031 16.28 16.15 1.63 11.8 23.87

3 agem 6031 47.05 47.08 13.38 24 70.03

4 females.Zscores 6031 0.02 0.01 1.01 -3.32 3.73

The next step involved merging the datasets

Merged.boys

No. of observations = 17158

Var. name obs. mean median s.d. min. max.

1 sex

2 muac 17158 20.43 18.5 5.29 11.6 45.5

3 agem 17158 108.02 94 64.61 24 299.9

4 males.Zscores 17158 0 0 1.03 -4 4

Merged.girls

No. of observations = 17497

Var. name obs. mean median s.d. min. max.

1 sex

2 muac 17497 20.37 18.9 4.95 11.8 49.9

3 agem 17497 113.7 97 71.41 24 300

4 females.Zscores 17497 0.01 0 1.04 -3 4

BCPE and BCT models were fitted using the gamlss function to the merged datasets separately for boys and girls.

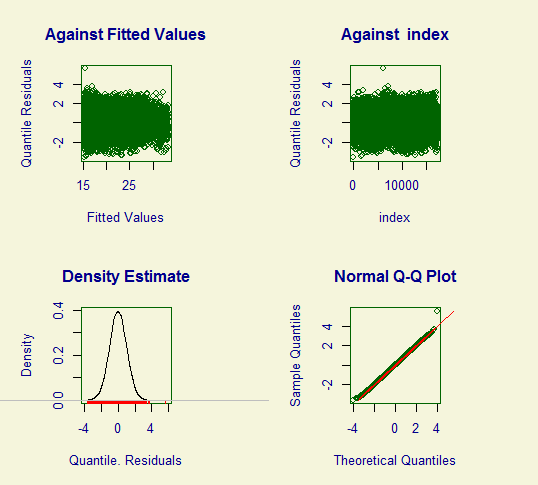
Model selection was done using AIC and GAIC criterion. The best models selected were BCPE in both groups.

m1<-gamlss(muac~cs(agem),sigma.fo=~cs(agem), nu.fo=~cs(agem),tau.fo=~cs(agem),data=merged.boys, family=BCPE)

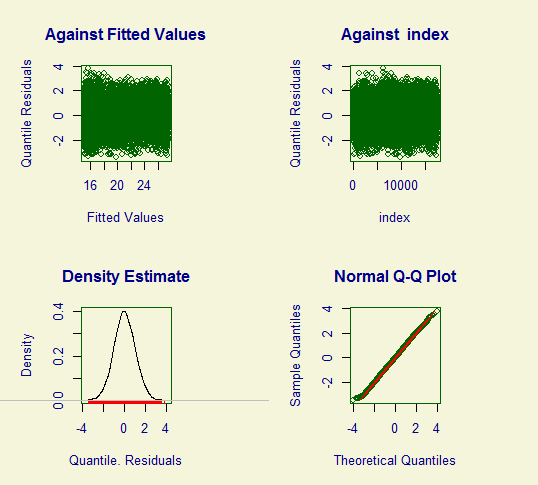
f1<-gamlss(muac~cs(agem),sigma.fo=~cs(agem),nu.fo=~cs(agem),tau.fo=~cs(agem), data=merged.girls, family=BCPE)

# Diagnostic plots: Worm Plots

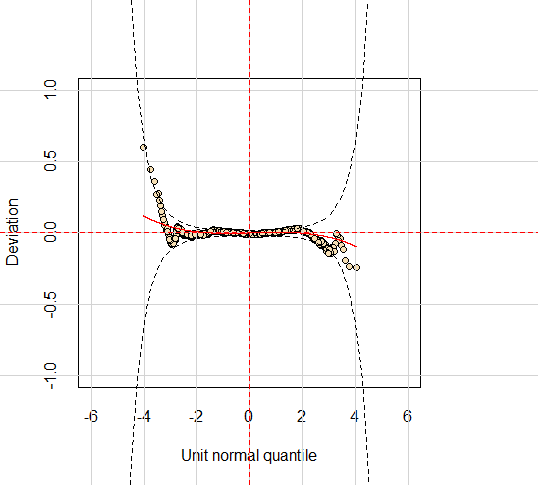
M1 plot



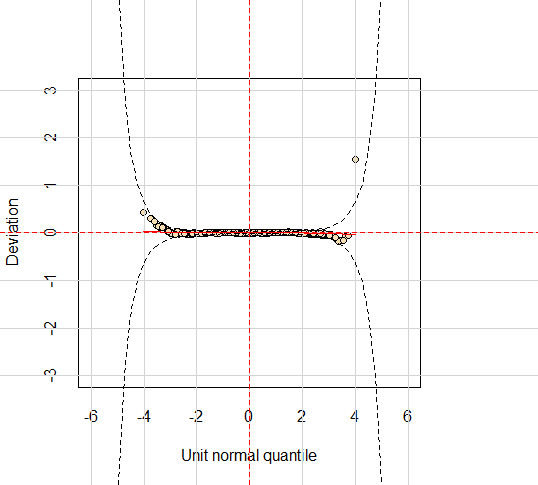
F1 plot



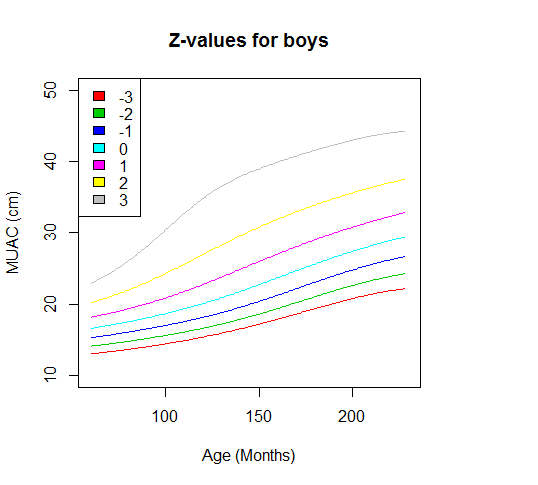
Worm plot for f1

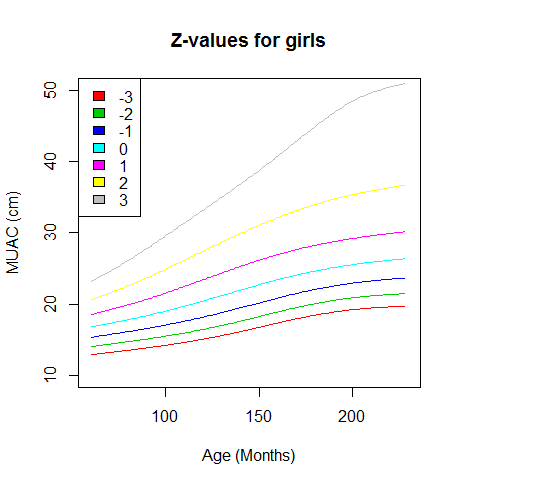


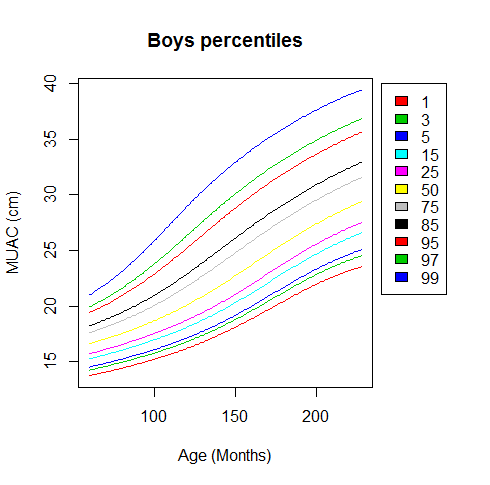
Worm plot for M1

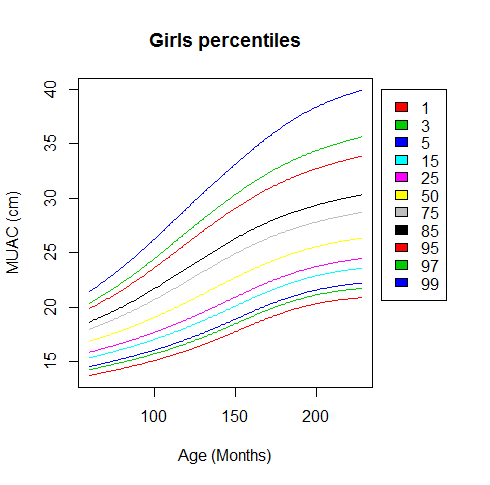


# Generate WHO-like Tables of Predicted Z-scores and percentiles with LMS at different ages given muac values.









# Arrow Dataset

## Predicting Z-scores for all individuals in the arrow dataset based on the merged dataset aboave (NHANES+simulated data)

No. of observations = 685

Var. name obs. mean median s.d. min. max.

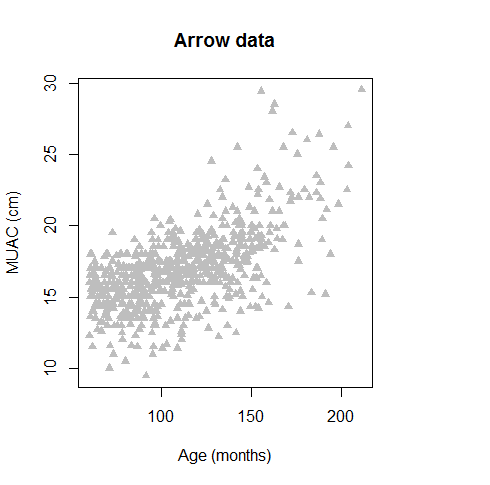
1 id 685 603.9 618 333.98 9 1206

2 sex 685 1.502 2 0.5 1 2

3 muac 685 16.88 16.7 2.68 9.5 29.5

4 agey 685 9.16 8.9 2.63 5.01 17.61

5 agem 685 109.87 106.74 31.61 60.16 211.35





arrow$sex :

Frequency Percent Cum. percent

Male 341 49.8 49.8

Female 344 50.2 100.0

Total 685 100.0 100.0

## Arrow.boys

No. of observations = 341

Var. name obs. mean median s.d. min. max.

1 id 341 595.25 632 329.63 10 1206

2 sex 341 1 1 0 1 1

3 muac 341 16.52 16.5 2.3 9.5 25.5

4 agey 341 8.97 8.61 2.66 5.03 16.54

5 agem 341 107.69 103.29 31.87 60.35 198.44

One individual could not be assigned a Z-score. I guess because the MUAC value was lower than that of the lowest muac value from the merged data set.

### id sex muac agey agem zscores\_m

821 Male 9.5 7.627652 91.53183 -Inf

I eliminated the individual in row 257 whose id is 821

No. of observations = 340

Var. name obs. mean median s.d. min. max.

1 id 340 594.58 626 329.89 10 1206

2 sex 340 1 1 0 1 1

3 muac 340 16.54 16.5 2.27 10 25.5

4 agey 340 8.98 8.63 2.66 5.03 16.54

5 agem 340 107.74 103.61 31.9 60.35 198.44

6 zscores\_m 340 -1.94 -1.7 1.58 -8.16 1.29

## Arrow.girls

No. of observations = 344

Var. name obs. mean median s.d. min. max.

1 id 344 612.48 613 338.5 9 1205

2 sex 344 2 2 0 2 2

3 muac 344 17.24 17 2.97 11 29.5

4 agey 344 9.34 9.2 2.6 5.01 17.61

5 agem 344 112.04 110.36 31.26 60.16 211.35

> summ(girlsdf) after predicting Z-scores for all girls

No. of observations = 344

Var. name obs. mean median s.d. min. max.

1 id 344 612.48 613 338.5 9 1205

2 sex 344 2 2 0 2 2

3 muac 344 17.24 17 2.97 11 29.5

4 agey 344 9.34 9.2 2.6 5.01 17.61

5 agem 344 112.04 110.36 31.26 60.16 211.35

6 zscores\_f 344 -1.57 -1.41 1.35 -6.58 1.61

See attached tables